











Identifying Valid Algorithms for Number of Lines of Anti-Neoplastic Therapy in the Danish National Patient Registry Among Patients with Advanced Ovarian, Gastric, Renal Cell, Urothelial, and Non-Small Cell Lung Cancer Attending a Danish University Hospital

Signe Sørup ¹, Bianka Darvalics ¹, Jakob Schöllhammer Knudsen¹, Anne Staub Rasmussen ¹, Cathrine Fønnesbech Hjorth ¹, Søren Viborg Vestergaard ¹, Azza Ahmed Khalil ², Leo Russo³, Dina Oksen ⁴, Emmanuelle Boutmy ⁴, Patrice Verpillat ⁴, Mikael Rørth^{1,5}, Deirdre Cronin-Fenton ¹

¹Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University & Aarhus University Hospital, Aarhus, Denmark;

²Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; ³Worldwide Medical and Safety, Pfizer, Collegeville, PA, USA; ⁴Global Epidemiology, Merck Healthcare KGaA, Darmstadt, Germany; ⁵Department of Oncology, Rigshospitalet, Copenhagen, Denmark

Correspondence: Signe Sørup, Department of Clinical Epidemiology, Aarhus University Hospital, Olof Palmes Allé 43-45, Aarhus, DK-8200, Denmark, Tel +45 871 68230, Fax +45 87 16 72 15, Email SigneSorup@clin.au.dk

Purpose: To develop algorithms to identify number of lines of anti-neoplastic therapy per patient based on the Danish National Patient Registry (DNPR) and identify which algorithm has the highest percentage agreement with a reference standard of documentation in medical records.

Patients and Methods: We included 179 patients diagnosed between January 1, 2012, and December 31, 2016, with stage II, III, or IV urothelial cell carcinoma or stage III or IV epithelial ovarian cancer, gastric adenocarcinoma, renal cell carcinoma, or non-small cell lung cancer (NSCLC). We developed two algorithms for number of lines of anti-neoplastic therapy based on dates and treatment codes (eg, “treatment with cisplatin” or “cytostatic treatment”) in the DNPR. First, to denote a change in line of therapy the “Time-based algorithm” used the number of days between consecutive administrations. Second, the “Drug-based algorithm” used information on drug names if available or the number of days between consecutive administrations if no drug names were specified. We calculated the percentage agreement between the algorithms setting the number of allowed days between consecutive administrations from 28 to 50 and the reference standard – information on anti-neoplastic therapy drugs abstracted from medical records and subsequently coded according to lines of anti-neoplastic therapy.

Results: For the “Time-based algorithm”, the highest percentage agreement with the reference standard was found when using <45 days between consecutive administrations (67.6%; 95% CI: 60.1–73.8%). However, the percentage agreement was higher for the “Drug-based algorithm” using <45 days between consecutive administrations for registrations where the drug name was unspecified (90.5%; 95% CI: 85.0–93.7%).

Conclusion: The algorithm for number of lines of anti-neoplastic therapy that had the highest percentage agreement with the reference standard (medical records) incorporated both registration of specific drug names and <45 days between consecutive administrations if the drug name was unspecified in routinely recorded data from DNPR.

Keywords: positive predictive value, medical records review, duration of chemotherapy, medical oncologic treatments, Denmark

Introduction

Anti-neoplastic therapy comprises key treatments for many advanced cancers. Anti-neoplastic therapy includes chemotherapy and targeted therapies like angiogenesis inhibitors, kinase inhibitors, proteasome inhibitors, transferase inhibitors, polymerase inhibitors, histone deacetylase inhibitors, and others (arsenic trioxide, BCL-2 inhibitors). Treatment with anti-neoplastic therapy usually starts with a standard drug or combination of drugs, which is termed the first line of anti-neoplastic therapy.¹ In the advanced setting, the first line of anti-neoplastic therapy depends on patient and disease characteristics. Following disease progression or unacceptable side effects, patients might be offered a second line of anti-neoplastic therapy – usually consisting of other types of anti-neoplastic therapy or other combinations of anti-neoplastic therapy than those used in the first line.^{1–3} Data on the number of lines of anti-neoplastic therapy are important to evaluate treatment results, treatment-induced side effects, survival, and other health care utilizations.⁴ Yet, information on lines of anti-neoplastic therapy is rarely registered in electronic health care databases.⁵ Several electronic health care databases include other information about anti-neoplastic therapy (such as drug type and dates of treatment).^{6–9} Studies including information about lines of anti-neoplastic therapy often rely on obtaining this information from medical records,^{4,10,11} which is costly and time-consuming thereby limiting the size of the study population. The ability to define lines of anti-neoplastic therapy based on registers or other electronic databases is important for future cancer research.

In Denmark, information on dates of treatment with anti-neoplastic therapy is included in the Danish National Patient Registry (DNPR) covering all Danish hospitals.¹² Yet, the registry does not record data specifying the number of lines of anti-neoplastic therapy. We therefore developed two DNPR-based algorithms to estimate the number of lines of anti-neoplastic therapy and examined the agreement between the algorithms and information from medical records on number of lines of anti-neoplastic therapy. Furthermore, we assessed the positive predictive value (PPV) of each DNPR-based

