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

Laar, S.A. van

Part IV

General discussion and summaries


Fuhrman graad 1
stage III a
thyroxine
verhoogd
Tumorprogressie
headache
borstkanker
Extrahepatisch
pazopanib
oestrogeen receptor
Hepatectomie
adjuvante
koude rill
metastasis
remdesiv

Chapter 9

General discussion

1. Introduction

Real-world data (RWD) on systemic cancer therapy are needed to complement the registration trial results as trials have limited external validity and cannot be used to investigate rare or long-term effects and treatment use in clinical practice [1-3]. Especially with the increasing incidence of cancer, but also an increase in treatment options and costs, insights into treatment effects in clinical practice are necessary to make informed treatment decisions [4, 5]. The electronic health record (EHR) is a rich source of RWD that has the potential to create these insights [6, 7]. However, as the EHR is a complex data source, with both structured and unstructured stored data, manual data collection has been the standard, which is laborious, costly and error-prone and therefore not scalable [8-10]. Text-mining techniques, including natural language processing, can be used for information extraction not from structured data but also from unstructured text documents, and thus might improve the efficiency of RWD collection from EHRs [10-13].

This thesis aimed to investigate whether this text mining EHR data would be a suitable method to collect real-world data on oncologic drug treatments. Therefore, we reviewed available real-world data sources and applications, performed a validation study and investigated whether this tool can be applied to evaluate treatment effectiveness, including the overall survival, progression-free survival and recurrence-free survival and prognostic factors, adverse events, treatment patterns and guideline adherence in a variety of cancer (and Covid-19) populations. The short answer is that it can be; the nuanced answer will follow in the discussion below.

1.1 The electronic health record and text mining

First, we will discuss the source and method used in this thesis. Of all RWD sources investigated in **Chapter 2** (case report forms, administrative claims data, patient reported outcomes, registries, wearable devices, and social media data), the EHR was the source with the most potential to study a range of real-world aspects of cancer patients and their treatments. Even though other sources have specific advantages, for example, claims data contains information across healthcare facilities, and patient reported outcomes can give insights into the patients' perception, the EHR is the most extensive and nuanced source overall. Especially, since cancer patients regularly visit the hospital for their treatment and evaluation, and therefore continuous patient data regarding their disease and treatment is captured in structured and unstructured parts of the EHR, e.g.,

the prescribed medication, laboratory data, and free-text notes [7]. However, the review also showed that for research purposes, the EHR as a real-world source has its limitations, first since notes and values can be uninterpretable, inconsistent, or erroneous, and second since data may be missing [14, 15]. As missing, or otherwise limited, data may influence the method as well as and the results. These will also be addressed later on.

CTcue (CTcue B.V., Amsterdam, The Netherlands), the EHR text-mining software used in this thesis, is a rule-based text-mining tool, meaning that the user/researcher sets a selection of rules, queries, for what patients to identify and what data to extract. To gain insight in whether a text-mining tool could be useful for cancer treatment evaluation, we performed a validation study **(Chapter 3**). Next to inclusion criteria, based on medication use and diagnosis-treatment codes (DBCs), a range of patient characteristics, treatments, survival outcomes, and adverse events (AEs) were collected both with text mining and manually from the EHR to estimate time spend per method and the precision, recall and F1-score of the text-mining tool.

It showed that with text mining, the time spend on information extraction could be reduced sevenfold. Part of the benefit in time with this method is in the patient identification. Since this was highly accurate, only using text mining for patient identification can already be beneficial, especially when a combination of inclusion criteria defines the population and this tool can work as a structured filter. Others also found that using text mining for patient identification alone can greatly decrease the time and resources needed to review charts [16, 17]. However, text mining can be used for more.

This study also showed the range in accuracy and recall of specific variables. Overall we could conclude that the risk factors based on laboratory values and variables extracted from free text with limited variability in terminology (e.g., type of cancer) showed higher accuracy than the extracted comorbidities, AEs, and risk factors stored in unstructured data. This can shortly be explained by the difference in variability of the data in the EHR, both related to the criteria themselves and how they are registered. As for the data extraction, the queries are comprised by the researcher, even though well informed, it is more likely to include all relevant terms of patient variables with low variance than with high variance. This potential difference of accuracy per variable key for researchers to keep in mind when designing queries.

