



Universiteit
Leiden
The Netherlands

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

Laar, S.A. van

Citation

Laar, S. A. van. (2023, October 12). *Text mining real-world data to evaluate systemic anti-cancer therapy*. Retrieved from <https://hdl.handle.net/1887/3643700>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3643700>

Note: To cite this publication please use the final published version (if applicable).

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

Sylvia A. van Laar

The research presented in this thesis was performed at the department of Clinical Pharmacy and Toxicology of Leiden University Medical Center, Leiden, The Netherlands. Financial support for the publication of this thesis was provided by Afdelingsfonds Klinische Farmacie & Toxicologie.

Cover	Sylvia van Laar
Layout	Renate Siebes Proefschrift.nu
Printed by	Proefschriftmaken.nl De Bilt
ISBN	978-94-6469-561-8

© 2023 Sylvia van Laar

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronically or mechanically, including photocopying, recording or in any information storage or retrieval system, without prior permission of the author.

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

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 12 oktober 2023
klokke 13.45 uur

door

Sylvia Adriana van Laar

geboren te Den Haag
in 1993

Promotor

prof. dr. H.J. Guchelaar

Copromotores

dr. J. Zwaveling

dr. K.B. Gombert-Handoko

Promotiecommissie

prof. dr. A.J. Gelderblom

prof. dr. M.W.J.M. Wouters

prof. dr. P.H. van Erp, Radboudumc

prof. dr. M. Koopman, UMC Utrecht

Contents

Chapter 1	General introduction	7
------------------	----------------------	----------

PART I: Methods for real-world data collection

Chapter 2	Use of real-world data sources to evaluate cancer treatments: current practice and future perspectives <i>Review - Manuscript in preparation</i>	19
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 <i>Original research - Clin Pharmacol Ther. 2020; 108(3): 644-652</i>	39

PART II: Real-world treatment effectiveness and treatment patterns

Chapter 4	Real-world metastatic renal cell carcinoma treatment patterns and clinical outcomes in The Netherlands <i>Original research - Front Pharmacol. 2022; 13: 803935</i>	83
Chapter 5	Application of an electronic health record text mining tool: Treatment patterns and outcomes for Dutch patients with hepatocellular carcinoma <i>Original research - Manuscript in preparation</i>	101

PART III: Real-world treatment safety

Chapter 6	Application of electronic health record text mining: Real-world tolerability, safety, and efficacy of adjuvant melanoma treatments <i>Original research - Cancers. 2022; 14(21): 5426</i>	127
Chapter 7	Real-world evaluation of supportive care using an electronic health record text mining tool: G-CSF use in breast cancer patients <i>Original research - Support Care Cancer. 2022; 30: 9181-9189</i>	177
Chapter 8	Liver and kidney function in patients with Covid-19 treated with remdesivir <i>Original research - Br J Clin Pharmacol. 2021; 87(11): 4450-4454</i>	203

PART IV: General discussion, summary and appendices

Chapter 9	General discussion	215
Chapter 10	English summary	237
	Nederlandse samenvatting	242
Appendices	Curriculum Vitae	251
	List of publications	252
	Dankwoord	254

Chapter 1

General introduction

General introduction

In the Netherlands, every year, approximately 120,000 people are diagnosed with cancer, 600,000 live with cancer, and approximately 45,000 patients die from cancer making it the most common cause of death. Also, both the incidence and absolute cancer related mortality are still rising, since the aging population and higher obesity prevalence are putting people at significant risk for cancer. Though overall survival is improving [1]. Furthermore, the costs related to cancer care in the Netherlands are high and rising, and currently estimated at 6.5 billion euro's yearly [2]. Making the development and use of (cost-)effective treatments a continuously pressing topic.

A significant proportion of all cancer patients receives systemic cancer therapy with anti-cancer drugs either as curative or palliative treatment, however exact numbers are unknown. Since 1940, chemotherapeutic antitumor drugs have become available and are known for their toxicity to non-cancerous tissue and the occurrence of drug resistance. Targeted therapies are designed to target mutations in the pathways of specific tumor cell types and include for example, tyrosine kinase inhibitors (TKIs) and monoclonal antibodies (MABs) [3]. Approval of the first MAB, rituximab, was in 1997, and the first TKI, imatinib, followed in 2001 [3, 4]. Onwards, the MABs were directed towards T-cell protein receptors and tumor antigens downregulating the immune response, creating the immune checkpoint inhibitors (ICI), of which ipilimumab was the first in 2011 [3]. All these innovations in cancer treatments have resulted in 196 approved anticancer drugs by the European Medicines Agency (EMA) in November 2022 [5].