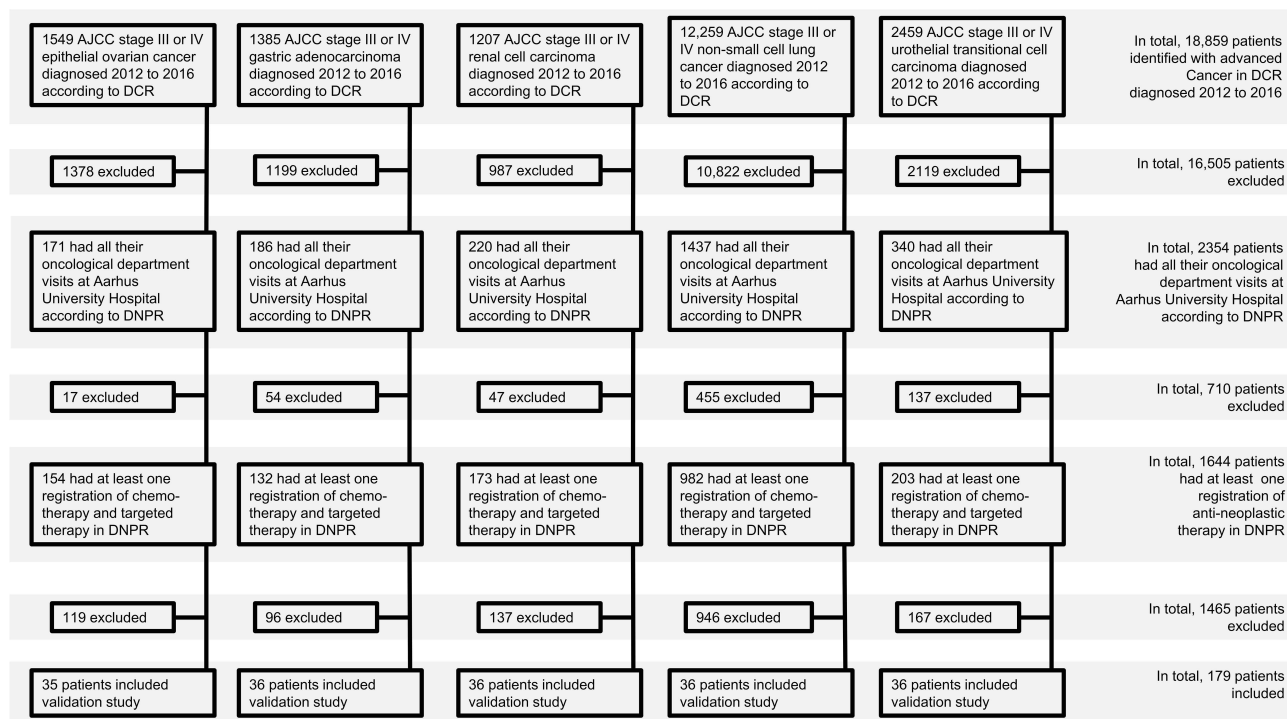


Figure 1 Flowchart for selection of patients for the validation of lines of anti-neoplastic therapy.

Notes: Stage was defined according to the 7th Edition of the AJCC Cancer Staging Manual. In accordance with guidelines issued by the Danish Health Data Authority, we were not permitted to report counts with fewer than 5 observations. Therefore, we are not able to give more details about the reasons for exclusion in the last step of the flowchart.

Abbreviations: AJCC, American Joint Committee on Cancer; DCR, Danish Cancer Registry; DNPR, Danish National Patient Registry.

algorithm based on the number of lines of anti-neoplastic therapy, start date of anti-neoplastic therapy lines, and duration of anti-neoplastic therapy lines.

Materials and Methods

The Danish national health care system provides the entire Danish population with unrestricted access to tax-supported health care with free-of-charge treatment at hospitals.¹³ Furthermore, there is a long tradition for nationwide registers on health care: the Danish Cancer Registry (DCR) was established in 1943¹⁴ and the DNPR in 1977.¹² We used the unique personal identification number assigned to all Danish residents to link information in these registers and the Danish Civil Registration System.^{15,16}

Population

This validation study is part of a larger research project using information from medical registers to characterize Danish patients with advanced cancer that may be targeted for treatment with programmed death receptor 1/programmed death ligand 1 immune checkpoint inhibitors. [Figure 1](#) shows the sampling strategy for the population for the present study. From the DCR, we identified all patients diagnosed with American Joint Committee on Cancer¹⁷ stage III or IV epithelial ovarian cancer, stage III or IV gastric adenocarcinoma, stage III or IV renal cell carcinoma, stage III or IV non-small cell lung cancer (NSCLC), or stage II, III, or IV urothelial cell carcinoma in Denmark between January 1, 2012 and December 31, 2016. Thus, all patients had advanced cancer at the date of diagnosis and the lines of anti-neoplastic therapy we observe relate to the advanced cancer setting. Among these patients, we selected all patients who had all their oncological department visits at Aarhus University Hospital, where we had access to patient medical records. We restricted it to patients who had at least one registration of anti-neoplastic therapy according to treatment codes in the DNPR. Among these, we made a stratified random sample of up to 36 patients for each cancer type stratified by year of diagnosis (2012–2016) and number of anti-neoplastic therapy lines according to the register-based Time-based algorithm with 28 days as illustrated in [sFigure 1](#) in online [Appendix 1](#). We included number of anti-neoplastic therapy lines in the stratification, to ensure the ability to validate the algorithm both for patients receiving few and many lines of anti-neoplastic therapy.

Follow-Up

We included the patients from the date of diagnosis as registered in the DCR until the first of the following events: death, emigration, new primary cancer (as any anti-neoplastic therapy given after this date could be related to the new primary cancer), or August 1, 2018 (the date when we obtained the first medical records for the data abstraction).

Collection and Abstraction of Data from Medical Records

For the selected patients, we obtained the medical records from the Department of Oncology at Aarhus University Hospital from the date of cancer diagnosis onwards. We developed an abstraction form and collected dates and types of anti-neoplastic therapy using the data capture software REDcap.¹⁸ The abstraction form is displayed in online [Appendix 2](#). Five abstractors did the abstraction. To minimize inter-abstractor variation and ensure consistency in abstraction, we established a priori guidelines and a codebook to guide the medical record abstraction. Furthermore, uncertain cases were adjudicated in a multidisciplinary team.

Reference Standard for Number of Lines of Anti-Neoplastic Therapy

The reference standard for number of lines of anti-neoplastic therapy was defined based on the information about dates and types of anti-neoplastic therapy retrieved from the medical records since the first date of cancer diagnosis according to DCR. We defined a line of anti-neoplastic therapy as a series of treatments with the same chemotherapy or targeted therapy drug (or combination of drugs), whereas change in drugs or addition of drugs was considered a change in anti-neoplastic therapy line. If a chemotherapy or targeted therapy drug was included in one line of anti-neoplastic therapy, omitted from a subsequent line, and was administered again more than 6 months after the last administration of that drug, we regarded it as a new line of anti-neoplastic therapy.

Since cisplatin and carboplatin can be used interchangeably, we regarded them as the same drugs for this definition.¹⁹ Likewise, we considered doxorubicin and doxorubicin (liposomal) and nab-paclitaxel and paclitaxel as the same drugs. Niraparib and neratinib were combined due to low numbers. Based on this, we counted the number of anti-neoplastic therapy lines for each patient.

