Text mining real-world data to evaluate systemic anti-cancer therapy

Laar, S.A. van

Chapter 3

An electronic health record text mining tool to collect real-world drug treatment outcomes: A validation study in metastatic renal cell carcinoma patients

Sylvia A. van Laar, Kim B. Gombert-Handoko,
Henk-Jan Guchelaar, Juliëtte Zwaveling

Clin Pharmacol Ther. 2020; 108(3): 644-652

Abstract

Introduction: Real-world evidence can close the inferential gap between marketing authorization studies and clinical practice. However, the current standard for real-world data extraction from electronic health records (EHR) for treatment evaluation is manual review (MR), which is time-consuming and laborious. Clinical Data Collector (CDC) is a novel natural language processing and text-mining software tool for both structured and unstructured EHR data and only shows relevant EHR sections improving efficiency.

Methods: We investigated CDC as a RWD collection method, through application of CDC queries for patient inclusion and information extraction on a cohort of metastatic renal cell carcinoma (mRCC) patients receiving systemic drug treatment. Baseline patient characteristics, disease characteristics, and treatment outcomes were extracted and these were compared to manual review for validation.

Results: 100 patients receiving 175 treatments were included using CDC, which corresponded to 99% with manual review. Calculated median overall survival was 21.7 months (95% CI 18.7–24.8) versus 21.7 months (95% CI 18.6–24.8) and progression-free survival 8.9 months (95% CI 5.4–12.4) versus 7.6 months (95% CI 5.7–9.4) for CDC versus MR respectively. Highest F1-score was found for cancer-related variables (88.1–100), followed by comorbidities (71.5–90.4) and adverse drug events (53.3–74.5), with most diverse scores on international mRCC database criteria (51.4–100). Mean data collection time was 12 minutes (CDC) versus 86 minutes (MR).

Conclusion: In conclusion, CDC is a promising tool for retrieving RWD from EHRs since the correct patient population can be identified as well as relevant outcome data as overall survival and progression-free survival.

1. Introduction

Randomized controlled trials (RCTs) are the gold standard to investigate efficacy of novel drug therapies and therefore RCTs are pivotal for drug marketing authorization applications [1-3]. However, in the accelerated approval pathway of the US Food And Drug Administration (FDA) and in the conditional marketing approval pathway of the European Medicines Agency (EMA), new and mostly expensive, anticancer drugs are increasingly approved based upon studies with surrogate end-points such as progression-free survival (PFS) or objective response rate (ORR), and a large part of these studies lack a standard-of-care control arm [4]. Consequently, the treatment effect in terms of overall survival (OS) is unclear at approval by the authorities. In addition, novel drugs are usually investigated in a highly selected patient population which may not be representative for the full cohort of patients who will receive the treatment in clinical practice [5]. This inferential gap between evidence from RCTs and clinical practice can be closed by the use of real-world data (RWD) as complementary information [1, 6-10]. These RWD may differ from outcome data from RCTs and may be valuable in assessing the effectiveness of a new drug in daily practice, for example, in patients with specific characteristics such as older patients or in patients with comorbidities.

An important source for RWD is the electronic health record (EHR) [6, 9, 11]. It contains individual longitudinal patient data collected during routine clinical practice and includes information about patients' demographics, health behavior, vital signs, encounters, laboratory data, medication orders, procedures, imaging, health problem lists, and free-text notes [12]. These free-text notes, in particular, contain very detailed and nuanced information about patients, their illnesses and treatment trajectory including efficacy and side effects of drug treatment. However, since these free-text notes are unstructured, they are less suitable for automated information extraction [13, 14]. Therefore, manual chart review is still the standard method for data collection from EHRs [12]. Unfortunately, this manual method is laborious, time-consuming and error-prone [12, 13, 15], and thus, not a durable approach for the structural collection of RWD from EHRs. Therefore, more advanced methods are highly warranted.

Natural language processing (NLP) and text-mining techniques are advanced methods of information extraction of free-text data [13, 14]. Although these methods are promising, they are not yet easy applicable as an alternative method to evaluate the

effectiveness of treatments in daily practice. Currently, these techniques are mostly used by a few health care institutions with strong informatics departments, where knowledge of informaticians can be combined with knowledge of clinicians [16]. For example, an NLP pipeline to extract urinary incontinence and erectile dysfunction was developed for patient-centered outcomes of prostate cancer treatment [17]. Additionally, a method combining NLP and machine learning techniques was developed by Sohn et al. to collect adverse drug events from psychiatry and psychology medical records [18]. Similar studies were performed for drug-named entity recognition, dosage information, and drug exposure extraction and all these studies were limited to one type of outcome [16].

The Clinical Data Collector (CTcue B.V., Amsterdam, the Netherlands) is an NLP and text mining-based tool, which is built to collect structured as well as unstructured data from EHRs and is currently available in hospitals in the Netherlands and Belgium. In contrast to other tools, CDC is designed for medical and pharmaceutical professionals to easily build queries themselves for information extraction on their topic of interest. Using these queries, only relevant parts of the EHRs are shown and results are directly collected into a table, thereby potentially improving the efficiency of retrieval of patient data [19].

CDC may be a useful extraction tool for retrieving RWD from EHRs. Therefore, we designed a validation study to assess the information extraction of clinical trial parameters from the EHR by CDC with customized queries. Since we are interested in the effectiveness data of specific oncological drug treatments we choose to perform this study in patients with metastatic renal cell carcinoma (mRCC) receiving systemic treatment.

2. Methods

In this observational, retrospective validation study, Clinical Data Collector (CDC) was applied to collect patient characteristics, treatment outcomes and ADEs during drug treatments for mRCC from EHRs. These data were compared to manually obtained data from the EHR. Patient inclusion, patient characteristics, treatment outcomes, ADEs and data collection time per patient were evaluated. The study was reviewed by the Medical Ethics Review Committee of the Leiden University Medical Center, who determined that the Medical Research Involving Human Subjects Act (WMO) was not applicable to this study.

2.1 Study population

Patients, 18 years and older, with metastatic renal cell carcinoma (mRCC) who received drug treatment with cabozantinib, pazopanib, sunitinib, everolimus, or nivolumab were included in the study. Patients underwent drug treatment between January 2015 until May 2019 in the Leiden University Medical Center, the Netherlands.

2.2 Collected variables

Variables that are generally presented in RCTs evaluating new drug therapies in mRCC were collected [20-22], namely general patient related characteristics (sex, age, length, weight, eGFR, ALAT and ASAT) and disease related characteristics (histological RCC subtype and prior nephrectomy) at baseline, including also four common comorbidities (hypertension, cardiovascular comorbidities, diabetes mellitus and COPD) and the International Metastatic Renal cell carcinoma Database Consortium (IMDC) criteria to predict prognostic categories (hypercalcemia, neutrophilia and thrombocytosis, anemia, performance status below 80% Karnofsky and time from diagnosis to systemic drug treatment below 1 year). Furthermore, treatment outcomes were collected, including tumor progression and overall survival since start of treatment and four common ADEs (hand-foot syndrome, liver toxicity, diarrhea and hypertension).