Moreover, inconsistencies, errors, and omissions in the EHR contributed to differences between the results collected with text mining and manual data retrieval. If structur-

ally stored information (used for text mining) was not consistent with the free-text notes, this could be adjusted in the manual data retrieval only. For example, we found differences in start- and stop dates of the treatments and this influenced outcomes by text mining. Since data collection was in all studies related to the moment of start or stop of treatments. Furthermore, even though some data was stored structurally, it was at that moment not yet extractable with the text-mining tool, including the vital signs as length and weight, therefore had to be extracted from the free-text. This was updated in newer versions of the tool. And these variables furthermore sometimes had large fluctuations, which we assigned to measurement- or typing errors, or rough estimations from the healthcare professional. These limitations of the EHR are also well recognized in the literature. Bowman et al. [18] described the discrepancies between structured data fields and free-text. And for drug dosing instructions and Hanauer et al. [19] underlined the variety and errors of numerical values registered in clinical notes. Ideally, the EHR would contain all relevant data, and without any errors, ready for secondary use. But that is not the current practice, and might not be reached anytime soon. That does not mean that the data becomes useless, in contrary, however, these limitations must be taken into consideration when performing (text mining) studies with EHR data and decide upfront how to deal with it.

1.2 Patient inclusion

Defining fitting patient inclusion queries, based on the patient inclusion- and exclusion criteria, is a crucial step in performing a text-mining study. All studies we performed were medication oriented. Therefore, we used medication prescriptions as the basis for the patient inclusion. We combined this with the diagnosis-treatment codes (Dutch: diagnose-behandelcombinatie-codes; DBC) in all studies reviewing oncologic treatments (**Chapter 4–7)**. The advantage of using prescribed medication and DBCs is that they are, in general, structured and reliable data, and therefore identify all potential patients. However, as treatments can have more than one specific indication and DBCs can still be too broad for the specific aimed diagnosis and for several other reasons, we found that the combination of medication and DBC was not specific enough to include no "false positives". Since an accurate patient population is the basis of a study, and is an important step in collecting reliable real-world data, we have added a third component. For all these studies, discriminating free-text searches were added to manually verify the population with context validation.

Context validation was used to exclude patients from the initial search. In **Chapter 4 and 5,** patients were excluded since they started treatment before the inclusion period, did not start the prescribed treatment, or received the treatment for a different type of cancer. An even more accurate registration of start of treatments and the indication could partly have prevented the patients ending up in our initial search. Furthermore, in **Chapter 6,** the added value of the verification was mostly to exclude melanoma patients that received palliative treatments, instead of the intended adjuvant treatment, as the DBC code does not discriminate between these two. Furthermore, in **Chapter 7,** the verification method was used to distinguish between a range of treatment courses, for example, the regular doxorubicin and cyclophosphamide (AC) course and a dose dense AC course. The advantage of this semi-automatic method was that with these three components already a large part of the manual labor is replaced, and this semi-automatic concept can also be used in other studies as a stepping stone to full automized patient inclusion. McKenzie et al. [20], for example, also used a semi-automatic approach to identify patients, in this case whether lung cancer patients developed radiation pneumonitis. However, all manual labor limits the scalability of the method. Therefore, further optimization and validation of the inclusion queries of these studies are needed before they can be implemented on a large scale without manual verification. Only in **Chapter 8,** patient inclusion was possible without additional verification since remdesivir use and a positive PCR for Covid-19 – both structured data – was already highly specific. This illustrates, furthermore, the potential for automatic patient identification.

1.3 Baseline characteristics

Baseline characteristics are key in every study. They are needed to describe the population in order to provide context to outcomes. This can be done by comparing the study population to other populations or by distinguishing subgroups within the population. For patients with cancer, these characteristics can further be roughly categorized in patient-, disease or tumor-, and treatment related characteristics. In our studies, these were collected with different approaches linked to the start date of the specific treatment.

Patient-related characteristics mainly include sex, age, length, weight, a selection of relevant laboratory values, including renal- and liver function, values that are already known to be predictive for an outcome of that type of disease, and the performance

status. The vital signs and laboratory values were found to be extracted with high accuracy from their structured source. By design of the queries, automatically the most recent value before the start of the treatment was selected for data extraction. This significantly reduced the time spend on data collection and the risk of typing errors when manually transferring these values to an electronic collection report form, also found by others [21]. These patient data already generate added value.

In **Chapter 8**, the structured laboratory values were sufficient to characterize patients according to their kidney-, and liver function. And in **Chapter 5** we could determine that at least 28% of the patients would not be eligible for participation in the pivotal trials based on values from structured data. However, sometimes these well-defined and/or structured data will not cover all desired patient characteristics in a study. In **Chapter 7**, investigating the use of G-CSFs in patients with intermediate risk of febrile neutropenia (FN), we limited the inclusion of the risk factors for FN to structured variables. To ensure high accuracy risk factors that were too broad or unclear to extract in high accuracy were excluded. One patient characteristic, the performance status, is notorious for being underreported in the EHR [22, 23]. In **Chapter 4 to 7** between 35% to 55% of the performance statuses could not have been retrieved. However, this rate is comparable to other real-world studies [22, 23]. As the performance status of patients is often a significant predictor of outcomes, irrespective of a specific cancer type, it would be beneficial for real-world treatment evaluations if this specific characteristic was better registered.