Register-Based Algorithms of Number of Lines of Anti-Neoplastic Therapy

In Denmark, cancer treatment is delivered in hospitals, and anti-neoplastic therapy is coded in the DNPR using Danish treatment codes.^{12,13} Danish treatment codes can be very specific (eg, “treatment with cisplatin”) or very broad (eg, “cytostatic treatment”). Thus, we could not use information only on the specific anti-neoplastic therapy drugs given to identify the number of lines of anti-neoplastic therapy. We examined the validity of two different algorithms to identify the number of lines of anti-neoplastic therapy. First, the Time-based algorithm that only incorporated the number of days between administrations of anti-neoplastic therapy. Second, the Drug-based algorithm incorporated the specific anti-neoplastic therapy drugs if registered in the DNPR and number of days between administrations of anti-neoplastic therapy when the specific drugs were not registered in the DNPR. We validated both algorithms setting the number of days between two consecutive administrations of chemotherapy and/or targeted therapy to all numbers between and including 28 and 50. Both algorithms are further described and illustrated in online [Appendix 1 \(sFigures 1 and 2a–m\)](#).

For each anti-neoplastic therapy line, the start date of the line was defined as the first date of administration of anti-neoplastic therapy in that line. The last date of the line was defined as the last date of administration of anti-neoplastic therapy in that line plus the number of days allowed between consecutive administrations of anti-neoplastic therapy in the algorithm (ie, numbers between and including 28 and 50). However, for the Drug-based algorithm, the next line of anti-neoplastic therapy might occur earlier than the last date of the line as defined in the previous sentence. In such cases, the last date of the line was defined as the day before the initiation of the subsequent line of anti-neoplastic therapy. The duration of each anti-neoplastic therapy line was calculated as the last date of that line minus the start date of that line.

Statistical Analyses

We examined the percentage agreement (and associated 95% confidence interval [CI] using the Jeffrey’s method²⁰) between the register-based algorithms for number of lines of anti-neoplastic therapy and the reference standard from the medical records for all cancer types combined and separately by type of cancer.

For the Time-based algorithm and the Drug-based algorithm with the greatest percentage agreement according to number of days between consecutive administrations of anti-neoplastic therapy for all cancer types combined, we estimated three PPV. First, the PPV of having at least 1, 2 or 3 anti-neoplastic therapy lines (see exact definitions [sTable 1, Online Appendix 3](#)). Second, the PPV for the start date of each line plus or minus 30 days, 60 days, and 90 days (see exact definitions [sTable 2, Online Appendix 3](#)). Third, the PPV for the duration of line 1, 2 or 3 plus or minus 30 days, 60 days, and 90 days (see exact definitions [sTable 3, Online Appendix 3](#)).

The algorithms were developed and analyzed using SAS version 9.4.

Ethics

We obtained permission to perform this validation study from the Danish Data Protection Agency and the Danish Patient Safety Authority (3-3013-2046/1). Under Danish law, ethical permission is not required for registry-based research. This study followed the EU General Data Protection Regulation. The register-based data were obtained from the Danish Health Data Authority.

Results

We included 179 patients in the validation of register-based algorithms for anti-neoplastic therapy lines ([Figure 1](#)). The patients were followed from the date of diagnosis up to death (n=120; 67.0%), emigration (0), new primary cancer (n=11; 6.1%), or end of follow-up (August 1, 2018) (n=48; 26.8%), whichever came first. [Table 1](#) outlines characteristics of the included patients.

Table 1 Description of the Included Patients with Advanced Epithelial Ovarian Cancer, Gastric Adenocarcinoma, Renal Cell Carcinoma, Urothelial Cell Carcinoma, and Non-Small Cell Lung Cancer Diagnosed in Denmark 2012 to 2016

	Epithelial Ovarian Cancer (N=35)	Gastric Adenocarcinoma (N=36)	Renal Cell Carcinoma (N=36)	Urothelial Cell Carcinoma (N=36)	Non-Small Cell Lung Cancer (N=36)	All Cohorts Combined (N=179)
Year of cancer diagnosis, N (%)						
2012	7 (20.0%)	7 (19.4%)	7 (19.4%)	7 (19.4%)	7 (19.4%)	35 (19.6%)
2013	8 (22.9%)	8 (22.2%)	8 (22.2%)	9 (25.0%)	8 (22.2%)	41 (22.9%)
2014	7 (20.0%)	7 (19.4%)	7 (19.4%)	7 (19.4%)	7 (19.4%)	35 (19.6%)
2015	7 (20.0%)	8 (22.2%)	8 (22.2%)	8 (22.2%)	8 (22.2%)	39 (21.8%)
2016	6 (17.1%)	6 (16.7%)	6 (16.7%)	5 (13.9%)	6 (16.7%)	29 (16.2%)
Stage at diagnosis, N (%)^a						
Stage II	Not included	Not included	Not included	< 10	Not included	6 (3.4%)
Stage III	15 (42.9%)	< 10	< 10	< 5	< 15	50 (27.9%)
Stage IV	15 (42.9%)	24 (66.7%)	25 (69.4%)	20 (55.6%)	21 (58.3%)	105 (58.7%)
Undefined stage ^b	5 (14.3%)	< 5	< 5	7 (19.4%)	< 5	18 (10.1%)
Female, N (%)	35 (100.0%)	10 (27.8%)	15 (41.7%)	12 (33.3%)	18 (50.0%)	90 (50.3%)
Age at diagnosis (years), Median (q1-q3)	68.3 (60.2–75.9)	66.3 (59.4–69.0)	62.8 (58.5–71.2)	68.8 (60.9–75.6)	68.5 (64.3–76.1)	67.7 (59.7–73.5)
Follow-up time (months), Mean (SD)^c	30.0 (20.1)	17.2 (13.3)	23.3 (22.8)	28.0 (21.4)	18.2 (17.4)	23.3 (19.8)

Notes: In accordance with the data protection guidelines issued by the Danish Health Data Authority, we do not report cell counts with fewer than 5 observations. As well, the cell count may not be identifiable based on counts in the other cells; for this reason we also report “N<10” etc. in some cells. ^aStage was defined according to the 7th Edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual. ^bSome information on Tumor, Node, or Metastases codes was missing, but based on available information the patient had at least stage II (urothelial cell carcinoma)/III (epithelial ovarian cancer, gastric adenocarcinoma, renal cell carcinoma, non-small cell lung cancer). ^cTime from date of cancer diagnosis as registered in the Danish Cancer Registry until the first of the following events: death, emigration, new primary cancer, or August 1, 2018.