2.3 Manual reference

To create a gold standard, manual chart review was performed by a pharmacist, who is experienced in working with the EHR both as healthcare professional and as reviewer. Data were collected from the EHR (HiX, Chipsoft B.V., Amsterdam, the Netherlands), which has no build-in term search, in an electronic case report form (eCRF) (Castor EDC, Amsterdam, the Netherlands). For each patient the time to collect data manually in the eCRF was recorded.

2.4 Clinical Data Collector

CDC is a software tool which is linked with the EHR in the hospital. EHR data are transformed by an application programming interface (API), to enable structured search using the search engine by medical professionals (user). Figure 3.1 shows the communication lines between on-premises isolation platform.

Both the patient population and data points can be defined using CDC queries. Structured data can be extracted from the EHR with specified queries per datatype

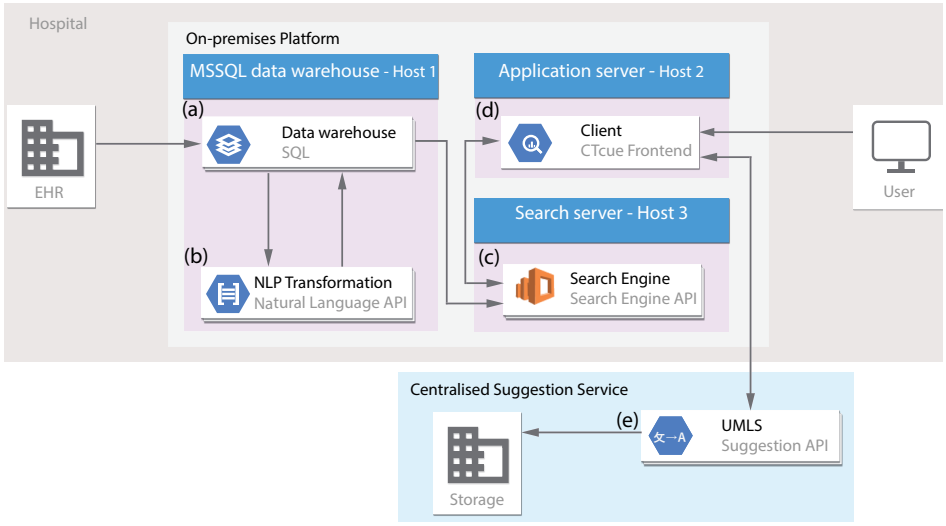

Figure 3.1. Architecture of the Clinical Data Collector on-premises isolation platform. **(a)** Copy of electronic health record (EHR) data transferred, stored, and cleaned in a local MS SQL Server relational database. **(b)** Natural language processing (NLP) transformation application programming interface (API) pseudonymizes data. **(c)** Search engine is compatible with the structure used in data warehouse. **(d)** Client to build queries by a user. Results window in CDC shows only parts of EHR documents containing defined criteria by user. **(e)** Text mining of (combinations of) keywords is supported by an online thesaurus.

(e.g., medication requests and lab results). Additionally, information extraction from unstructured text is enabled through text mining based on keywords. After running the designed queries, all data are combined in a generic dataset. When a data point is selected, the EHR context is shown, which enables the user to manually validate results. The handling of structured and unstructured EHR data is shown in Figure 3.2. The results can be exported into a CSV-file or XLSX-file.

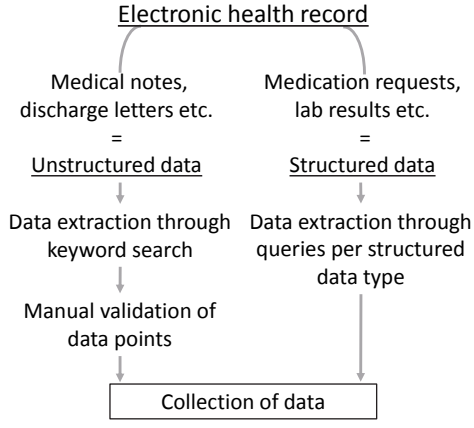

Figure 3.2. Data extraction approach from structured and unstructured data using Clinical Data Collector.

Queries for patient inclusion and data collection were defined as follows. Patients were included only in CDC for data extraction with both a Diagnosis Treatment Combination (DTC)-code for kidney tumors as well as an initial prescription of at least one of the five drug treatments. A DTC is a code used for hospital costs reimbursement in the Netherlands [23]. As both variables were stored as structured data, corresponding structured data queries were applied. The remaining structured data (e.g., medication requests and lab results) were extracted using these queries as well. For example, the last known measurement result before the start of drug treatment could be automatically selected through linkage with the treatment initiation date. Additionally, queries enabling keyword search were used to select relevant parts of unstructured text in EHRs only. A combination of keywords resulting from the suggestion application programming interface (API), common known synonyms, variants, abbreviations, and typing errors were manually set for this free-text search. Also, combinations of queries to select structurally stored data and free-text search queries were used to improve recall of some variables. Three query examples are shown in Supplementary File S3.1. The completeness of the queries was assessed inspecting the test results section in CDC of 10 random patients and a set of test results was compared to a test set of manual results, before finalizing the queries.

After applying patient inclusion criteria using CDC, preselected patients were screened for final inclusion. Subsequently, for data extraction, all variables fully based on structured data were automatically extracted. Variables fully or partially based on unstructured data were manually verified before extraction, using the selected parts of the EHR shown in the results display of CDC, resulting in a semi-automatic extraction procedure. Patient screening and data validation was performed by the same pharmacist that performed the manual review. The time spent on final patient inclusion and verification of data was measured for CDC. This was compared to the time that was spent per patient task for manual chart review.

2.5 Analysis & statistics

To establish accuracy of data retrieval, results were compared with manual review. For categorical patient characteristics and ADEs, precision, recall and F1-scores were calculated. There is no consensus on thresholds for accuracy scores that an information extraction tools should meet. However, we set thresholds for both precision and recall at 90%, to limit the chance on incorrect conclusions when data is used for treatment

evaluation. This is in line with thresholds set by Hernandez-Boussard et al. [24]. Since a part of the IMDC-criteria are measurement values, with the answer being a binary question, these will also be analyzed by calculating precision, recall and F1-score.

$$Precision = \frac{True\ positives}{True\ positives + false\ positives}$$

$$Recall = \frac{True\ positives}{True\ positives + false\ negatives}$$

$$F_1 - score = 2 * \frac{Precision * recall}{Precision + recall}$$

Next, for all continuous patient characteristics, Bland-Altman plots were composed, to describe agreement between CDC and MR. Per patient the difference in extracted value was plotted against the mean value of both methods for this patient. Also, mean differences between data collected using CDC and manual review were determined. Kaplan-Meier plots for PFS and OS were composed for all treatments combined. Data were combined since the aim of our study was to validate whether CDC PFS and OS results are equivalent to manual review. For PFS, time from start treatment until significant tumor progression during treatment according to RECIST 1.1 [25] was used or death from any cause. Patients were censored when treatment ended without tumor progression or when patients were still on treatment at the end of inclusion. Furthermore, for OS, time from start treatment until death from any cause was calculated. Patients were censored when alive at the end of the inclusion period. Since the included patients could have received multiple lines of treatments, patients could occur multiple times in both plots. Statistical analysis was performed in SPSS version 25 (IMB corp., Armonk, NY, USA).