For decades, the randomized controlled trial (RCT) has been the basis of the marketing access of anticancer agents. RCTs can be characterized by their strict patient in- and exclusion criteria, and highly protocolized treatment regimens and rigid statistical analysis plans. As the blinding and randomization further reduce the chance of bias, this method has become the golden standard to investigate treatment efficacy and safety before granting marketing access [6-8]. However, RCTs have limited external validity, as the in- and exclusion criteria are strict, and therefore may result in a highly selected patient population, not representative for the patient population in clinical practice [9]. Furthermore, as the unmet medical need of patients with cancer remains, new treatments often receive faster marketing approval through the conditional marketing approval pathway of the EMA. These approvals are often initially based on surrogate

end-points as progression-free survival, recurrence-free survival or response rates, all not always good predictors for overall survival [10, 11]. Regarding treatment safety, rare or long-term safety outcomes may be undetected during the trial period. Furthermore, it is not uncommon that new anti-cancer drugs show moderate improvements of overall survival only, that disappear if patients in clinical practice differ from those included in RCTs. More concrete, 40% of oncologic drugs with an orphan designation authorized between 2000 and 2017 did not show any clinically relevant gain in overall survival [12]. And, even though the RCT is the golden standard, there is a trend that initial registration trials only have a single-arm and a non-randomized cohort design [13]. Therefore, results from the registration trials are not always representative for the effectiveness of anticancer treatments in clinical practice, thus a knowledge gap remains.

Real-world data (RWD) are defined as all observatory healthcare data not collected in conventional trials, and in contrary to trials, RWD are generated in daily practice [14, 15]. Even though RWD are not new, the interest in RWD to generate real-world evidence, increased significantly in recent years, since they comprehend the valuable information needed to complement trial results [16]. RWD have the potential to gain insights in patients, their characteristics, treatments and outcomes in clinical practice. All of these are important variables since both individual patient characteristics and specific treatment details can influence the overall outcomes. This is extremely relevant in the field of cancer treatments, as the treatments usually enter the market with limited evidence from clinical studies, but also since the number of treatment options are rapidly expanding and treatments come with high costs [17]. Therefore, being able to determine which patients might or might not benefit (most) from a specific treatment, and if treatments are really cost-effective in clinical practice is of high importance [18]. Therefore, continuously generated RWD have the potential to be used for the evaluation of new therapies and to establish whether efficacy depends on certain patient characteristics.

The electronic health record (EHR) is the hospitals' digital version of the once paper patient files. In the EHR a vast amount of RWD is stored. Initially, EHRs were used purely for billing purposes, but the function and functionalities have widely expanded and now it is used, e.g., to keep track of patients' health information, test results, decision supports, and communication with healthcare partners [19]. As a result, EHRs contain longitudinal data on patients, their diseases and treatments collected during

routine care, especially on cancer patients, which frequently visit the hospital [20]. One part of the data is stored structurally in the EHR, including the demographics, laboratory data, medication orders and vital signs of patients. However, 80% of the information is stored in unstructured data – free-text notes, including visit summaries and correspondence – and these notes contain valuable details and nuances on a patients' treatment trajectory [21, 22]. As these notes are unstructured, manual chart review is still the standard method of data extraction, but this is time-consuming, laborious, and error-prone [21, 23, 24]. Since, this method cannot be implemented on a large scale to systematically review cancer treatments, a more advanced data collection method is warranted.

Text mining is a general term for a wide range of techniques that are used for information extraction of data from files with unstructured text. Natural language processing is the most important one, since it can be implemented to parse, segment, extract, or analyze text data. The specific algorithms used can be divided into 1) rule-based algorithms that are manually comprised, based on knowledge of the field and the expected used terminology, and 2) statistical-based algorithms, where the algorithm is trained based on large datasets with machine learning [24-27]. Text mining has been implemented in many other fields to capture data from text, however, until now the use of text mining of EHRs has been limited [28, 29]. We hypothesize that text mining has a high potential to improve the efficiency of data capture from EHR, and therefore the collection of RWD on cancer treatments.

Aim and outline of this thesis

The general aim of this thesis is to investigate if a text-mining tool is suitable for the collection of real-world data from EHRs to evaluate cancer treatments in clinical practice. More specific, we aim to investigate, first, if it is possible to capture data on cancer treatments with the same accuracy as manual review by text mining the EHR. And second, if text mining can be applied to evaluate a variety of aspects of cancer treatments in clinical practice, including treatment patterns, effectiveness, adverse events and guideline adherence.