Abbreviations: N, numbers; q1, first quartile; q3, third quartile; SD, standard deviation.

Description of the Reference Standard for Number of Lines of Anti-Neoplastic Therapy

According to the reference standard for number of lines of anti-neoplastic therapy from the medical chart review, 8 patients (4.5%) had no record of any chemotherapy or targeted therapy, 119 patients (66.5%) received first line of anti-neoplastic therapy only, 34 patients (19.0%) received a maximum of two lines of anti-neoplastic therapy, 13 patients (7.3%) received a maximum of three lines of anti-neoplastic therapy, and 5 patients (2.8%) received a maximum of up to four lines of anti-neoplastic therapy before the end of follow-up.

Agreement for Number of Lines of Anti-Neoplastic Therapy

Figure 2 shows the percentage agreement between the reference standard for the number of lines of anti-neoplastic therapy compared with the register-based Time-based algorithm and Drug-based algorithm by the allowed number of days between consecutive administrations of chemotherapy and/or targeted therapy between 28 and 50 days. For the Time-based algorithm, the percentage agreement was greatest when the number of days between consecutive administrations of anti-neoplastic therapy was 44–46 days (67.6%; 95% CI: 60.1–73.8). For all five cancer types, the percentage

agreement with the Time-based algorithm was also greatest around 44–46 days (Figure 3). The lowest percentage agreement with the Time-based algorithm with 44–46 days was for patients with renal cell carcinoma (50.0%; 95% CI: 34.2–65.8) and the highest percentage agreement was for patients with urothelial cell carcinoma (80.6%; 95% CI: 65.6–90.9; Figure 3).

For the Drug-based algorithm, the percentage agreement with the reference standard was 90.5% (95% CI: 85.0–93.7) irrespective of the number of days between consecutive administrations of anti-neoplastic therapy without recording of the specific anti-neoplastic drug (Figure 2). This occurred because all the treated patients selected for the validation study had recorded codes for specific anti-neoplastic drugs in the DNPR on all dates of administration of anti-neoplastic therapy, although we had not selected based on availability of codes for specific anti-neoplastic therapy drugs in the DNPR. The percentage agreement between the Drug-based algorithm and the reference standard was lowest for patients with NSCLC (75.0%; 95% CI: 59.3–86.8) and highest for patients with gastric adenocarcinoma or urothelial cell carcinoma (97.2%; 95% CI: 87.7–99.7; Figure 3).

PPV for Different Number of Anti-Neoplastic Therapy Lines, Start Date of the Lines and Duration of the Lines

We evaluated the PPV for different number of anti-neoplastic therapy lines, start date of the lines, and duration of the lines for the Time-based algorithm and the Drug-based algorithm with 45 days, because the highest percentage agreement for the Time-based algorithm was at 44–46 days.

The PPV for at least one line of anti-neoplastic therapy according to both the Time-based algorithm and the Drug-based algorithm with 45 days was 97.7% (95% CI: 94.7–99.2) in all cohorts combined (Table 2a and b). The PPV declined with increasing number of anti-neoplastic therapy lines, but declined less for the Drug-based algorithm with 45 days (PPV for third anti-neoplastic therapy line, 78.3%; 95% CI: 58.7–91.2; Table 2b) than for the Time-based algorithm with 45 days (PPV for third anti-neoplastic therapy line, 31.8%; 95% CI: 15.5–52.6; Table 2a). The PPV for the second line of anti-neoplastic therapy for the Drug-based algorithm with 45 days was considerably lower for patients with

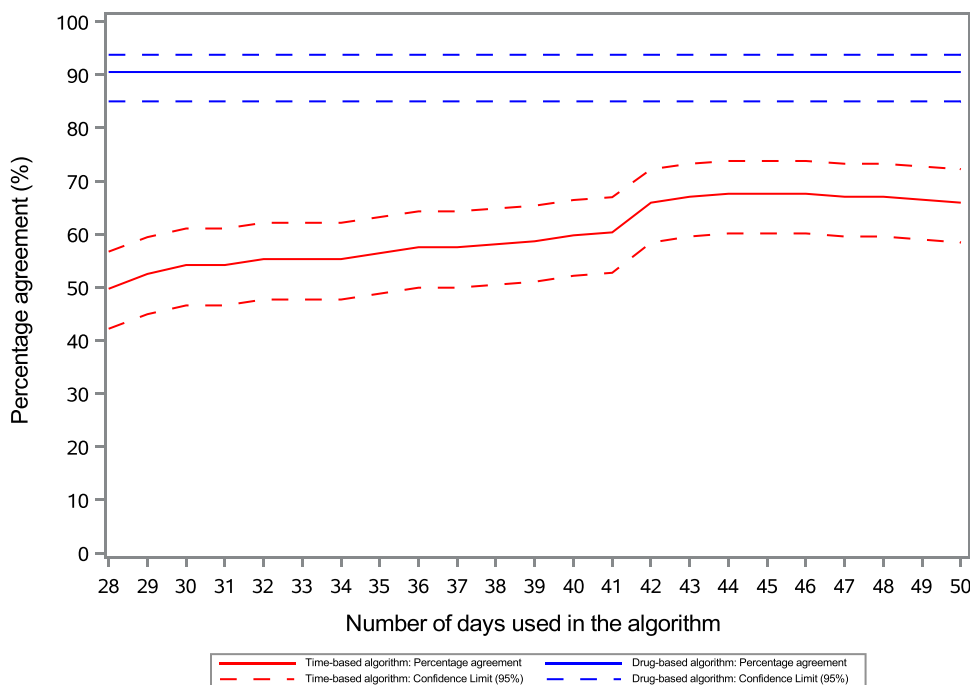


Figure 2 Percentage agreement between the Time-based algorithm and the Drug-based algorithm, respectively, and the reference standard (medical records) for number of lines of anti-neoplastic therapy for all cancer cohorts combined.