3. Results

3.1 Patient inclusion

First, we investigated whether CDC was able to trace all patients who met the inclusion criteria. Using inclusion queries in CDC, 133 patients were initially selected based on treatment use and DTC-code. Of these, 33 patients were excluded, which resulted in 100 patients included by CDC. For manual review, 119 patients were initially selected

based on drug prescriptions of cabozantinib, everolimus, nivolumab, pazopanib and sunitinib in the EHR. These drug treatments represent several treatment lines for mRCC. Of these, 19 patients were excluded, and therefore 100 patients were included in the manual dataset. Most of the patients who were excluded in both methods were selected based on a new prescription of follow-up treatment, however they did not initiate the treatment in defined inclusion period.

A total of 99 out of 100 patients selected using CDC corresponded with the patients included manually. This difference was caused by an incorrect registered DTC-code in the EHR of an unselected patient. Figure 3.3 shows the complete patient inclusion flowchart.

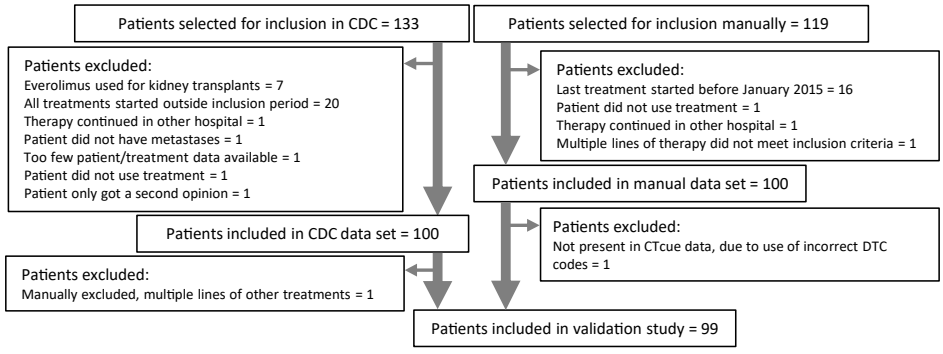

Figure 3.3. Flowchart of patient inclusion.

3.2 Information extraction

Validation parameters were collected and accuracy scores were calculated in order to qualify the usefulness of CDC with respect to manual retrieved outcome data. Table 3.1 presents an overview of the collected variables per drug treatment for both methods. First, both manual review and CDC identified 175 treatments, of which 174 were identical. The two differences in treatments were due to prescribing errors. One patient did not start treatment with nivolumab according to free-text documentation, which was manually recorded, while documented in the structured medication overview and therefore extracted by CDC, and vice versa for a treatment of sunitinib. Clear cell RCC was the most frequently reported histological subtype, with 152 patients by manual review (MR) and 151 patients by CDC. 14 patients were manually identified as rarer subtypes, while 9 remained unclear. CDC reported 6 and 18 patients respectively. The reported values of other cancer-related variables were also similar. The most reported ADE by manual review was liver toxicity (n=69), however diarrhea was mostly reported

by CDC (n=51). Hand-foot syndrome was the least reported by both methods (MR: n=26; CDC: n=19). Furthermore, the number of reported ADEs showed the largest difference between both data retrieval methods for liver toxicity (MR: n=69; CDC=39) and the smallest for hand-foot syndrome (MR: n=26; CDC=19). Further, of all IMDC score parameters, used to determine the mRCC prognosis, the incidence of anemia was far the most reported by both methods (MR: n=103; CDC: n=105). The reported incidence is quite similar between methods for anemia and thrombocytosis (absolute difference of resp. 0.4% and 0.6%). Though, an absolute difference of 22% was shown in reported patients which received systemic treatment within a year after diagnosis. A substantial amount of missing data was reported on the IMDC-criteria calcium (MR: n=9; CDC: n=9), neutrophil (MR: n=22; CDC: n=13), and performance status (MR: n=19; CDC: n=64). Finally, the means of all continuous variables were similar. Values for age (years), length (cm), weight (kg), ALAT (U/L), and ASAT (U/L) all differed less than one measurement unit. The reported means for eGFR showed a difference of 2.9 ml/min/1.73m^2. Moreover, for all variables some missing data was found, however, length (CDC: n=6), weight (MR: n=11; CDC: n=27), and kidney function (CDC: n=20) were most prominent.

Table 3.1. Collected variables for each treatment per method

	Manual review (n=175)[a]	Clinical Data Collector (n=175)[a]
Drug treatment		
Cabozantinib, n (%)	27 (15.4)	27 (15.4)
Everolimus, n (%)	17 (9.7)	17 (9.7)
Nivolumab, n (%)	40 (22.9)	41 (23.4)
Pazopanib, n (%)	70 (40.0)	70 (40.0)
Sunitinib, n (%)	21 (12.0)	20 (11.4)
Male, n (%)	128 (72.7)	129 (73.3)
Cancer-related variables		
Histological subtype of renal cell carcinoma		
Clear cell (%)	152 (86.9)	151 (86.3)
Papillary, n (%)	7 (4.0)	3 (1.7)
Sarcomatoid, n (%)	3 (1.7)	3 (1.7)
Mixed, n (%)	4 (2.3)	0 (0)
Unclear, n (%)	9 (5.1)	18 (10.3)
Prior nephrectomy, n (%)	114 (65.1)	117 (66.9)
Progression on treatment, n (%)	101 (57.7)	98 (56.0)
Death since start treatment, n (%)	99 (56.7)	99 (56.7)

Table 3.1 continues on next page.