Part 1: Methods for real-world data collection

First, in **Chapter 2**, we will provide an overview of the current state of the use of real-world data. In this chapter all commonly used real-world data sources will be summarized, and we will reflect on the strengths and limitations of each source. Furthermore, we summarize the different aspects of oncologic treatments that can be reviewed with these data as the real-world effectiveness, adverse events, cost-effectiveness and treatment utilization patterns. In **Chapter 3** we will validate the text-mining tool CTcue for data collection from EHRs by performing a study in a population of renal cell carcinoma patients and compare these data with results of a manual data collection.

Part 2: Real-world treatment patterns and effectiveness

The focus of part 2 of this thesis will be on the application of text mining EHR to study treatment effectiveness and treatment patterns. In **Chapter 4** we will apply the validated tool from chapter 3 in an additional hospital and summarize the treatment patterns and the first-line effectiveness results of a metastatic renal cell carcinoma population. Next, based on the rule-based queries from chapter 4, the treatment patterns and effectiveness of population of patients with metastatic hepatocellular carcinoma will be studied in **Chapter 5**. Since hepatocellular carcinoma has many known potential prognostic factors, these are added in this study, to further investigate which patients are most likely to benefit from treatment.

Part 3: Real-world treatment safety

Part 3 includes three chapters regarding the safety of treatments in clinical practice. Compared to palliative treatments, the acceptance of adverse events is lower for adjuvant treatments in the curative setting. Therefore, we will investigate in **Chapter 6** the tolerability, safety, and preliminary efficacy of adjuvant melanoma treatments. Furthermore, some adverse events are severe but can be prevented with prophylactic treatment. An example is the occurrence of febrile neutropenia during certain types of breast cancer treatments. This can be prevented by use of granulocyte colony-stimulating factors (G-CSFs), which is indicated for a high-risk and intermediate risk group. In **Chapter 7** we will evaluate with text mining whether G-CSFs are applied according to the guideline. At last, we use the data-extraction method to fast discover the severity of adverse events of a potential treatment of Covid-19. In **Chapter 8** we react on a signal that the treatment of Covid-19 with remdesivir could result in renal- and liver damage and investigate the changes in renal- and liver function after treatment initiation.

Part IV: General discussion, summaries and appendix

This thesis will conclude with a general discussion on future perspectives in **Chapter 9**, and summaries in English and Dutch are provided in **Chapter 10**. Table 1.1 gives an overview of the investigated topics per chapter illustrating overlap and differences between studies.

Table 1.1. Overview of topics per chapter

Chapter	2	3	4	5	6	7	8
Type of study							
Review	•						
Original study		•	•	•	•	•	•
Study topics							
Method validation		•					
Treatment effectiveness		•	•	•	•		
Treatment safety		•			•	•	•
Treatment patterns			•	•			
Treatment tolerability					•		
Guideline adherence						•	
Prognostic factors				•		•	
Disease							
Renal cell carcinoma		•	•				
Hepatocellular carcinoma				•			
Melanoma					•		
Breast cancer						•	
Covid-19							•
Treatments							
Immune checkpoint inhibitors			•	•	•		
Tyrosine kinase inhibitors			•	•	•		
Chemotherapy						•	
Granulocyte-colony stimulating factors						•	
Antiviral treatment							•

References

1. NKR cijfers. 10 December 2022; Available from: <https://iknl.nl/nkr-cijfers>.
2. J. Praagman, E.S., L. van Disseldorp, V. Lemmens, *Kanker in Nederland - trends & prognoses tot en met 2032*. 2022.
3. Falzone, L., S. Salomone, and M. Libra, *Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium*. Frontiers in Pharmacology, 2018. **9**.
4. Cohen, P., D. Cross, and P.A. Jänne, *Kinase drug discovery 20 years after imatinib: progress and future directions*. Nature Reviews Drug Discovery, 2021. **20**(7): p. 551-569.
5. Pantziarka, P., et al., *An Open Access Database of Licensed Cancer Drugs*. Frontiers in Pharmacology, 2021. **12**.
6. 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.
7. Bothwell, L.E. and S.H. Podolsky, *The Emergence of the Randomized, Controlled Trial*. 2016. **375**(6): p. 501-504.
8. Verweij, J., et al., *Innovation in oncology clinical trial design*. Cancer Treatment Reviews, 2019. **74**: p. 15-20.
9. 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.
10. Pasalic, D., et al., *Progression-free survival is a suboptimal predictor for overall survival among metastatic solid tumour clinical trials*. Eur J Cancer, 2020. **136**: p. 176-185.
11. 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.
12. Schuller, Y., et al., *Oncologic orphan drugs approved in the EU – do clinical trial data correspond with real-world effectiveness?* Orphanet Journal of Rare Diseases, 2018. **13**(1): p. 214.
13. 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.
14. 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.
15. 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.
16. 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.
17. Vokinger, K.N., et al., *Prices and clinical benefit of cancer drugs in the USA and Europe: a cost-benefit analysis*. The Lancet Oncology, 2020. **21**(5): p. 664-670.
18. Wild, C.P., et al., *Cancer Prevention Europe*. Molecular Oncology, 2019. **13**(3): p. 528-534.
19. Kim, E., et al., *The Evolving Use of Electronic Health Records (EHR) for Research*. Seminars in Radiation Oncology, 2019. **29**(4): p. 354-361.
20. Cowie, M.R., et al., *Electronic health records to facilitate clinical research*. Clin Res Cardiol, 2017. **106**(1): p. 1-9.
21. 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.
22. Fessele, K.L., *The Rise of Big Data in Oncology*. Semin Oncol Nurs, 2018. **34**(2): p. 168-176.