Next, often it is relevant to further subdivide the investigated cancer type based on its histology-, tumor receptor-, and/or oncogene mutation profile since these can be predictive for the overall survival or the clinical outcome of medical treatments. These details are often stored in unstructured text. And, even though these characteristics are well-defined, and therefore have the potential to be extracted with high accuracy with text mining, we encountered several challenges. In **Chapter 4** we identified the tumor histology by including free-text key words not only for the most prominent clear-cell type, but also papillary- and chromophobic renal cell carcinoma (RCC). However, as a high variety of other rare subtypes are present, we scanned the texts by adding a general query instead of adding all potential variables, and identified by context analysis the sarcomatoid subtype in the population. With this, we were able to observe that specifically sunitinib treated patients had a non-clear cell histology and therefore identify a potential bias by indication in this group that is underrepresented

in trials [24]. This was not only reflected in clinical practice in which sunitinib was first choice for non-clear cell histology, but also could partially explain differences in results between the sunitinib and pazopanib treated group in our study population.

Furthermore, we had to adapt search queries regarding disease characteristics, since punctuation, in general, is ignored by the text-mining program, and the symbols as "+" and "-", often used to indicate presence or absence of a specific mutation or receptor cannot be included in searches. Removal of punctuation is often done in the pre-processing of text data for text mining [25]. In **Chapter 6 and 7** we extracted the presence or absence of a specific receptor or mutation. Since punctuation is ignored by the text-mining program, including, for example, "PR+" a common method to note progesterone receptor positive, would not result in the desired outcomes. Therefore, other combinations of the specific gene combined with options used textually to note the presence or absence of a mutation were searched for. Specifically, in **Chapter 6** these results were used to show that all patients receiving dabrafenib plus trametinib had a BRAF mutation, but also up to half of the patients receiving immunotherapy. This was relevant, since for this group of melanoma patients both therapies can be considered.

Also for cancer staging we had to adapt the search method to the practice. Cancer staging can be classified with the TNM staging system, composed of the primary tumor site and size (T), regional lymph node involvement (N) and the presence of metastasis (M) [26]. As this is a general standard to register the stage, we expected that is would be easy to use this for the determination of the disease stage. However, due to the complexity in the reporting of this score with subcategories, e.g., T1aN2M1b, the variants reported in the EHR were too diverse to include in the search. Other attempts for retrieving the TNM classification by text mining showed varying accuracy ranging from 64–94% for the retrieval of the different T, N, and M components [27, 28]. In **Chapter 6**, only disease stage was used instead. With this, enough insight in disease severity could be established by retrieving stage from the EHR for most patients.

Treatment-related variables do not only describe the specific treatment investigated, but can also be used to put this treatment in perspective by including previous and following treatments. All systemic cancer treatments could be extracted from the medication lists. Even though the validation study in **Chapter 3** highlighted that start- and stop date may not always be correct, omissions will be highly unlikely since patients then would not receive their treatment. Vice versa, verification showed that not all

patients started their intended treatment, due to multiple reasons as already further progressive disease. With these treatment-related variables, a variety of insights can be created. The field of RCC and HCC treatments are undergoing major developments. Therefore, we investigated treatment patterns in **Chapter 4 and 5.** These patterns add relevant insights, as for the HCC population the data illustrate the introduction of new treatments after a decade with only sorafenib as treatment option. And in the RCC population, compared to the first trials of sunitinib and pazopanib, significantly more patient received follow-up treatments, and the type of treatments also changed. In our study, these follow-up treatments might have added to improved overall survival. However, this does not only highlight the relevance of real-world data on treatment patterns, but also the complexity in interpreting results. With the introduction of a new treatment, more treatment lines can be followed, which contribute, next to patient- and cancer characteristics, to the outcome. Even though estimations are made, for example by investigating the (improvement of) overall survival between time periods, or investigating multiple lines of one treatment [29, 30] it is difficult to estimate, based on these complex data, the added benefit of every individual treatment to the overall survival of a patient.

However, not all treatments could not be extracted by using the medication list. Other types of treatment as radiology interventions (e.g. TACE in **Chapter 5**) or surgical interventions (e.g. nephrectomies in **Chapter 3/4**) were determined by text mining. And in **Chapter 7** we had to perform free-text search to identify granulocyte colony-stimulating factor (G-CSF) as G-CSF treatment is extramural therapy and only documented in unstructured text. However, as use of G-CSF was also a study outcome and we wanted to ensure high accuracy, we did not only perform context validation, but also manual validation in the EHR of G-CSF use for the patients that were identified as potential users by text mining. This was not only followed by context validation but also manual validation in the EHR to determine whether patients started G-CSF treatment.