Table 3.1. *Continued*

	Manual review (n=175)[a]	Clinical Data Collector (n=175)[a]
Comorbidities		
Hypertension, n (%)	91 (52.3)	114 (65.1)
Cardiovascular comorbidities, n (%)	43 (24.6)	27 (15.4)
Diabetes Mellitus, n (%)	39 (22.3)	34 (19.4)
COPD, n (%)	12 (6.9, n=172)	15 (8.6)
Adverse drug events		
Hand-foot syndrome, n (%)	26 (14.8)	19 (10.8)
Liver toxicity, n (%)	69 (39.2)	39 (22.2)
Diarrhea, n (%)	43 (24.4)	51 (29.0)
Hypertension, n (%)	64 (36.4)	46 (26.1)
IMDC score parameters		
Hypercalcemia, n (%)	28 (16.9, n=166)	24 (14.5, n=166)
Anemia, n (%)	103 (59.2, n=174)	105 (60.0, n=175)
Neutrophilia, n (%)	32 (20.9, n=153)	40 (24.9, n=162)
Thrombocytosis, n (%)	23 (13.4, n=172)	22 (12.8, n=172)
Performance status <80% Karnofsky, n (%)	28 (17.9, n=156)	13 (11.7, n=111)
Time from diagnosis to systemic therapy <1 year, n (%)	89 (50.9)	49 (28.0)
Continuous variables		
Age, years, mean	65.0	65.2
Length, cm, mean	176.2 (n=173)	176.6 (n=169)
Weight, kg, mean	80.6 (n=164)	81.2 (n=148)
ALAT, U/L, median	21 (n=173)	21 (n=174)
ASAT, U/L, median	22 (n=172)	22 (n=174)
eGFR, ml/min/1.73m^2, mean	64.9 (n=174)	62.0 (n=155)

COPD: Chronic obstructive pulmonary disease; eGFR: estimated glomerular filtration rate; ALAT: alanine transaminase; ASAT: aspartate aminotransferase; IMDC: international metastatic renal cell carcinoma database consortium.
[a] In case of missing data, number of known variables is presented.

To assess the quality of data extraction of categorical variables by CDC, the precision, recall, and F1-scores, summarizing both precision and recall, were calculated and presented in Table 3.2. In general, the highest scores on data retrieval were established in cancer-related variables and lowest in ADEs. Besides, results for IMDC-criteria were most diverse with higher scores for continuous structured variables. The highest score for precision of 100% was obtained for sex and platelet levels above normal, and the lowest precision of 39.1% was obtained for performance status. Similar, the highest recall of 100% was reached for sex, platelet levels, and cardiovascular disease and the lowest score of 63.2% was obtained for hand-foot syndrome.

Table 3.2. Performance scores on collection of categorical variables

	Precision (%)	Recall (%)	F1-score (%)
Sex	100*	100*	100*
Cancer related variables			
Death since start treatment	100*	100*	100*
Prior nephrectomy	96.5*	94.0*	95.2*
Progression during treatment	93.1*	96.0*	94.5*
Histological subtype of renal cell carcinoma	89.3	88.5	88.1
Comorbidities			
Diabetes Mellitus	84.6	97.1*	90.4
COPD	91.6*	73.3	81.1
Cardiovascular comorbidities	62.8	100*	77.1
Hypertension	80.2	64.6	71.5
Adverse drug events			
Diarrhea	81.4	68.6	74.5
Liver toxicity	49.3	87.2	63.0
Hypertension	51.6	71.7	60.0
Hand-foot syndrome	46.2	63.2	53.3
IMDC-criteria			
Thrombocytosis	100*	100*	100*
Anemia	99.0*	98.1*	98.6*
Hypercalcemia	80.1	91.3*	85.7
Neutrophilia	90.0*	72.9	80.5
<1 year from diagnosis to systematic treatment	53.9	98.0*	69.6
Karnofsky performance status <80%	39.1	75.0	51.4

* Meet the set threshold for accuracy of 90%.
COPD: chronic obstructive pulmonary disease; IMDC: International Metastatic renal cell carcinoma Database Consortium.

Outcome parameters were validated by determining progression-free survival and overall survival. Progression during treatment could be predicted with a precision of 93.1% and recall of 96% by CDC (Table 3.2). In addition, calculated median PFS was 8.90 months (95% CI 5.38–12.43) versus 7.59 months (95% CI 5.74–9.44) for CDC versus manual review, respectively (Figure 3.4A), which was not significantly different. Until the 7th month the curves for PFS overlap, subsequently they split slightly.

Death after start treatment was 100% similar extracted by CDC as by manual review (Table 3.2) and calculated median OS was 21.72 months (95% CI 18.69–24.75) versus 21.72 months (95% CI 18.59–24.84), which was equal for both methods (Figure 3.4B). Although CDC reports 77 events with respect to 75 for CDC versus manual review, the curves almost fully overlap.

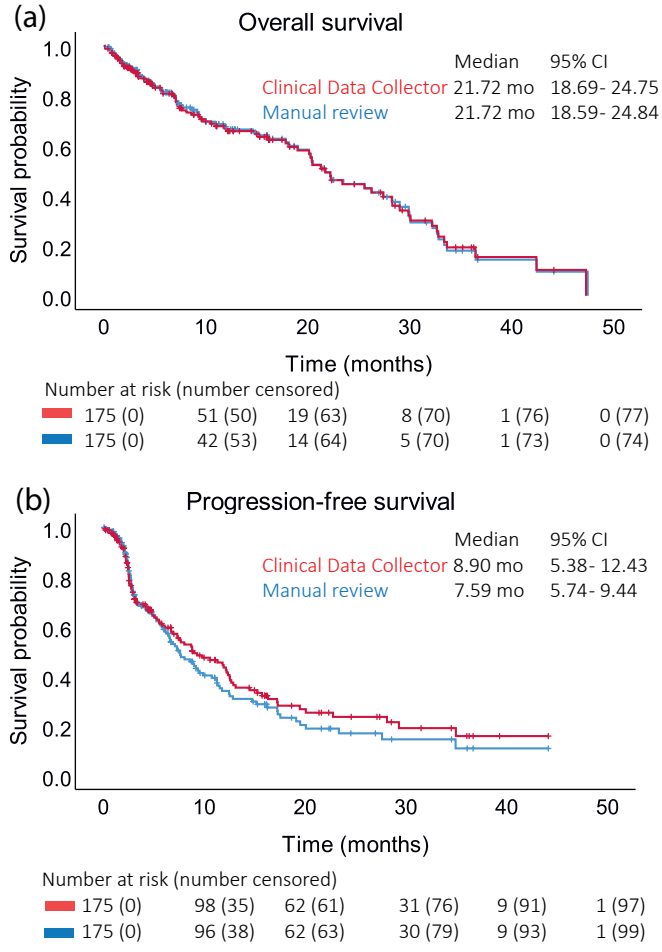

Figure 3.4. Kaplan-Meier plots of overall survival **(a)** and progression-free survival **(b)** determined from manual review and Clinical Data Collector data for cabozantinib, everolimus, nivolumab, pazopanib and sunitinib combined.

Bland-Altman plots with mean values of the continuous variables plotted against difference per value were composed to assess the quality of continuous data extraction by CDC (Figure 3.5). Since all confidence intervals include 0, differences of means were not significant. Data for age, ALAT, and ASAT showed the best concurrence between both methods.