23. 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.
24. 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.
25. 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.
26. 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.
27. Sun, W., et al., *Data Processing and Text Mining Technologies on Electronic Medical Records: A Review*. Journal of Healthcare Engineering, 2018. **2018**: p. 4302425.
28. Wang, Y., et al., *Clinical information extraction applications: A literature review*. J Biomed Inform, 2018. **77**: p. 34-49.
29. 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.

Part I

Methods for real-world data collection



stage III a

Fuhrman graad I

thyroxine

levertr

tumorprogressie

headache

borstkanker

Extrahepatisch

pazoparib

oestrogeen receptor

Hepatectomie

metastasis

koude rill

remdesiv

Chapter 2

Use of real-world data sources to evaluate cancer treatments: current practice and future perspectives

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

Manuscript in preparation

Abstract

Over the last two decades the interest in real-world data (RWD), specifically on cancer treatments, has significantly increased since they can bridge the knowledge gap in evidence not addressed in randomized clinical trials. All sources of RWD used in studies have their own strengths and limitations and therefore can be used to investigate various aspects of oncologic drug treatments in daily practice. This review first gives an overview of the currently available real-world sources used to evaluate cancer treatments in clinical practice. It includes sources as the case report form, electronic health records, administrative claims data, patient reported outcomes, cancer registries, mobile applications, wearable devices, and social media data and their strengths and limitations. Next, it elaborates on the potential applications of RWD, and summarizes how data from these sources can, in varying degrees, be used to estimate real-world effectiveness, monitor pharmacovigilance, calculate cost-effectiveness, describe general treatment utilization patterns, and support personalized treatment choice of cancer treatments. And, it gives a perspective of how RWD use can be optimized as the field of RWD on oncologic treatments is rapidly growing. The improvements include, but are not limited to, more harmonized and complete RWD sources, more connected or merged datasets, and more use of artificial intelligence in data processing. Furthermore, improved privacy legislation can even further enhance the insights generated with RWD. Concluding, a variety of RWD sources can be used to generate significant insights into cancer treatments in clinical practice. With further improvements RWD use can reach its full potential.

1. Introduction

Randomized controlled trials (RCTs) are the golden standard in determining the efficacy and safety of medical drug treatments due to its proven and rigorous study design, which makes RCTs crucial in market authorization for drug treatments [1]. However, RCTs also have potential weaknesses: by design, they are time and resource-intensive, and only include a selected patient group; they are often not designed for follow-up of long-term toxicity, frequently use surrogate parameters, and they have limited external validity [2, 3].

In general, all data regarding the effects of health interventions in patients not collected in RCTs are defined as real-world data (RWD) [4, 5]. RWD includes data primarily generated for studies in clinical practice, and the reuse or secondary use of data already generated for other (healthcare) purposes. As health-related information is increasingly stored digitally, these routinely generated data, e.g., electronic health records, administrative claims- and patient-derived data, can also be used as RWD sources [4, 6, 7]. Evidence derived from RWD, real-world evidence (RWE), in general, has greater external validity and generalizability RCTs. For example, by including larger populations, RWD can gain insight in treatment effectiveness and long-term and rare toxicity. Also, topics not included in trials can be studied, such as treatment patterns, effectivity in rare cancer tumours, and the effects in specific subgroups can be analysed [2]. In this way, RWE can add valuable information to the evidence which is obtained from RCTs [8].

In the last two decades an exponential growth is seen in publications with RWD and RWE on PubMed. It increased from 12 articles in 2002 to 3,310 articles in 2021, including opinion and perspective papers, original studies, and (systematic) reviews [9]. Furthermore, in 2021 around a quarter of the publications included the topic “cancer” (Figure 2.1). All sources of RWD used in studies have their own strengths and limitations and can be used to investigate various aspects of oncologic drug treatments in daily practice [10]. In addition, the increase in the number of studies results in a rapidly changing field. Therefore, the aim of this review is to give an overview of the current options for real-world sources used to evaluate cancer treatments in clinical practice, and how it can be optimized in the future.