To summarize, all these examples of baseline characteristics show the variability in documentation and some accompanying challenges. Over the chapters we have shown all approaches we have used in an attempt to optimize efficiency and accuracy in the data extraction of the baseline variables, which can be used as basis for new studies.

1.4 Treatment effectiveness

The most important endpoint to assess the effectiveness of cancer treatments is the measurement of overall survival (OS). Determining the overall survival time per patient based on real-world data from the health records is also relatively easy since the start of treatment and time of death are structurally and accurately stored in the EHR. The accuracy was confirmed in our validation study (**Chapter 3)** which showed an almost identical Kaplan-Meier curve when calculating the median OS based on manual versus text-mining data. We first used this to estimate the effectiveness of RCC treatments in **Chapter 4** and reused the queries in **Chapter 5** now only regarding HCC treatments. This resulted in the OS of first-line treatments pazopanib, sunitinib, nivolumab (RCC), and sorafenib (HCC). As only data that can be extracted from structured sources is needed to calculate the effectiveness, this could be something to use on a large scale to gain very general insights into whether treatments are effective in clinical practice, especially since this has not always been the case [31]. And it could, for example, be used to yearly benchmark hospital results. Of course, only effectiveness data will not be enough, but with these data trends can be identified.

Since the median time to OS, fortunately, often takes years, surrogate endpoints are often used in trials for an earlier estimation of treatment efficacy. Often-used surrogate end-points are progression-free survival (PFS), in case of palliative disease, and recurrence-free survival (RFS) or disease-free survival (DFS) for curative disease. For the extraction of the PFS of pharmacological treatments for RCC and HCC, and RFS for adjuvant melanoma (**Chapter 4–6)**, a more elaborate method was used. Even though the measurement of tumor response is standardized in the Response Evaluation Criteria In Solid Tumors (RECIST), the documentation in the radiology reports was unstructured and very high dense and variable in information. Therefore, these radiology reports were not fit for direct text mining. In our approach, we first identified relevant free-text fields to determine reason to end treatment. This was linked to the last date of use of a treatment according to the prescription data. Following, in case of recurrence or progression, the radiology report diagnosing the case was selected, to have the most accurate date of the event. Even though this method still contains some manual labor, it also is a significant reduction and shown to be accurate in the validation study. This method is, in basis, comparable to a mix of the clinician-anchored approach radiology-anchored approach of Griffith et al. [32], who investigated methods to identify real-world progression in EHRs, and showed that the combination was most optimal.

The difficulty with survival results in our studies, irrespective of the collection method, was the interpretation of the results because of the large confidence intervals due to small sample sizes and sometimes limited follow-up time (e.g., the recurrence-free survival in **Chapter 6**). Mostly, we could detect trends but lacked statistical significance to make conclusions. Therefore, it might be beneficial to not only aim to include more patients, but also to – a priori – determine the sample size and power necessary to detect meaningful insights [33]. Even when sample sizes are not reached, this can put the outcomes better in perspective. Also, it is known that not all surrogate endpoints are not always good predictors of overall survival, even the well-established progression-free survival [31, 34]. Therefore, it remains a challenge to quickly establish treatment effectiveness based on real-world data, since it takes time for enough patients to reach the median OS.

1.5 Treatment safety

Even though adverse events are monitored before the marketing approval of a treatment, post approval monitoring is important to identify 1) adverse events occurring in subgroups, especially patients not meeting trial eligibility criteria, 2) adverse events that occur later after the start of a medication, and 3) rare events [35]. In this thesis we also used text mining to estimate the incidence of adverse events.

In **Chapter 3**, we extracted regular occurring adverse events during treatment with targeted therapies, such as diarrhea, liver toxicity, hypertension, and hand-foot syndrome to test text mining for AE extraction. In general, we set a threshold for accuracy of 90%, and for none of these adverse events this threshold was reached. This relative low accuracy be attributed to the different and tangled way AEs are recorded in the EHRs. What further might contribute is the – often – temporary nature of adverse events. Wang et al. [36] already stated that ADE identification is complex. However, we argue that a larger range of uncertainty may be acceptable for the detection of adverse events, compared to treatment effectiveness, as at least the results can indicate whether an AE occurs distinctively more, less, or comparable to the trials.

The first study we implemented automated data extraction to evaluate potential adverse events was slightly out of scope of this thesis, nevertheless fully in line with the period in which this thesis was written and the urgent need for data on pharmacological treatments for Covid-19. In **Chapter 8,** fast data extraction was performed to test the hypothesis that remdesivir could result in nephrotoxicity and hepatotoxicity

in Covid-19 patients. Especially since patients with poor kidney- and liver function were contra-indications. No severe nephrotoxicity, and in two patients grade 3 hepatotoxicity, was found in the first fifteen days after treatment initiation. Furthermore, the incidence of decreased eGFR was comparable with data reported in the RCTs last year [37, 38]. And none of the patients that started with transaminases above ULN, including patients meeting exclusion criteria, had grade 2 or higher transaminase elevation. Based on these results of structured data, we concluded that patients meeting contra-indications seem not to develop more renal- or hepatotoxicity on remdesivir treatment, and this may be no reason for these contra-indications to be absolute. It has to be mentioned that since later no effectiveness was shown for remdesivir treatment of Covid-19, it is no longer advised in The Netherlands [39]. However, in general, it shows how fast(er) access to and publication of data is of relevance for adverse events.