Notes: The percentage agreement is the same for the drug-based algorithm irrespective of the number of days used in the algorithm, this occurred because all the treated patients selected for the validation study had recorded codes for specific anti-neoplastic drugs in the DNPR on all dates of administration of anti-neoplastic therapy.

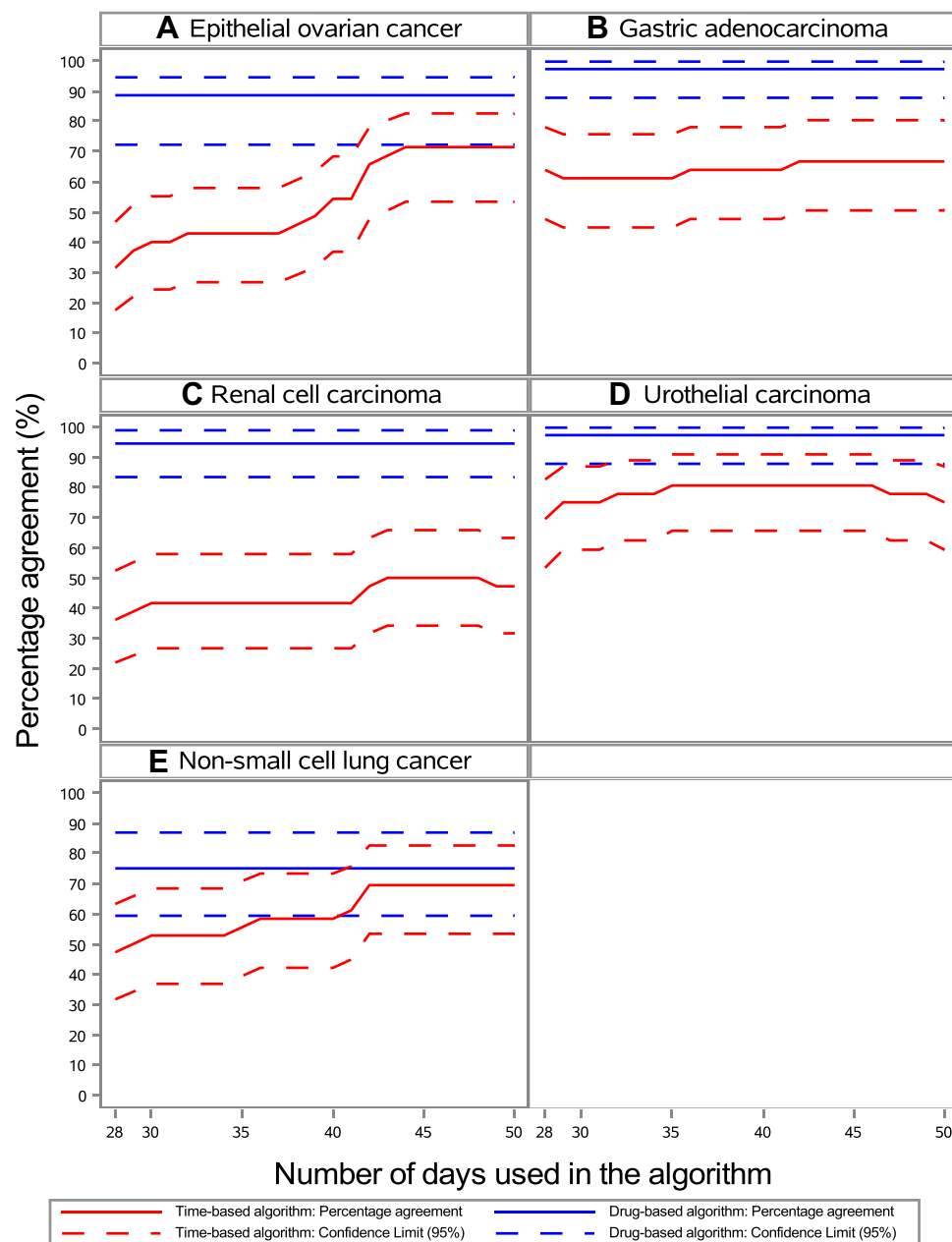


Figure 3 Percentage agreement between the algorithms and the reference standard (medical records) for number of lines of anti-neoplastic therapy for patients with advanced epithelial ovarian cancer (A), gastric adenocarcinoma (B), renal cell carcinoma (C), urothelial cell carcinoma (D), and non-small cell lung cancer (E).

NSCLC (33.3%; 95% CI: 10.4–65.2%) compared with the other cohorts (Table 2b). Likewise, the PPV for the start date of the line and duration of the line depended on the algorithm and the number of lines (Table 3a and b).

Discussion

We found that the algorithm for number of lines of anti-neoplastic therapy based on codes for specific anti-neoplastic drugs from DNPR (Drug-based algorithm) had a higher percentage agreement with the reference standard (from the patient medical records) for number of lines of anti-neoplastic therapy than the algorithm only based on the dates of anti-neoplastic therapy registered in DNPR (Time-based algorithm). The PPV for number of lines of anti-neoplastic therapy, the PPV for the start date of each line, and the PPV for the duration of each anti-neoplastic therapy line diminished with increasing number of lines of anti-neoplastic therapy according to the Drug-based algorithm.