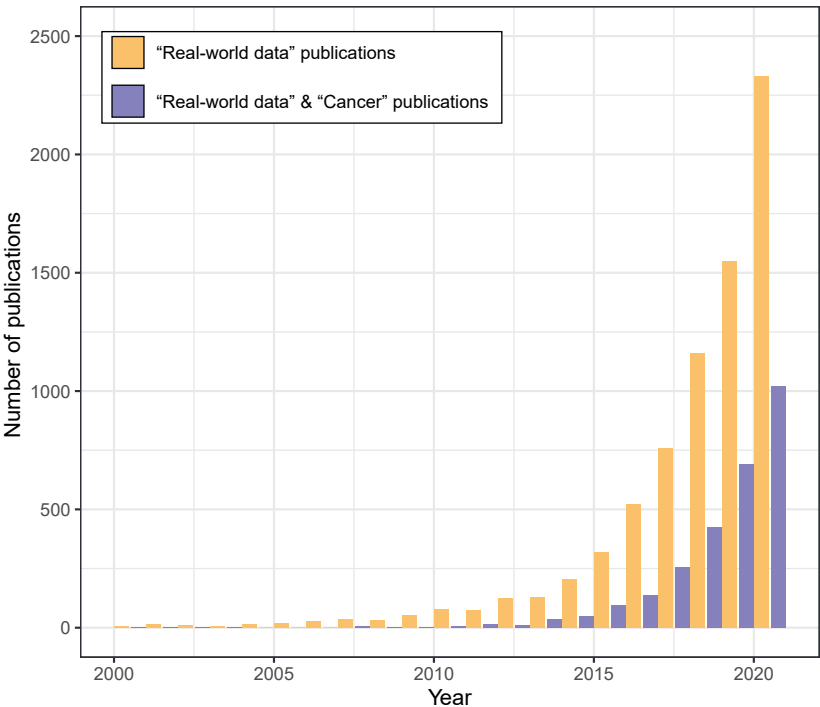


Figure 2.1. Number of publications on “Real-world data” and “Real-world data” combined with “cancer” per year on PubMed.

2. Sources

Multiple sources contain RWD on cancer treatments, and each source has its advantages and challenges regarding the time and effort it takes to collect the data, the quality of the data, the coverage of information and therefore the potential risk of bias. Table 2.1 shows a summary of basic characteristics and the strengths and limitations per source.

2.1 Case report form

The standard method to generate data prospectively, and primarily for study purposes, is by using case report forms (CRFs). Traditionally CRFs were paper forms, but they are now mostly digital (electronic CRFs [eCRFs]). Based on a study protocol, they are used to collect all study results, which is mandatory by the Good Clinical Practice guidelines for clinical trials [11, 12]. However, CRFs can also be used to collect data from clinical practice. The process of data collection, recording, monitoring, and auditing is done by staff, which makes the use of CRFs to collect data a time intensive and costly process [11].

Table 2.1. Strengths and limitations of real-world data sources

Real-world data source	Strengths	Limitations
	Case report forms - All essential data can be collected - High data quality	- Expensive - Laborious - Data is not readily available
	Electronic health records - Extensive and nuanced longitudinal patient, disease, and treatment data	- Often limited to one healthcare center - Complex data source - Risk of missing or incorrect data
	Administrative claims data - Longitudinal data across healthcare facilities - Insights in costs	- High variety in sensitivity - Limited to reimbursable healthcare
	Patient reported outcomes - Direct insight in patients' perception of treatments and quality of life - Multiple validated questionnaires available	- Use of PROMS is not yet standard of care - Patients' willingness and capacity to participate can bias outcomes
	Registries - Large, structured databases with patient, disease, and treatment data - Reflects treatments on population level	- Laborious - Data may lag current practice
	Wearable devices - High density longitudinal biometric data	- Data generation is limited to users - Still in development
	Social media data - Insight in patients' perception on treatments and disease	- Limited to social media users - Missing data on patient demographics

2.2 Electronic medical records

A vast amount of RWD is collected from electronic health records (EHRs). Primarily, EHRs were implemented for billing purposes, but has expanded to many core functions including keeping track of patients' health information, test results, decision support, and improvement of communication with care partners [13]. As a result, EHRs contain

longitudinal data on patients, their diseases and treatments collected during routine care, especially on cancer patients, which frequently visit the hospital [14]. It includes data on primary care-, hospital and emergency department visits, demographics, vital signs, laboratory data, medication orders, procedures, imaging, health behaviour, and free-text notes, e.g., summarizing encounters and correspondence [15]. These free-text notes can make up to 80% of the healthcare information, and contain valuable details and nuances on a patients' treatment trajectory [16]. EHR data has limitations when used for research purposes. Since accurate and correct documentation may be difficult in clinical practice, data may be missing, erroneous, uninterpretable, or inconsistent, in particular in free-text notes. Also, EHRs of a single center may not represent the complete medical history of a patient and a patient population [17, 18]. Furthermore, high variety in expressions in terminology – the linguistic complexity – and the level of detail or specificity used to describe a specific case can limit the accuracy of automatic data extraction [17-19].