Next, in **Chapter 7,** we investigated the use of G-CSFs to prevent the adverse event of chemotherapy-induced febrile neutropenia in breast cancer patients. Even though neutropenia can be identified based on laboratory results, the "febrile" aspect is not always documented as elevated body temperature in the structured vital signs, even though this is potentially a reason for hospitalisation. Since this adverse event was a crucial outcome, we identified the potential cases of neutropenia and febrile neutropenia with text mining, and additionally performed manual validation by EHR review. This is another method to use text mining and simultaneously reduce the risk of incorrect data by only performing this validation on the selection of patients [40]. Furthermore, even though the low use of G-CSFs in the intermediate risk group (<20% FN risk), the FN incidence was moderate (12%), but a substantial amount of the patients developed grade 3-4 neutropenia (31%). Therefore, we concluded that higher G-CSF use could further have prevented cases of severe neutropenia, and with it concomitant chemotherapy dose delay or reductions [41, 42].

In the real-world evaluation of adjuvant nivolumab, pembrolizumab, and dabrafenib plus trametinib for melanoma in **Chapter 6**, we not only investigated the ten most frequent adverse events reported in the pivotal trials (including the severity), but also the treatment-limiting adverse events in this population. To estimate the incidence of the common adverse events, we used extensive free-text search. With this we, for instance, found that manifestations of thyroid-related AE after use of both nivolumab of pembrolizumab seemed to occur more often in this population than in the trials. Even though a thyroid-related AE not necessarily leads to treatment termination, or

can be classified as severe, patients often need lifelong hormone replacement [43], and therefore the occurrence of this type of AE might have a severe impact on an individual patient's quality of life. Furthermore, we were able – to a limited extent – to identify the severity of adverse events based on the context corresponding with the CTCAE. De Meza et al. [44] did report higher rates of severe AEs in clinical practice compared to the RCTs. However, due to the low incidence of the individual severe AEs, we could not determine if this was the case for our study population.

To identify the treatment-limiting AE, we used the parts of EHR used to select the reason to end treatment and were able to select the specific treatment-limiting AE in the program. This showed that, in general, nivolumab and pembrolizumab had less treatment-limiting toxicity compared to dabrafenib plus trametinib. Furthermore, for both immunotherapies, mostly one or two concurring immune-related AEs (irAEs) were treatment limiting. These irAEs are characterized by auto-reactive T-cells and can affect many organs. This contrasts with combinations of two up to six treatment-limiting AEs on dabrafenib plus trametinib treatment [45-47]. Most outstanding treatment-limiting adverse event due to dabrafenib plus trametinib was pyrexia, occurring in more than half of the patients who ended treatment due to toxicity. Even though pyrexia is a well-known AE of dabrafenib plus trametinib, the rate in which it occurred seemed higher than in the trial [48]. Overall, we could use the analysis of treatment-limiting adverse events to conclude that, in this population, adjuvant immunotherapy was better tolerated than dabrafenib plus trametinib, which is relevant to consider when initiating adjuvant melanoma treatment.

Overall, the different chapters show that, AE can be gathered with multiple approaches of text mining EHRs. This is also done by others. But, e.g., Mashima et al. [49] used a machine-learning approach to develop an algoritm to detect nausea/vomiting and diarrhea. The difficulty with AEs is that for every event a query has to be developed. And, since there are many possible AEs, it will take a lot of effort to develop and validate the extraction methods of AEs.

2. Future perspectives

The amount of available real-world data in the EHR of patients with cancer receiving systemic pharmacologic treatments is constantly growing, and this thesis shows that with text mining these data can be captured for several research and evalua-

tion purposes. These data have the potential to replace or add data to the evidence generated in traditional research methods by being able to generate datasets quicker, cheaper or datasets that are bigger, and generate insights in data that were previously not available. However, by improving registration at the source and by application of text-mining techniques, the use of RWD can reach its full potential.