3.3 Extraction time

The total time spent on patient inclusion and information extraction using CDC was 12 minutes per patient, in contrast to 86 minutes spent per patient during manual

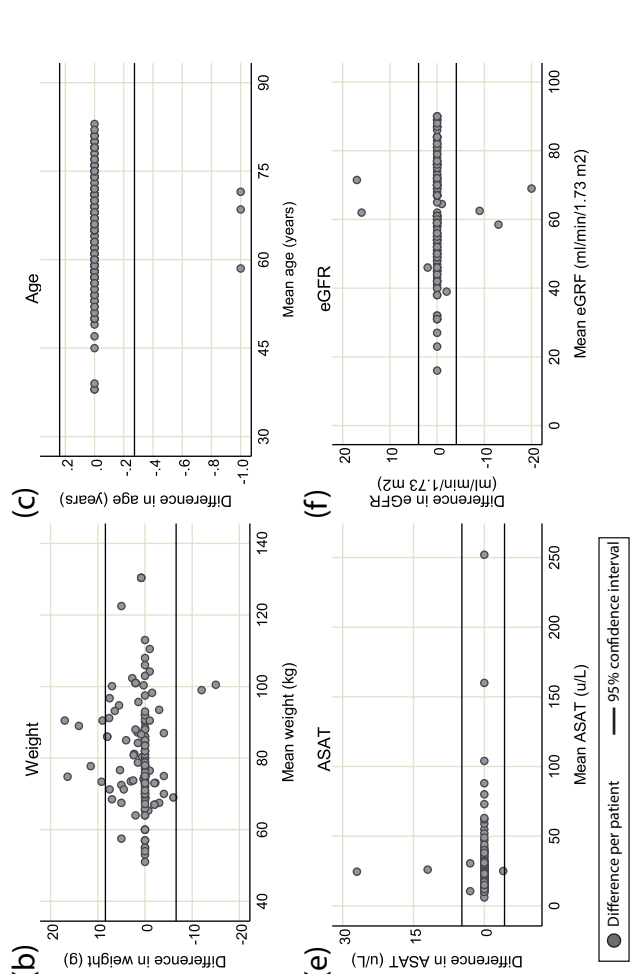

Figure 3.5. Bland-Altman plots of continuous variables collected using CDC versus manual with mean difference and 95% confidence interval. **(a)** Length: -0.21 cm (-4.2 to 4.8), **(b)** weight: 1.1 kg (-6.6 to 8.7), **(c)** Age: -0.17 years (-0.27 to 0.24), **(d)** Estimated glomerular filtration rate (eGRF) 0.22 ml/min/1.73m^2 (-5.3 to 5.8), **(e)** Alanine transaminase (ALAT) 0.19 U/L (-3.2 to 3.6), **(f)** Aspartate aminotransferase (ASAT) 0.24 (-4.0 to 4.5).

review. This indicates that use of CDC could result in a sevenfold time reduction for the information extraction.

4. Discussion

This study shows that main treatment outcomes such as PFS and OS can be accurately collected using CDC as NLP and text-mining software. These most important outcomes met the set standard of 90% for recall and precision. Furthermore, the Kaplan-Meier plots, including time to event, showed no significant differences. Therefore we conclude that CDC can be adequately applied to retrieve RWD from EHRs in order to add effectiveness data to complement the efficacy data already obtained from RCTs. We conclude that CDC shows to be a technical solution for more consistent and timely data collection [26]. To our knowledge, this is the first study that investigated the use of an information extraction tool to assess drug treatment outcomes in clinical practice.

Of all the extracted categorical patient characteristics, disease as well as drug related characteristics, prevalence of a prior nephrectomy, thrombocytosis or anemia before start of treatment met our standard of >90% for recall and precision. Although not all categorical data met the standard, in general, cancer-related variables and structured IMDC-criteria could be extracted reliably with CDC. Recall and precision were lower for comorbidities, ADEs, and unstructured IMDC-criteria. The differences between both data collection methods may be explained by the characteristics in the EHR for various types of data. First, variables with less variance in free-text registration options in the EHR, e.g., the structured IMDC-criteria such as laboratory values and cancer-related variables, showed higher accuracy. When data are retrieved using CDC, parts of the EHR are presented containing the predefined keywords. When there is low variety in words used to document variables, chances are higher that all relevant terms are covered in the CDC-queries. Also, variables which are stable, e.g., histological subtype of a tumor, seem to be more accurately extracted by CDC than variables of a temporary nature, e.g., comorbidities and ADEs. Wang et al. [16] already stated that ADE identification is complex, although identification tools as CDC can be complementary to manual review. Additionally, variables registered in the EHR with typing errors could be missed, unless they are specifically entered as search key. Since real-world oncologic treatment studies in general focus on primary outcomes as PFS and OS to study treatment effectiveness, we can accept a larger uncertainty for patient

characteristics and adverse events. However, for follow-up studies focusing on these secondary outcome parameters, improvement of queries with the already available software, or advancing CDC by automatizing synonym handling will be beneficial.

The use of CDC resulted in a sevenfold reduction in time for information extraction per patient, therefore the use of CDC can highly improve the efficiency in retrieving real-world data. In the 12 minutes spent per patient, verification of preselected patients and verification of variables was performed. Time spent for preparation of both methods was not taken into account. Manual review was prepared by constructing an eCRF and for applying CDC, queries were built. Whereof the latter was perceived as more time consuming, especially the construction of queries for unstructured data. However, these CDC queries can be used repeatedly for example in the same patient population at a later moment in time or in other hospitals.

We observed that inconsistencies in the EHR caused differences between both datasets. Firstly, we observed that the information regarding an event in structured data was occasionally not consistent with the description in free-text notes. This led to differences in data retrieval, since some structurally stored variables were extracted by CDC from their dedicated location only, whereas the same variables extracted by manual review could be verified consulting free-text notes. For example, errors in using DTC-codes could be manually corrected, which was the case in one of our patients. This also applies for inconsistencies between the structured medication list and free-text notes. Since information extraction by CDC was directly linked to the treatment period as deduced from the medication list (structured data), incomplete registration of medication use may influence the extracted values. To illustrate, in our study, start data for drug treatment was not consistent between structured and free-text notes for 72 treatments and differed from one day (34 cases) to more than one year (one case). Bowman (2013) also described the discrepancies between structured data fields and free-text, for example, for drug dosing instructions [27]. Furthermore, variables such as length, weight, and performance status had to be extracted from unstructured text since these data are not stored in an easily accessible EHR file for CDC. As the CDC extracts information exactly meeting the search criteria only, data can be missed, introducing differences. This may explain the large fraction of missing data and low accuracy scores for the Karnofsky performance status, since these scores are often not literally stated in the EHR. This was also recognized by Hanauer et al. [28], who underlined the variety and errors of numerical values registered in clinical note. For

a few patients, differences in length and weight between CDC and manually retrieved data were remarkably large. We expect these data were subject to measurement errors, typing errors or were just rough estimations by a physician.

It should be realized that EHR data are real-world data, and not a clean data file such as an eCRF created for research [1]. Therefore, discrepancies as well as errors are not completely unavoidable, especially when data are collected in retrospect. Awareness of these errors is necessary when effectiveness of a drug treatment in real life is assessed using data extracted automatically with a tool such as CDC. However, the results of our study show that despite discrepancies in a few cases, overall, continuous variables were the same between both methods.