2.3 Administrative claims data

Administrative claims data are generated for reimbursement purposes. However, since the billing codes from claims for insurance providers include coded information on diagnoses and healthcare services provided, they can also give insight in provided cancer treatments [20]. What makes claims data unique, is that they capture longitudinal data from individuals across healthcare facilities, including in- and outpatient visits, and dispensing data of prescription medication from outpatient or community pharmacies [20, 21]. Therefore, these data can be used to study specific elements on healthcare delivery such as the hospital volumes, healthcare providers, and prescribing patterns [20]. Claims data only contains reimbursable healthcare, therefore excludes, e.g., over-the-counter medicines. Furthermore, they are more likely to present more severe clinical conditions, i.e., conditions that require surgery, or persistent healthcare increase the accuracy of correct coding. Therefore, the sensitivity of specific outcomes in claims data varies [20, 21]. Furthermore, compared to other sources, claims data often lack clinical, laboratory and patient's behavioral data [20].

2.4 Patient reported outcome measures

Patient reported outcome measures (PROMs) are standardized questionnaires through which patients themselves directly share their experience and perception of a specific disease or treatment by reflecting e.g., on health-related quality of life, symptoms, and

treatment adherence and satisfaction. These questionnaires can be generic for quality of life, e.g., the EQ-5D [22], or disease specific, like the questionnaire for patients with cancer: EORTC-QLQ-C30 [23, 24]. Primarily, they were developed to assess the effectiveness of treatment in clinical trials [25, 26]. However, the applications have further broadened, by including them in clinical practice, improving the communication regarding and assessment and management of patients' symptoms [25]. Furthermore, these PROMs create the possibility to retrospectively review all patient reported experiences. However, due to the lack of time and knowledge and inadequate digital infrastructures clinicians have difficulty with successfully integrate, interpret and act upon the PROMs data, which must be taken into account when using PROMs as RWD source [27]. Furthermore, it must be taken into account that the willingness and capacity of patients to continuously participate with the questionnaires can bias results [28].

2.5 Cancer registries

In cancer registries data from all patients with a certain type of cancer and/or with a specific treatment in clinical practice is systematically and continuously stored without specific prospectively defined outcomes [20, 29-31]. They are often population-based on a regional or national level, but can also be health-care setting based (e.g., one hospital) [31]. In registries, in general, a wealth of information on patients' sociodemographic and clinical characteristics is collected, making them a unique source of data on all patients in a certain region and therefore ideal to retrospectively investigate, e.g., cancer incidence and treatment effectiveness in a specific geographical area, social group or time period [31, 32]. For the collection of data, registries are mostly dependent on EHR data from hospitals, which are usually manually entered [33], but also data from independent (pathology) labs, physician offices, administrative claims data, biomarker data, pharmacy claims, patient reported outcomes and general demographic data can be included [20]. Even though registries are large and structured databases, they lack uniformity. Therefore, not every registry will contain the same variables [33]. Also, information on disease progression or recurrence and adverse drug reactions is not always structurally collected [20]. And, as data collection from various sources is time-consuming, the results from registries can lag 1-2 years behind current practice [20, 34].

2.6 Mobile applications and wearable devices

Wearable devices are items worn around the body (e.g., around the wrist, upper arm, waist, or hip) that can monitor and objectively collect biometric data (e.g., heart rate, respiratory rate, oxygen saturation, physical activity, sleep patterns) [35, 36]. These devices also can be used for the evaluation of cancer treatments. Multiple recent review articles mention that these tools have the potential to investigate adverse events, treatment adherence, symptoms, practice of physical activity, and overall, the quality of life [35-37]. Limitations in the use of these wearable items include, firstly, that users must have a certain level of digital literacy, which may be lacking in the elder generations. Furthermore, the devices themselves must be worn and function without technical problems, in order to generate data. Therefore the consumer-grade tools have to be validated and monitored [36]. Also, consensus still must be found on the terminology regarding wearable devices, in order to increase uniformity of the outcome measures in studies [37].