2.1 Improving EHR

The quality of the data is essential to establish meaningful real-world evidence. However, the challenge of the EHR is that the primary purpose for data registration is not research. Therefore, the data has limitations. Part of the solution can be improving and harmonizing the data capture by generating more user-friendly and clear front-ends of the used EHR software so that healthcare professionals easily can store information for secondary use. However, fundamentally, it remains impossible to capture everything. Thus, relevant baselines should be established and harmonized over all the used EHRs [50]. However, these alterations should not add to the burden of registration. By requesting only the registration of relevant variables, healthcare professionals will still be motivated [51]. Reflecting on cancer treatment care, it might be relevant to systematically capture the diagnosis and stage for all patients at the start of every treatment. But, in case of laboratory values, one might argue that it is reasonable that only relevant parameters for patient care measured. Furthermore, regarding adverse events, not every symptom might be recognized as an adverse event, and writing down all symptoms can be burdensome. However, optimized data registration in the EHR, making it suitable for text mining, can potentially reduce time spent on validating data extracted for research purposes and registries, and therefore motivate healthcare professionals.

Also, other methods of data collection and entry into the EHR can be implemented. For example, integration of patient reported outcome measures (PROMs) and wearable devices into the EHR can add extra perspective. Not only for research purposes, but also/especially for healthcare practice. Some devices and questionnaires have been piloted or implemented in EHR systems already. However, limited interoperability of systems is currently still a barrier [52]. Also, in the future, manual data entry can potentially be replaced with new techniques as automatic speech-recognition based clinical documentation, in which conversations between patients and healthcare professionals are captured and automatically documented. However, these speech to text-techniques are still in development [53].

2.2 Development of text-mining use

(Semi-)automatic data extraction, as shown in this thesis, can significantly reduce time needed to extract relevant real-world data for treatment evaluation. But, as this is still a relative new technique, further developments can be made, both technical and in the used procedures. The challenge in the field of NLP is generating a query set or algorithm that is specific and precise enough to fit the often heterogenous EHR (72). The rule-based extraction, used in this thesis, is in the basis a coded set of rules for data extraction. And especially use of elaborate searches in free-text notes is of added value. However, currently, the data retrieval is limited to the synonyms provided, and recognition of typing errors or the sentiment (in short: is the mention of a variable positive, negative, or neutral) is not yet possible in the used software. To enable these, often machine-learning based NLP is necessary, a system based on a statistical learning algorithm that identifies the relevant context based on annotated training data (70). With large data sets as training data, and more uniform data, use of machine learning to establish the algorithms might be the next step.

Furthermore, not only the literal use of text mining can be improved, but also the agreement on procedures followed when implementing. For example, in **Chapter 3** we choose to formally validate individual queries by comparison to manual data collection and calculate precision, accuracy and F1-scores, and set 90% as benchmark. However, no guideline or standard is set on how and when to manually validate, including a minimal percentage or number of records/data points, what benchmarks needs to be met, and how to verify if these queries can also be applied in other centers. This should not be a one-size-fits all agreement, but be dependent on the variables, research questions and population size. For example, higher accuracy can be relevant for the primary end-points.

2.3 Active use of real-world data in pursuit of value-based healthcare

This thesis showed that text mining EHRs can create insights into real-world data on cancer treatments: the primary aim of this research. The data can be used – as previously discussed – to evaluate treatment effectiveness, adverse events, patient populations, treatment patterns, and guideline adherence, reflecting current care. By text mining, more data can be gathered for evaluation, creating the potential for more insights into clinical practice of oncologic treatments. These real-world data are of the most value when actually actively used.

More concretely, ideally, real-world data is collected regularly and with or without the use of text mining. The results should be linked to concepts of value-based healthcare, in which measuring outcomes (and costs) is only one step of the cycle. A step further in this direction is using these data to refine funding recommendations needed to challenge currently escalating costs [54]. In case benefits in clinical practice do not match the results on a national level, they potentially could be reduced accordingly. Additionally, real-world data can play a role in other payment schemes, such as outcome-based payments. Lorgelly et al. highlighted that not only survival is an outcome of importance for cancer patients, but also other tumor progression and quality of life aspects – which should and can be included [55]. Reducing costs is only one aspect. Using these data to regularly reflect on outcomes on the hospital-, regional- or (inter)national level and change or verify treatment policies is a second. And actively implementing insights from real-world data for shared patient and clinician decision making [56].

As this discussion highlights once more, there is a high need for real-world data to evaluate cancer treatments in clinical practice, and the EHR is a source containing these. This thesis shows that text mining can extract the data that create insights on a variety of aspects, including patient characteristics, treatment outcomes, adverse events and treatment patterns by a variety of different cancer patient populations and their treatments. However, this method is relatively new, and still can be optimized by improving the source, the EHR, and further developing the use of text mining. But, ultimately text mining can be a critical component in the generation of more real-world data. This must result in better insights in all aspects of a cancer treatment, especially what treatment would benefit an individual patient best in every phase of their disease, and therefore optimize systemic cancer therapy.