This study validated the use of CDC for patient inclusion and data extraction directly on a real-world EHR, on a wide range of variables, as reported in RCTs. Comparison to the gold standard manual reviewed data showed accurate results. A limitation of the study design is that it focused on one type of cancer and its treatments in one Dutch hospital. Also, in this first study on the accuracy of CDC, data collection was initially performed by one person. Patients were only included for a maximum of approximately 4 years, therefore not all end-points were reached by the time of inclusion ended.

5. Conclusion

We conclude that by using CDC the efficiency of real-world data collection can be improved considerably, since patients could be adequately included and treatment outcomes and all structured data could be collected with no significant difference from manual review. Although, information extraction of unstructured data showed varying results on accuracy, we assume that with some effort suboptimal queries can be optimized for data collection. In the future these queries can be applied to obtain RWD for several other oncologic drug treatments as well as exported to other centers, which, in particular, can improve efficiency regarding larger and multi-center patient cohorts.

References

1. Franklin, J.M. and S. Schneeweiss, *When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?* Clin Pharmacol Ther, 2017. **102**(6): p. 924-933.
2. Bothwell, L.E. and S.H. Podolsky, *The Emergence of the Randomized, Controlled Trial.* 2016. **375**(6): p. 501-504.
3. Verweij, J., et al., *Innovation in oncology clinical trial design.* Cancer Treatment Reviews, 2019. **74**: p. 15-20.
4. Chen, E.Y., V. Raghunathan, and V. Prasad, *An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate.* JAMA Internal Medicine, 2019. **179**(7): p. 915-921.
5. Lakdawalla, D.N., et al., *Predicting Real-World Effectiveness of Cancer Therapies Using Overall Survival and Progression-Free Survival from Clinical Trials: Empirical Evidence for the ASCO Value Framework.* Value in Health, 2017. **20**(7): p. 866-875.
6. de Lusignan, S., L. Crawford, and N. Munro, *Creating and using real-world evidence to answer questions about clinical effectiveness.* J Innov Health Inform, 2015. **22**(3): p. 368-73.
7. Skovlund, E., H.G.M. Leufkens, and J.F. Smyth, *The use of real-world data in cancer drug development.* Eur J Cancer, 2018. **101**: p. 69-76.
8. Stewart, W.F., et al., *Bridging the inferential gap: the electronic health record and clinical evidence.* Health Aff (Millwood), 2007. **26**(2): p. w181-91.
9. Khozin, S., G.M. Blumenthal, and R. Pazdur, *Real-world Data for Clinical Evidence Generation in Oncology.* J Natl Cancer Inst, 2017. **109**(11).
10. Liu, Q., A. Ramamoorthy, and S.M. Huang, *Real-World Data and Clinical Pharmacology: A Regulatory Science Perspective.* Clin Pharmacol Ther, 2019. **106**(1): p. 67-71.
11. Sherman, R.E., et al., *Real-World Evidence - What Is It and What Can It Tell Us?* The New England journal of medicine, 2016. **375**(23): p. 2293-2297.
12. Casey, J.A., et al., *Using Electronic Health Records for Population Health Research: A Review of Methods and Applications.* Annu Rev Public Health, 2016. **37**: p. 61-81.
13. Assale, M., et al., *The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records.* Front Med (Lausanne), 2019. **6**: p. 66.
14. Ford, E., et al., *Extracting information from the text of electronic medical records to improve case detection: a systematic review.* Journal of the American Medical Informatics Association : JAMIA, 2016. **23**(5): p. 1007-1015.
15. Haerian, K., et al., *Detection of pharmacovigilance-related adverse events using electronic health records and automated methods.* Clin Pharmacol Ther, 2012. **92**(2): p. 228-34.
16. Wang, Y., et al., *Clinical information extraction applications: A literature review.* J Biomed Inform, 2018. **77**: p. 34-49.
17. Hernandez-Boussard, T., et al., *Mining Electronic Health Records to Extract Patient-Centered Outcomes Following Prostate Cancer Treatment.* AMIA Annu Symp Proc, 2017. **2017**: p. 876-882.
18. Sohn, S., et al., *Drug side effect extraction from clinical narratives of psychiatry and psychology patients.* J Am Med Inform Assoc, 2011. **18 Suppl 1**: p. i144-9.
19. *CTcue*. 25 june 2019]; Available from: https://ctcue.com/.
20. Choueiri, T.K., et al., *Cabozantinib Versus Sunitinib As Initial Targeted Therapy for Patients With Metastatic Renal Cell Carcinoma of Poor or Intermediate Risk: The Alliance A031203 CABOSUN Trial.* J Clin Oncol, 2017. **35**(6): p. 591-597.

21. Motzer, R.J., et al., *Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma.* N Engl J Med, 2015. **373**(19): p. 1803-13.
22. Sternberg, C.N., et al., *A randomised, double-blind phase III study of pazopanib in patients with advanced and/or metastatic renal cell carcinoma: final overall survival results and safety update.* Eur J Cancer, 2013. **49**(6): p. 1287-96.
23. Janssens, P.M.W., *Managing the demand for laboratory testing: Options and opportunities.* Clinica Chimica Acta, 2010. **411**(21): p. 1596-1602.
24. Hernandez-Boussard, T., et al., *Real world evidence in cardiovascular medicine: assuring data validity in electronic health record-based studies.* J Am Med Inform Assoc, 2019.
25. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* European journal of cancer (Oxford, England : 1990), 2009. **45**(2): p. 228-247.
26. Cave, A., X. Kurz, and P. Arlett, *Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe.* Clinical Pharmacology & Therapeutics, 2019. **106**(1): p. 36-39.
27. Bowman, S., *Impact of electronic health record systems on information integrity: quality and safety implications.* Perspect Health Inf Manag, 2013. **10**: p. 1c.
28. Hanauer, D.A., et al., *Complexities, variations, and errors of numbering within clinical notes: the potential impact on information extraction and cohort-identification.* BMC Med Inform Decis Mak, 2019. **19**(Suppl 3): p. 75.

Supplementary material

Supplementary Table S3.1A. Patient inclusion

Inclusion criterium	Type of data extraction	Searched terms	Time window	Comments
Treatment prescription	Medication request	Cabozantinib, Everolimus, Nivolumab, Pazopanib, Sunitinib		Start after jan 1, 2015
DBC* renal cell carcinoma	DBC	Specialism: 0313 with Diagnosis: 834 Specialism: 0306 with Diagnosis: 010		Diagnose behandel combinatie = diagnosis treatment combination, a code used for reembursements in the Netherlands

Supplementary Table S3.1B. Data collection

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
Sex	Male or female	Standard patient characteristic in CDC				
Type of renal cell carcinoma	Clear cell	Keyword search	Manually added: heldercellig niercelcarcinoom heldercellig adenomcarcinoom heldercellig niercarcinoom	Pathology report, radiology report, poli clinical letter or clinical letter		
	Papillary	Keyword search	From keyword synonym library: Papillair niercelcarcinoom papillary renal cell carcinoma chromophil renal cell carcinoma papillary (chromophil) renal cell carcinoma Manually added: papillair adenocarcinoom papillair niercarcinoom	Pathology report, radiology report, poli clinical letter or clinical letter		
	Chromofobic	Keyword search	Manually added: Chromofoob niercelcarcinoom Chromofoob adenocarcinoom Chromofoob niercarcinoom	Pathology report, radiology report, poli clinical letter or clinical letter		