Table 2 Positive Predictive Values for the Number of Anti-Neoplastic Therapy Lines According to Type of Cancer and for All Cohorts Combined for the Time-Based Algorithm 45 Days (a) and the Drug-Based Algorithm 45 Days (b)

	PPV for the Number of Lines (95% CI) ^a					
	Epithelial Ovarian Cancer (N=35)	Gastric Adenocarcinoma (N=36)	Renal Cell Carcinoma (N=36)	Urothelial Cell Carcinoma (N=36)	Non-Small Cell Lung Cancer (N=36)	All Cohorts Combined (N=179)
a) Identify the line number according to Time-based algorithm 45 days						
First line	100.0% (89.7–100.0%)	100.0% (90.3–100.0%)	100.0% (90.3–100.0%)	97.1% (87.1–99.7%)	100.0% (89.7–100.0%)	97.7% (94.7–99.2%)
Second line	78.3% (58.7–91.2%)	71.4% (45.5–89.5%)	35.3% (16.3–58.9%)	57.1% (31.9–79.7%)	78.3% (58.7–91.2%)	57.0% (46.0–67.5%)
Third line	NR	NR	NR	NR	NR	31.8% (15.5–52.6%)
b) Identify the line number according to Drug-based algorithm 45 days						
First line	100.0% (89.7–100.0%)	100.0% (90.3–100.0%)	100.0% (90.3–100.0%)	97.1% (87.1–99.7%)	91.4% (78.9–97.5%)	97.7% (94.7–99.2%)
Second line	100.0% (80.5–100.0%)	92.9% (71.2–99.2%)	100.0% (59.0–100.0%)	90.0% (61.9–98.9%)	33.3% (10.4–65.2%)	86.0% (75.3–93.1%)
Third line	NR	NR	NR	NR	NR	78.3% (58.7–91.2%)

Notes: ^aThe PPV is estimated as N with the given line number both according to the algorithm and the reference standard divided by N with the given line number according to the algorithm.

Abbreviations: N, numbers; NR, not reportable (due to great uncertainty); PPV, positive predictive value.

Table 3 Positive Predictive Values for the Start Date of the Anti-Neoplastic Therapy Lines, and Duration of Anti-Neoplastic Therapy Lines According to the Time-Based Algorithm 45 Days (a) and the Drug-Based Algorithm 45 Days (b) for All Cohorts Combined

	PPV for the Start Date of the Line (95% CI) ^a			Duration in Days of the Line According to the Algorithm, Median, q1-q3 ^b	PPV for the Duration of the Line (95% CI) ^c		
	±30 Days	±60 Days	±90 Days		±30 Days	±60 Days	±90 Days
a) Identify the line number according to Time-based algorithm 45 days							
First line (N=175)	97.1% (93.9–98.9%)	97.1% (93.9–98.9%)	97.1% (93.9–98.9%)	111 (81–157)	65.7% (58.5–72.4%)	71.4% (64.4–77.7%)	76.0% (69.3–81.9%)
Second line (N=79)	36.7% (26.7–47.7%)	38.0% (27.9–49.0%)	41.8% (31.4–52.8%)	101 (66–164)	31.6% (22.2–42.4%)	40.5% (30.2–51.5%)	44.3% (33.7–55.3%)
Third line (N=22)	13.6% (4.0–32.1%)	18.2% (6.5–37.6%)	18.2% (6.5–37.6%)	94 (45–167)	13.6% (4.0–32.1%)	22.7% (9.2–42.9%)	31.8% (15.5–52.6%)
b) Identify the line number according to Drug-based algorithm 45 days							
First line (N=175)	97.1% (93.9–98.9%)	97.1% (93.9–98.9%)	97.1% (93.9–98.9%)	133 (87–202)	86.9% (81.3–91.2%)	94.3% (90.1–97.0%)	95.4% (91.6–97.8%)
Second line (N=57)	80.7% (69.1–89.3%)	82.5% (71.1–90.6%)	82.5% (71.1–90.6%)	90 (66–144)	71.9% (59.4–82.3%)	82.5% (71.1–90.6%)	84.2% (73.2–91.9%)
Third line (N=23)	69.6% (49.3–85.2%)	69.6% (49.3–85.2%)	73.9% (53.9–88.3%)	71 (45–140)	60.9% (40.6–78.6%)	69.6% (49.3–85.2%)	69.6% (49.3–85.2%)

Notes: ^aThe PPV is estimated as N who have started the given anti-neoplastic therapy line number according to the algorithm and who have also started the given anti-neoplastic therapy line number according to the reference standard and for whom the start date of the given anti-neoplastic therapy line number according to the reference standard is within plus or minus 30/60/90 days from the start date according to the algorithm divided by N who have started the given anti-neoplastic therapy line number according to the algorithm. ^bDuration is estimated as the last date of that line minus the start date of that line. ^cThe PPV is estimated as N who have started the given anti-neoplastic therapy line number according to the algorithm and who have also started the given anti-neoplastic therapy line number according to the reference standard and for whom the duration of the given anti-neoplastic therapy line number according to the reference standard is within plus or minus 30/60/90 days from the duration according to the algorithm divided by N who have started the given anti-neoplastic therapy line number according to the algorithm.

Abbreviations: N, numbers; q1, first quartile; q3, third quartile; PPV, positive predictive value.

The present study relied on information from the medical records as the reference standard. In general, the records did not include information about the line of anti-neoplastic therapy, so we had to define lines of anti-neoplastic therapy based on the recorded anti-neoplastic therapy drugs. It is possible that the information in the medical records was incomplete – for example, the administration of anti-neoplastic therapy may not have been recorded. In addition, errors might have occurred in entering information from the medical record into REDCap. Such errors could explain why 8 patients did not have any registrations of anti-neoplastic therapy in the medical records, although they had been selected based on registrations of anti-neoplastic therapy in the DNPR. Another explanation for no registration of anti-neoplastic therapy could be that the selection of patients did not censor at the date of new primary cancer, therefore the registration of anti-neoplastic therapy in DNPR might have occurred after the diagnosis of a new primary cancer. Overall, it is a limitation that the reference standard cannot be assumed to hold perfect information about lines of anti-neoplastic therapy. However, we believe it was the most valid available information on lines of anti-neoplastic therapy.

The sample size for this study was largely decided based on the resources available for data extraction from the medical records. The precision of the estimates would increase if we had been able to include additional patients, which could have been particularly valuable when splitting the analyses on cancer type for the PPV related to second or third line of anti-neoplastic therapy.