References

1. Makady, A., et al., *What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews.* Value in Health, 2017. **20**(7): p. 858-865.
2. 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.
3. Ramamoorthy, A. and S.-M. Huang, *What Does It Take to Transform Real-World Data Into Real-World Evidence?* Clinical Pharmacology & Therapeutics, 2019. **106**(1): p. 10-18.
4. *NKR cijfers*. 10 December 2022]; Available from: https://iknl.nl/nkr-cijfers.
5. J. Praagman, E.S., L. van Disseldorp, V. Lemmens, *Kanker in Nederland - trends & prognoses tot en met 2032*. 2022.
6. Kim, E., et al., *The Evolving Use of Electronic Health Records (EHR) for Research.* Seminars in Radiation Oncology, 2019. **29**(4): p. 354-361.
7. Cowie, M.R., et al., *Electronic health records to facilitate clinical research.* Clin Res Cardiol, 2017. **106**(1): p. 1-9.
8. 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.
9. 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.
10. 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.
11. 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.
12. Wong, A., et al., *Natural Language Processing and Its Implications for the Future of Medication Safety: A Narrative Review of Recent Advances and Challenges.* Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 2018. **38**(8): p. 822-841.
13. Sun, W., et al., *Data Processing and Text Mining Technologies on Electronic Medical Records: A Review.* Journal of Healthcare Engineering, 2018. **2018**: p. 4302425.
14. Hersh, W.R., et al., *Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research.* Medical Care, 2013. **51**.
15. Gianfrancesco, M.A. and N.D. Goldstein, *A narrative review on the validity of electronic health record-based research in epidemiology.* BMC Medical Research Methodology, 2021. **21**(1): p. 234.
16. Labrosse, J., et al., *Text Mining in Electronic Medical Records Enables Quick and Efficient Identification of Pregnancy Cases Occurring After Breast Cancer.* JCO Clin Cancer Inform, 2019. **3**: p. 1-12.
17. Maguire, F.B., et al., *A text-mining approach to obtain detailed treatment information from free-text fields in population-based cancer registries: A study of non-small cell lung cancer in California.* PLOS ONE, 2019. **14**(2): p. e0212454.
18. Bowman, S., *Impact of electronic health record systems on information integrity: quality and safety implications.* Perspect Health Inf Manag, 2013. **10**: p. 1c.
19. 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.

20. McKenzie, J., et al., *A Semiautomated Chart Review for Assessing the Development of Radiation Pneumonitis Using Natural Language Processing: Diagnostic Accuracy and Feasibility Study.* JMIR Med Inform, 2021. **9**(11): p. e29241.
21. Feng, J.E., et al., *Transcription Error Rates in Retrospective Chart Reviews.* Orthopedics, 2020. **43**(5): p. e404-e408.
22. Board, R., et al., *Metastatic melanoma patient outcomes since introduction of immune checkpoint inhibitors in England between 2014 and 2018.* Int J Cancer, 2021. **148**(4): p. 868-875.
23. Moser, J.C., et al., *Real-world survival of patients with advanced BRAF V600 mutated melanoma treated with front-line BRAF/MEK inhibitors, anti-PD-1 antibodies, or nivolumab/ipilimumab.* Cancer Med, 2019. **8**(18): p. 7637-7643.
24. Ahrens, M., et al., *Non-Clear Cell Renal Cell Carcinoma - Pathology and Treatment Options.* Oncol Res Treat, 2019. **42**(3): p. 128-135.
25. Abbe, A., et al., *Text mining applications in psychiatry: a systematic literature review.* International Journal of Methods in Psychiatric Research, 2016. **25**(2): p. 86-100.
26. Edge, S.B. and C.C. Compton, *The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM.* Annals of Surgical Oncology, 2010. **17**(6): p. 1471-1474.
27. Najafabadipour, M., et al., *Reconstructing the patient's natural history from electronic health records.* Artificial Intelligence in Medicine, 2020. **105**: p. 101860.
28. Spasić, I., et al., *Text mining of cancer-related information: Review of current status and future directions.* International Journal of Medical Informatics, 2014. **83**(9): p. 605-623.
29. Venugopal, B., et al., *Early Clinical Experience with Cabozantinib for Advanced Renal Cell Carcinoma in the UK: Real-World Treatment Pathways and Clinical Outcomes.* Clinical Genitourinary Cancer, 2022. **20**(1): p. 94-94.e10.
30. Ratta, R., et al., *Exposure to Multiple Lines of Treatment and Survival of Patients With Metastatic Renal Cell Carcinoma: A Real-world Analysis.* Clinical Genitourinary Cancer, 2018. **16**(4): p. e735-e742.
31. Davis, C., et al., *Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13.* BMJ, 2017. **359**: p. j4530.
32. Griffith, S.D., et al., *Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer.* Advances in Therapy, 2019. **36**(8): p. 2122-2136.
33. Rivera, D.R., et al., *The Friends of Cancer Research Real-World Data Collaboration Pilot 2.0: Methodological Recommendations from Oncology Case Studies.* Clinical Pharmacology & Therapeutics, 2022. **111**(1): p. 283-292.
34. Belin, L., et al., *Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review.* British Journal of Cancer, 2020. **122**(11): p. 1707-1714.
35. Lavertu, A., et al., *A New Era in Pharmacovigilance: Toward Real-World Data and Digital Monitoring.* Clinical Pharmacology & Therapeutics, 2021. **109**(5): p. 1197-1202.
36. Wang, Y., et al., *Clinical information extraction applications: A literature review.* J Biomed Inform, 2018. **77**: p. 34-49.
37. Wang, Y., et al., *Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial.* The Lancet, 2020. **395**(10236): p. 1569-1578.
38. Goldman, J.D., et al., *Remdesivir for 5 or 10 Days in Patients with Severe Covid-19.* New England Journal of Medicine, 2020. **383**(19): p. 1827-1837.