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
	Undefined type†	Keyword search	From keyword synonym library: Niercelcarcinoom niercel Carcinoom Niercelca Niercelcarcinoom stadium Niercelkanker nierkanker Niercel adenocarcinoom Adenocarcinoom niercel adenocarcinoom van de nier Grawitz-tumor Hypernefroom Renal cell cancer Renal cell carcinoma	Pathology report, radiology report, poli clinical letter or clinical letter		* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error
Date of diagnosis	Oldest date of occurrence	Keyword search	From keyword synonym library: Niercelcarcinoom niercel Carcinoom Niercelca Niercelcarcinoom stadium Niercelkanker nierkanker Niercel adenocarcinoom Adenocarcinoom niercel adenocarcinoom van de nier Grawitz-tumor Hypernefroom Renal cell cancer Renal cell carcinoma			

			Manually added: Chromofoob niercelcarcinoom chromofoob adenocarcinoom Chromofoob niercarcinoom papillair niercelcarcinoom papillair adenocarcinoom papillair niercarcinoom heldercellig niercelcarcinoom heldercellig adenocarcinoom heldercellig niercarcinoom	
Fuhrman grade	Grade 1	Keyword search	Manually added: Fuhrman graad 1	
	Grade 2	Keyword search	Manually added: Fuhrman graad 2	
	Grade 3	Keyword search	Manually added: Fuhrman graad 3	
	Grade 4	Keyword search	Manually added: Fuhrman graad 4	
	Undefined grade†	Keyword search	Manually added: Fuhrman	† Especially these results were manually verified and if possible, defined outcomes were assigned.
Nephrectomy	Nephrectomy	Keyword search	From keyword synonym library: Nefrectomie Nefrectomieen Nephrectomy Kidney excision	

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
	No nephrectomy	Keyword search	Manually added: geen nefrectomie			
Length	Numerical value in cm	Measurement	Lengte 100 cm - 250 cm			
		Keyword search	Lengte 100 cm - 250 cm lengte	Measurement forms		
Start date treatment	First prescription start date	Medication request	Treatment			
Age at start of treatment	Numerical value in years |	Medication request	Treatment			| it is possible to automatically collect age at related dates for every data point in CDC
End prescription treatment	Latest prescription end date	Medication request	Treatment			

Stop treatment reason	Progression †	Keyword search	Manually added: progressie PD		Start date of documentation between 30 days before "End prescription treatment" until 30 days after "End prescription treatment"	* (informal) abbreviation † Especially these results were manually verified and if possible, defined outcomes were assigned.
		Report type		Radiology report		
	Side effects †	Keyword search	Manually added: Stop stoppen staken And per treatment: pazopanib pazo sunitinib suni cabozantinib cabo everolimus eve nivolumab nivo	Form: consult and Specialism: medical oncology		
	Death †					
	Other †					
Stop treatment	Date of documentation of "stop treatment reason"					

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
Progression during treatment	Progression ◊	Keyword search	From keyword synonym library: Tumorprogressie tumor progressie progressie tumor progressie van tumor progressie van tumoren neoplasma progressie Tumor progression Manually added: toename laesie nieuwe laesie te zien nieuwe laesies te zien groei groei tumor toename progressie pd	Radiology report	Start date of documentation between 30 days before "End prescription treatment" until "End prescription treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error ◊ Progression according to RECIST 1.1, therefore progression defined in radiology reports are included only. Additional limited search in free text is performed, in case progression in the radiology report is not noted with identified terms and therefore potentially missed.
		Keyword search	From keyword synonym library: Tumorprogressie tumor progressie progressie tumor progressie van tumor progressie van tumoren neoplasma progressie Tumor progression			
		Report type		Radiology report		

Date of progression	Date of documentation of "progression during treatment"					
Weight at treatment start	Numerical value in kg	Measurement ‡	Gewicht gew. Gew gw weegt Only measurement values between 40 and 300 kg		Measuement start date should be before or on the same day as "start date treatment"	‡ measurement search also in free text * (informal) abbreviation
		Keyword search	From keyword synonym library: woog gewicht weegt manually added: gew. Gw gew			
Egfr at treatment start	Numerical value in ml/min/1.73m2	Measurement	Egfr MDRD egfr CKD EPI		Measurement start date should be before or on the same day as "start date treatment"	Egfr: estimated glomerular filtration rate egfr MDRD: egfr using modification of diet in renal disease formula egfr CKD-EPI:egfr using chronic kidney disease epidemiology collaboration formula

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
ALAT at treatment start	Numerical value in u/L	Measurement	ALAT GPT ALAT		Measurement start date should be before or on the same day as "start date treatment"	ALAT: alanine aminotransferase GPT: glutamic-pyruvate-transaminase
ASAT at treatment start	Numerical value in u/L	Measurement	ASAT GOT ASAT		Measurement start date should be before or on the same day as "start date treatment"	ASAT: aspartate transaminase GOT: glutamic oxaloacetic transaminase
Calcium at treatment start	Numerical value in mmol/L	Measurement	Calcium albumine gecorrigeerd calcium alb. Gecorrigeerd		Measurement start date should be before or on the same day as "start date treatment"	* (informal) abbreviation
Hemoglobin at treatment start	Numerical value in mmol/L	Measurement	Hemoglobine hb		Measurement start date should be before or on the same day as "start date treatment"	* (informal) abbreviation

Trombocytes at treatment start	Numerical value in mmol/L	Measurement	Trombocyten	Measurement start date should be before or on the same day as "start date treatment"	
Neutrophils at treatment start	Numerical value in mmol/L	Measurement	Neutrofielen	Measurement start date should be before or on the same day as "start date treatment"	
WHO score at treatment start	Numerical value	Measurement	To extract who performance status (0-5) WHO WHO graad WHO performance WHO-klasse WHO klasse WHO-score WHO score performance score performance only measurement values ≤5 To extract karnofsky performance status (0–100) Karnofsky Karnofsky index Karnofsky performance Performance score performance Only measurement values ≤100	Measurement start date should be before or on the same day as "start date treatment"	X measurement search also in free text

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
Hypertension as comorbidity	Hypertension	Medication request	Antihypertensiva Overige antihypertensiva ACE-remmers Angiotensine-ii-antagonisten Calciumantagonisten Beta-blokkers Low-ceiling diuretica		Start date should be from 1 year before until the same day as "start date treatment"	
		Measurement	RR bovendruk RR NIPB Only measurement values >140			RR: riva-rocci NIPB: non-invasive blood pressure
		Keyword search	At least one of: Manually added: Voorgeschiedenis VG together in a document with: Hypertensie			* (informal) abbreviation
		Keyword search	Manually added: Hypertensie	Within form: consult; question: Anamnsese		
Cardiovascular comorbidity	Cardiovascular disease	Keyword search			Start date should be from 1 year before until the same day as "start date treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error