We only included patients treated at the Department of Oncology at Aarhus University Hospital and therefore our results may not be directly transferable to other hospitals, which may code anti-neoplastic therapy differently. A previous study validating treatment codes for colorectal cancer patients in the DNPR against the reference standard from medical records at all hospitals in the Region of Southern Denmark, found that the PPV for all anti-neoplastic procedures was slightly higher at Odense University Hospital (96%; 95% CI: 95–96) compared with the regional hospitals (the lowest PPV was 84% [95% CI: 76–91] at South-West Jutland hospital).²¹ Nonetheless, cancer care is delivered according to national Danish Cancer Patient Pathways,²² which should ensure uniform treatment across Denmark. It has been mandatory to report cancer treatment to DNPR since 2001.¹²

We found that the PPV of at least one line of anti-neoplastic therapy according to DNPR was 97.7%, which corresponded to the PPV for initiating anti-neoplastic therapy. This is in line with two previous studies of colorectal cancer patients, reporting PPVs between 94% and 100% for registration of anti-neoplastic treatment.^{21,23}

In the present study, we examined alternative algorithms for the number of lines of anti-neoplastic therapy based on data in DNPR. In future studies, we would recommend the Drug-based algorithm given the high percentage agreement with the reference standard. In the present study, it was not possible to assess the most valid number of days between dates of administration of anti-neoplastic therapy in the Drug-based algorithm, because all patients had their specific treatments registered in the DNPR. Based on our Time-based algorithm, we recommend using 45 days in the final Drug-based algorithm. We have created and made available three SAS syntax files, which can be used by other researchers to define the number of lines of anti-neoplastic therapy according to the Drug-based algorithm with 45 days ([Appendices 4–6](#)). Use of the syntax files, including modifying these for use in non-Danish settings, is described in [Appendix 7](#). It is also possible to modify the algorithm to include other types of systemic anti-cancer treatment and describe the type of drugs used in each line. For instance, Sørup et al modified the Drug-based algorithm to also include immunotherapy and described the types of drugs used in different lines of treatment.²⁴ However, if feasible, we recommend validating the algorithm in the intended study population or at least consider carefully the limitations and recommendations described in the following paragraphs and in [Appendix 7](#).

The percentage agreement between the Drug-based algorithm and the reference standard varied by cancer type. The validity of the algorithm may differ for other types and stages of cancer than those included in the present study, which only included patients diagnosed with incident advanced cancer – this implies that anti-neoplastic therapy was given with palliative intent. However, no information on curative or palliative intent of anti-neoplastic therapy is registered in the DNPR. It is important to be aware that our algorithm defines maintenance therapy with a different type of drug as a new therapy line; this might overestimate the number of lines of anti-neoplastic therapy.

The use of specific treatment codes for anti-neoplastic therapy might vary over the years. The aforementioned study of patients with colorectal cancer diagnosed between 2009 and 2010, found that a low proportion of patients treated at Aarhus University Hospital had specific codes registered for chemotherapy in the DNPR,²³ while all patients included in

the present study had all registrations of anti-neoplastic therapy with specific codes in the DNPR (diagnosed 2012 to 2016). Another study showed that only one hospital out of four hospitals sometimes used an unspecific code for anti-neoplastic therapy, while the other hospitals always used specific anti-neoplastic treatment codes for colorectal cancer patients treated May 2016 to April 2018.²¹ Furthermore, recently introduced anti-neoplastic therapy might not have specific treatment codes in the first years after introduction.²⁵

In future studies, we recommend examining the proportion of the population who have all registrations of anti-neoplastic therapy with treatment codes for specific drugs. It is important to note that the PPV for the second line of anti-neoplastic therapy was 86% for patients with all registration of anti-neoplastic therapy in the DNPR with specific drugs (Drug-based algorithm), while the PPV was substantially lower when using the Time-based algorithm (57%). Therefore, in a study only including patients on second line of anti-neoplastic therapy, it might be relevant to restrict the study to patients who have all registrations of anti-neoplastic therapy with treatment codes for specific drugs. Overall, it is important to note that the PPVs for number of lines of anti-neoplastic therapy, the start date of each line, and the duration of each anti-neoplastic therapy line all diminished with increasing number of lines of anti-neoplastic therapy. This may be related to discrepancies between dates in the DNPR and medical record for the first line therapy, which would in turn influence the date of the second line, and subsequent therapy lines. Furthermore, it is likely more difficult to catch changes in anti-neoplastic therapy than initiation of anti-neoplastic treatment.

The algorithm and conclusions cannot be applied directly to other countries because they are based on the national Danish treatment codes. However, the results indicate that if only dates of anti-neoplastic therapy are available, the validity of an algorithm for number of lines of anti-neoplastic therapy might be low, while the validity might be higher if there is information on the specific drugs administered at each date. Nonetheless, it is important that all administrations of anti-neoplastic therapy are correctly registered.

Algorithms have been developed for number of lines of therapy based on registers or electronic databases in other settings. One study from the USA based on SEER-Medicare data developed an algorithm for lines of chemotherapy and biologic treatment for metastatic colon cancer patients.⁹ A French study defined an algorithm for treatment lines for multiple myeloma patients based on drugs administered and days between administrations.⁸ Neither of these studies validated the algorithms.^{8,9} A study of breast, lung and colorectal cancer patients from Kaiser Permanente, Colorado, USA, found a high agreement in the number of lines of therapy between the reference standard and an algorithm based on drug names, dates of dispensing, and manual adaptation to the general rules based on known treatment plans and removal of non-antineoplastic drugs.⁷ A Canadian study validated a supervised learning approach to identify the best performing proxy for number of lines of systemic therapy using data from pharmacy records on drug names, protocol codes, and dates of prescription.⁶ The approach worked well and the best proxy differed between the different types of cancers (chronic lymphocytic leukemia, follicular lymphoma, diffuse large B-cell lymphoma, and advanced cancers of varying histology).⁶

Conclusion

Our findings show a higher validity for the register-based algorithm for number of lines of anti-neoplastic therapy considering the specific type of anti-neoplastic therapy compared with an algorithm only considering the dates of anti-neoplastic therapy across different cancer types. In future register-based Danish studies using information on anti-neoplastic therapy lines, we recommend using the Drug-based algorithm with 45 days to identify the number of lines of anti-neoplastic therapy and associated start date and duration of each line. Preferably, the algorithm should be validated for the specific population under study, as the validity may vary according to study period, type of cancer, and introduction of new drug types. If a separate validation is not feasible, it is important to consider the proportion of the study population who have all registrations of anti-neoplastic therapy in the DNPR with specific drugs and potentially restrict the study to this subpopulation.