39. FMS/SWAB. *Remdesivir*. 3 October 2022; Available from: https://richtlijnendatabase.nl/richtlijn/covid-19/behandeling/medicamenteuze_behandeling_voor_patienten_met_covid-19/remdesivir.html.
40. Clemons, M., et al., *A multicentre, randomised trial comparing schedules of G-CSF (filgrastim) administration for primary prophylaxis of chemotherapy-induced febrile neutropenia in early stage breast cancer.* Ann Oncol, 2020. **31**(7): p. 951-957.
41. Pettengell, R., et al., *Neutropenia occurrence and predictors of reduced chemotherapy delivery: results from the INC-EU prospective observational European neutropenia study.* Supportive Care in Cancer, 2008. **16**(11): p. 1299-1309.
42. Liutkauskiene, S., et al., *Retrospective analysis of the impact of anthracycline dose reduction and chemotherapy delays on the outcomes of early breast cancer molecular subtypes.* BMC Cancer, 2018. **18**(1): p. 453.
43. Darnell, E.P., et al., *Immune-Related Adverse Events (irAEs): Diagnosis, Management, and Clinical Pearls*. Current Oncology Reports, 2020. **22**(4): p. 39.
44. de Meza, M.M., et al., *Adjuvant treatment for melanoma in clinical practice – Trial versus reality.* European Journal of Cancer, 2021. **158**: p. 234-245.
45. Postow, M.A., R. Sidlow, and M.D. Hellmann, *Immune-Related Adverse Events Associated with Immune Checkpoint Blockade.* N Engl J Med, 2018. **378**(2): p. 158-168.
46. Khan, S. and D.E. Gerber, *Autoimmunity, checkpoint inhibitor therapy and immune-related adverse events: A review.* Seminars in cancer biology, 2020. **64**: p. 93-101.
47. Mor, A. and M. Strazza, *Bridging the Gap: Connecting the Mechanisms of Immune-Related Adverse Events and Autoimmunity Through PD-1.* Frontiers in cell and developmental biology, 2022. **9**: p. 790386-790386.
48. Atkinson, V., et al., *Improved pyrexia-related outcomes associated with an adapted pyrexia adverse event management algorithm in patients treated with adjuvant dabrafenib plus trametinib: Primary results of COMBI-APlus.* European Journal of Cancer, 2022. **163**: p. 79-87.
49. Mashima, Y., et al., *Using Natural Language Processing Techniques to Detect Adverse Events From Progress Notes Due to Chemotherapy.* Cancer Informatics, 2022. **21**: p. 11769351221085064.
50. Rolland, B., et al., *Toward Rigorous Data Harmonization in Cancer Epidemiology Research: One Approach.* Am J Epidemiol, 2015. **182**(12): p. 1033-8.
51. Zegers, M., et al., *Perceived burden due to registrations for quality monitoring and improvement in hospitals: a mixed methods study.* 2022.
52. Dinh-Le, C., et al., *Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions.* JMIR Mhealth Uhealth, 2019. **7**(9): p. e12861.
53. Tran, B.D., et al., *How does medical scribes' work inform development of speech-based clinical documentation technologies? A systematic review.* J Am Med Inform Assoc, 2020. **27**(5): p. 808-817.
54. Chan, K., et al., *Developing a framework to incorporate real-world evidence in cancer drug funding decisions: the Canadian Real-world Evidence for Value of Cancer Drugs (CanREValue) collaboration.* BMJ Open, 2020. **10**(1): p. e032884.
55. Lorgelly, P., et al., *Outcome-Based Payment Schemes: What Outcomes Do Patients with Cancer Value?* Patient, 2020. **13**(5): p. 599-610.
56. Kann, B.H., A. Hosny, and H. Aerts, *Artificial intelligence for clinical oncology.* Cancer Cell, 2021. **39**(7): p. 916-927.