	At least one of: Manually added: Voorgeschiedenis VG Together in a document with: at least one of from keyword synonym library: STEMI N-stemi nstemi non-stemi hartaanval hartinfarct hart infarct Hartinfarcten hart infarct myocardinfarct myocard infarct myocardinfarcten myocard infarcten cardiaal infarct Manually added: atriumfibrilleren afib atriumfibrillatie atriale fibrillatie atriumfibrilleren atrium fibrilleren atriumfibrileren atriumfibrillen

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
			Hartfalen falen hart hartinsufficientie hart insufficientie cardiaal insufficientie decompensatio cordis decomp cordis decompensatie hartdecompensatie linksdecompensatie rechtsdecompensatie			
		Keyword search	From keyword synonym library: STEMI N-stemi nstemi non-stemi hartaanval hartinfarct hart infarct Hartinfarcten hart infarct myocardinfarct myocard infarct myocardinfarcten myocard infarcten cardiaal infarct	Within form: consult; question: Anamnsese		* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error STEMI: ST evelvation myocardial infarction

		Medication request	Manually added: atriumfibrilleren afib atriumfibrillatie atriale fibrillatie atriumfibrilleren atrium fibrilleren atriumfibrileren atriumfibrillen hartfalen falen hart hartinsufficientie hart insufficientie cardiaal insufficientie		
			Decompensatio cordis decomp cordis decompensatie hartdecompensatie linksdecompensatie rechtsdecompensatie		
Diabetes mellitus as comorbidity	Diabetes mellitus		Diabetesmiddelen Insulines en analogen bloedglucoseverlagende middelen		Start date should be from 1 year before until the same day as "start date treatment"

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
		Keyword search	At least one of: Voorgeschiedenis VG together in a document with at least one of: From keyword synonym library: Diabetes mellitus Diabetes DM2 DM1 DMT2 DMT1 Diabets Diabeet			* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error
		Keyword search	From keyword synonym library: Diabetes mellitus Diabetes DM DM1 DM2 suikerziekte	Within form: consult; question: Anamnese		
COPD as comorbidity	COPD	Medication request	Parasympaticolytica sympaticomymetica+parasympaticolytica		Start date should be from 1 year before until the same day as "start date treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error

		Keyword search	At least one of: Voorgeschiedenis VG Together in a document with at least one of: From keyword synonym library: Chronisch obstructieve longaandoening chronisch obstructieve luchtwegaandoening COPD			
		Keyword search	From keyword synonym library: Chronisch obstructieve longaandoening chronisch obstructieve luchtwegaandoening COPD	Within form: consult; question: Anamnese		
Hypertension as adverse event	Hypertension	Medication request	Antihypertensiva Overige antihypertensiva ACE-remmers Angiotensine-ii-antagonisten Calciumantagonisten Selectieve beta-blokkers Diuretica		Start date should be from 1 week after "start date treatment" until 30 days after "End prescription treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error PRES: posterior reversible encephalopathy syndrome
		Measurement	Bloeddruk bovendruk RR NIBP > 140 mmhg Bloeddruk onderdruk diastolische bloeddruk > 90 mmhg			

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
		Keyword search	From keyword synonym library: Hypertensie hoge bloeddruk hoog bloeddruk bloeddruk hoog bloeddruk hoge bloeddruk verhoogd bloeddruk verhogen verhoogd bloeddruk verhoogde bloeddruk bloeddruk verhoogde arteriele hypertensie HBP HTN hypertension high blood pressure systemic hypertension systemic arterial hypertension high blood pressure hypertensive disease	From specialism: medical oncology		
		Keyword search	From keyword synonym library: hypertensie hoge bloeddruk verhoogde bloeddruk manually added: PRES	Within form: consult; question: anamnese (anamnesis), lichamelijk onderzoek (physical examination), aanvullend onderzoek (additional examination) of specialism: medical oncology		

Liver toxicity as adverse event	Liver toxicity	Keyword search	Manually added: verhoogde leverwaarden hoge leverwaarden leverwaarden verhoogd verhoogde transaminases hoge transaminases transaminases verhoogd verhoogde aminotransferases hoge aminotransferases aminotransferases verhoogd verhoogd alat hoog alat stijging van alat alat stijging verhoogd asat hoog asat stijging van asat asat stijging leverenzymstijging leverenzymstijgingen levertoxiciteit leverenzymstoornis leverenzymstoornissen leverenzym stoornissen leverenzymstoornis Leverchemie stoornis leverfunctiestoornissen leverfunctie stoornissen leverfunctiestoornis leverfunctie stoornis	Exclusion: policlinical letters from specialism: medical oncology	Start date should be from 1 week after "start date treatment" until 30 days after "End prescription treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
		Measurement	ALAT GPT ALAT For female >175 U/L For male >225 U/L ASAT GOT ASAT for female >115 U/L for male >175 U.L			
Diarrhea as adverse event	Diarrhea	Medication request	Loperamide		Start date should be from 1 week after "start date treatment" until 30 days after "End prescription treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error
		Keyword search	From keyword synonym library: diarree diarre diaree diarrhee diarrhea manually added: dunne def dunne defecatie			

Hand-foot syndrome as adverse event	Hand-foot syndrome	Keyword search	From keyword synonym library: palmar-plantar erythrodysesthesia syndrome palmar-plantar erythrodysaesthesia syndrome palmar-plantar erythodysthesia palmoplaintair erytrodysesthesiesyndroom palmoplantaire erythrodysesthesie-syndroom Hand-foot syndrome Hand and foot syndrome secondary to chemotherapy Hand- en voetsyndroom secundair aan chemotherapie Chemotherapy-induced acral erythema erytheem erythemen erythemateuze dermatose erythema rode huid erythemateuze aandoening Huidschilfering schilfering huid huidexfoliatie	Start date should be from 1 week after "start date treatment" until 30 days after "End prescription treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occuring typing error

Supplementary Table S3.1B continues on next page.

Supplementary Table S3.1B. *Continued*

Data point	Collected answer	Type of data extraction	Searched terms	Keyword search restricted to	Time window	Comments
			Bulla bullae blaar blaren blaarvorming blaar vorming blaarvormingen bulleuze laesies bleb blister bulla manually added: hand-voetsyndroom handvoet hand voet syndroom hvs hfs ppe rode handen rode voeten schilferen schilferende huid			

	Keyword search	Manually added: hand voet syndroom hand voetsyndroom handvoet hand voet hand foot syndrome hand foot hfs palmoplantaire erytrodysesthesie syndroom palmoplantair erytrodysesthesiesyndroom palmar plantar erythodysthesia	Within form: consult; question: anamnese (anamnesis), lichamelijk onderzoek (physical examination), aanvullend onderzoek(additional examination) of specialism: medical oncology