Acknowledgments

We would like to thank Johan Frederik Håkonsen Arendt, MD, PhD for contribution to obtain the data for the study, project nurse Henriette Kristoffersen for her invaluable help in retrieval of medical records and abstraction of data

from the medical records, and statistician Thomas Bøjer Rasmussen for review of the algorithms for number of lines of anti-neoplastic therapy. This work was partly supported by Merck (CrossRef Funder ID: 10.13039/100009945) by an institutional grant to and administered by Aarhus University Hospital, as part of an alliance between Merck and Pfizer.

Disclosure

SS, BD, JSK, ASR, CFH, SVV, MR, and DCF report no personal conflicts of interest in this work. AAK reports honoraria from AstraZeneca A/S for speaking at a Nordic-Baltic scientific meeting, “1ST Nordic/Baltic multidisciplinary scientific exchange meeting on treatment of inoperable stage III NSCLC patients” (2019). LR is an employee of Pfizer and DO, EB, and PV are employees of Merck Healthcare KGaA, Darmstadt, Germany. The authors report no other conflicts of interest in this work.

References

1. Explanation of first-line and second-line chemotherapy regimens; 2021. Available from: <https://callaix.com/firstline>. Accessed September 2, 2021.
2. Reck M, Popat S, Reinmuth N, et al. Metastatic non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2014;25 Suppl 3:iii27–iii39. doi:10.1093/annonc/mdu199
3. Escudier B, Porta C, Schmidinger M, et al. Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27(suppl 5):v58–v68. doi:10.1093/annonc/mdw328
4. Davies J, Patel M, Gridelli C, de Marinis F, Waterkamp D, McCusker ME. Real-world treatment patterns for patients receiving second-line and third-line treatment for advanced non-small cell lung cancer: a systematic review of recently published studies. *PLoS One*. 2017;12(4):e0175679. doi:10.1371/journal.pone.0175679
5. Sacher AG, Le LW, Lau A, Earle CC, Leighl NB. Real-world chemotherapy treatment patterns in metastatic non-small cell lung cancer: are patients undertreated? *Cancer*. 2015;121(15):2562–2569. doi:10.1002/ncr.29386
6. Weymann D, Costa S, Regier DA. Validation of a cyclic algorithm to proxy number of lines of systemic cancer therapy using administrative data. *JCO Clin Cancer Inform*. 2019;3:1–10. doi:10.1200/CCI.19.00022
7. Carroll NM, Burniece KM, Holzman J, McQuillan DB, Plata A, Ritzwoller DP. Algorithm to identify systemic cancer therapy treatment using structured electronic data. *JCO Clin Cancer Inform*. 2017;1:1–9. doi:10.1200/CCI.17.00002
8. Palmaro A, Gauthier M, Despas F, Lapeyre-Mestre M. Identifying cancer drug regimens in French health insurance database: an application in multiple myeloma patients. *Pharmacoepidemiol Drug Saf*. 2017;26(12):1492–1499. doi:10.1002/pds.4266
9. Bikov KA, Mullins CD, Seal B, Onukwugha E, Hanna N. Algorithm for identifying chemotherapy/biological regimens for metastatic colon cancer in SEER-medicare. *Med Care*. 2015;53(8):E58–E64. doi:10.1097/MLR.0b013e31828fad9f
10. Di Lorenzo G, Buonerba C, Bellelli T, et al. Third-line chemotherapy for metastatic urothelial cancer: a Retrospective Observational Study. *Medicine*. 2015;94(51):e2297. doi:10.1097/MD.0000000000002297
11. Schonemann KR, Mortensen MB, Bjerregaard JK, Fristrup C, Pfeiffer P. Characteristics, therapy and outcome in an unselected and prospectively registered cohort of patients with gastro-oesophageal cancer. *Acta Oncol*. 2014;53(3):385–391. doi:10.3109/0284186X.2013.820839
12. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
13. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol*. 2019;11:563–591. doi:10.2147/CLEP.S179083
14. Gjerstorff ML. The Danish cancer registry. *Scand J Public Health*. 2011;39(7 Suppl):42–45. doi:10.1177/1403494810393562
15. Schmidt M, Pedersen L, Sorensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
16. Mainz J, Hess MH, Johnsen SP. The Danish unique personal identifier and the Danish civil registration system as a tool for research and quality improvement. *Int J Qual Health Care*. 2019;31(9):717–720. doi:10.1093/intqhc/mzz008
17. Compton CC, Byrd DR, Garcia-Aguilar J, Kurtzman SH, Olawoye A, Washington MK. *AJCC Cancer Staging Atlas*. New York: Springer; 2012.
18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010
19. Ho GY, Woodward N, Coward JI. Cisplatin versus carboplatin: comparative review of therapeutic management in solid malignancies. *Crit Rev Oncol Hematol*. 2016;102:37–46. doi:10.1016/j.critrevonc.2016.03.014
20. Brown LD, Cai TT, DasGupta A. Confidence intervals for a binomial proportion and asymptotic expansions. *Ann Stat*. 2002;30(1):160–201. doi:10.1214/aos/1015362189
21. Broe MO, Jensen PB, Mattsson TO, Pottegård A. Validity of antineoplastic procedure codes in the Danish national patient registry: the case of colorectal cancer. *Epidemiology*. 2020;31(4):599–603. doi:10.1097/EDE.0000000000001208
22. Probst HB, Hussain ZB, Andersen O. Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians—a national Danish project. *Health Policy*. 2012;105(1):65–70. doi:10.1016/j.healthpol.2011.11.001
23. Lund JL, Froslev T, Deleuran T, et al. Validity of the Danish national registry of patients for chemotherapy reporting among colorectal cancer patients is high. *Clin Epidemiol*. 2013;5:327–334. doi:10.2147/CLEP.S49773
24. Sørup S, Darvalics B, Khalil AA, et al. Treatment and survival in advanced non-small cell lung cancer, urothelial, ovarian, gastric and kidney cancer: a nationwide comprehensive evaluation. *Clin Epidemiol*. 2021;13:871–882. doi:10.2147/CLEP.S326470
25. Vesteghem C, Brøndum RF, Falkmer UG, Pottegård A, Poulsen LØ, Bøgsted M. High validity of the Danish national patient registry for systemic anticancer treatment registration from 2009 to 2019. *Clin Epidemiol*. 2021;13:1085–1094. doi:10.2147/CLEP.S332776

Clinical Epidemiology

Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>