2.7 Social media data

Next to all data collected in and around the healthcare setting, social media use has significantly increased. This concurrent expansion of generated digital information has the potential to be used for cancer treatment evaluation. A review of Kalf et al. [38] showed that data collection from social media sources as forum topics and discussions, Facebook pages of patient communities or support groups and Twitter conversations can be a potential source to study relative effectiveness, adverse events, symptoms, quality of life, and adherence behaviour of patients with cancer and their treatments. Social media listening can also be used to give insight in patients experiences with (new) treatments [39]. Social media data can be biased due to the specific user profile, e.g. younger and female patients, and post may be duplicated, topics can be recurrent, and reporting on social media can also induce the reporting from patients with similar events [38]. Also, the amount of demographic or clinical information of patients on social media is limited, e.g., regarding patients' age, sex, or specific disease characteristics [39].

3. Applications of RWD in the evaluation of drug treatments

All sources of RWD can be used for specific research purposes in various phases of the life cycle of a drug – from drug discovery to post-registration. Multiple RWD applications have been defined and categorized.

3.1 Effectiveness

In real-world studies the treatment outcomes, and more specifically the effectiveness of a treatment in a specific population, are studied frequently for new cancer treatments post-marketing. Thereby the effectiveness of treatments under different circumstances than the trial and trial population can be studied, e.g., elderly, patients with comorbidities or limited kidney-, liver-, or cardiac function. But also, for the investigation of long-term efficacy outcomes, RWD are relevant, since in RCTs the outcome of a treatment is mostly studied during a shorter period [40]. For an accurate determination of the treatment effectiveness in clinical practice, insight in patient characteristics, previous treatments a patient received and the accompanying outcomes as recurrence, progression, or mortality are necessary. Electronic health records data, especially when merged into a registry or other larger databases, are suitable for this purpose, when the caveats as inaccurate or incomplete data are considered [17, 41]. Also claims data are used for effectiveness studies, mostly regarding the overall survival after treatment initiation [42]. The growing quantity of effectiveness studies even enables systematically reviewing specific outcomes [43].

3.2 Pharmacovigilance

The monitoring of adverse events is continued after marketing authorization, which is an important part of pharmacovigilance. Indeed, not all adverse events (AEs) are identified in RCTs; especially rare and late AEs do often not surface during this phase [44]. Especially for oncologic treatments pharmacovigilance is important, as almost all treatments are accompanied with serious toxicity. New oncologic treatments frequently enter the market before their safety profile is complete, and more often fast track routes are applied to accelerate market entry [45]. Furthermore, new targeted treatments and immunotherapies come with their own unique, complex safety profiles [46]. AEs can arise as a signal from multiple sources indicating a potential causal association between a treatment and the specific AE. Spontaneous reporting systems, in which a reporter, e.g., physician, reports their suspicion of an adverse event, have been the cornerstone in the pharmacovigilance process. However, the disadvantages of this system are, in general, a combination of underreporting, overreporting of publicized events, biased reporting, incomplete patient information, and missing of a denominator to estimate the frequency of this AE [44, 46, 47]. Therefore, in the field of pharmacovigilance, there is increasing interest in various sources of RWD. Reviews of Crestan et al. [46] and Lavertu et al. [47] emphasize the potential use of big data collections as EHR,

claims, and social media data. And both groups underline the potential application of innovative methods for the synthesis and analysis of data, as in general the use of artificial intelligence, including but not limited to, data mining, cognitive computing, and natural language processing. Furthermore, for more insights in the patients' perspective on adverse events, PROMs are a relevant RWD source, as observed in the study of Stormoen et al., who investigated the burden of adverse events in a population of prostate cancer patients receiving medical treatment [48].

3.3 Cost-effectiveness

Demonstrating cost-effectiveness is especially relevant for cancer treatments, since new treatments often come with high prices even though the uncertainty on the added clinical value [49, 50]. RWD already has already extensively be used in the field of cancer treatments for cost-effectiveness. For example, almost all (96%) single technology appraisals of oncologic drug treatment performed by the National Institute for Health and Care Excellence of the United Kingdom contained some RWD in the cost-effectiveness analysis [51]. Moreover, insight into the clinical value of treatments with RWD can substantiate more accurate economic evaluations, and therefore influence reimbursement decisions and price negotiations [52]. Key components of the estimation are survival outcomes (discussed in “3.1 effectiveness”), quality of life, and costs [53]. The quality of life can best be evaluated by using patient reported outcomes, and increasing the use of PROMs will improve these estimations, as currently the quality of life is not often systematically recorded [54]. Furthermore, administrative claims data are the prime source for costs data [55]. Furthermore, detailed information as actual number of used vials of a specific treatment can be extracted from EHR, as is done in the real-world cost-effectiveness analysis of Van Kampen et al. [56].

3.4 Indication expansion

Often indications are expanded later in the drug treatment life-cycle [45]. Real-world evidence on the off-label use reflecting on the safety and efficacy of a treatment can be useful in this process. To illustrate, between 1 January 2018 until 31 December 2019, 12 of the 78 applications for the extensions of indications regarding antineoplastic and immunomodulating agents submitted at the European Medicines agency included RWE, mainly derived from registries and EHR data [57]. One example is the expansion of the application of palbociclib in men with hormone receptor-positive/human epidermal growth factor receptor 2-negative metastatic breast cancer in addition to the

registration for women. This was based on the study of Kraus et al. [58], who showed by reviewing EHRs, medical claims and prescription data, that palbociclib use in men was safe and consistent with the treatment of women.

3.5 General treatment prescribing and utilization patterns

How medication is used in clinical practice, can only be studied by analysing RWD. Treatment-related aspects include the specific choice of treatment, the dose, and the accompanying order in lines of treatment, reflected in treatment pattern studies in certain populations and moments [40]. These studies give insight in (shifts in) preference, and the adherence to treatment guidelines. Also, the time on treatment, treatment compliance and dose (reductions) over time add important understandings of how treatments are used and tolerated in clinical practice. For example, the review of Waser et al. [59] summarized twenty treatment pattern studies in non-metastatic non-small-cell lung cancer patients, to conclude that the patterns of increasing use of (neo)adjuvant treatment with increasing stage were, in general, in line with guidelines, and these studies mainly used data from registries, claims and medical records. Furthermore, Gan et al. [60] showed maintained effectiveness of cabozantinib in first- to fourth line treatment use, and associated toxicity-related dose reductions with improved outcomes based on a database with hospital and pharmacy records data. These studies also show that for comprehensive insights in patterns, larger datasets are needed, and therefore claims data, or larger registries containing EHR data are suitable sources.

3.6 Personalized treatment choice

Next to drug utilization in the general population, the characterisation of specific patient populations receiving treatments is relevant to interpret use and effectiveness of treatments of real-world cancer patients. It is well-known that trial populations are not always representative due to strict eligibility criteria and underrepresentation of certain demographics, e.g., related to race or age [61, 62]. Besides, disease related factors including biomarkers and non-related factors as a patients' physical state, genetics, comorbidities, comedication can influence outcomes [1, 63]. Detailed RWD, especially from health records, can be used to generate better insight in the patient characteristics and the accompanying outcomes can potentially help in identifying what treatment a specific patient might benefit the most from. Ideally, RWD is used to estimate the influence of all these individual factors on specific treatments and outcomes for clinical decision support in therapeutic decision making [64]. Machine learning algorithms

can potentially be the key for the pattern recognition in the growing amount of data, however due to the complexity of the data, this is not yet practice [65].

4. Future perspectives

The amount of available RWD of patients with cancer is constantly growing and can be found in several data sources. These RWD have the potential to replace or add data to the evidence generated by conventional research methods by being able to generate datasets that are quicker, cheaper, or larger, and generate data that were previously not available. This manuscript highlights the most common sources used to generate insights in oncologic treatment effects after registration, and the aspects of a treatment of which insight can be improved. It is shown that not every data source is suitable for all research questions, but a combination of sources can generate a wide range of data of oncologic patients and their treatments in clinical practice. However, improvements are needed before the use of RWD will reach its full potential.

4.1 Optimising source data

In the basis, the quality of the data is essential for collecting meaningful RWD which is suitable for analysis. However, a repeating disadvantage of all sources is that the primary purpose of the data capture was not research, and therefore the data quality can be limited. Maissenhaelter et al. [2] summarized the challenges in data quality in data completeness, data structure, data accuracy and the challenge of novel types of data. Part of the solution can be improving and harmonizing the data capture by generating more user-friendly, but clear front-ends of the used software, for example of EHRs. However, it will remain impossible to capture all patient data, therefore, minimum standards have to be established and harmonized over all used software programs [66]. Also, data entry can in the future potentially be replaced with new techniques as automatic speech-recognition based clinical documentation. However, these techniques are still in development [67]. Also, with more uniform data sources, combination of data and the implementation of application of smart analysis (4.3) will be made easier, as currently most data is siloed in their own systems [65]. However, only improving the data input will not solve all challenges with RWD.

4.2 Integration of sources and systems

The next challenge lies in the better integration of the sources and systems. This problem can be split into two parts, namely that that data from the same source type can be easily merged, and that different source types can be combined. If data sources are more harmonized, they can more easily be combined in joint data warehouses, and thereby the pool of RWD will be enlarged [2]. Subsequently, these large pools of RWD enable larger population studies. Registries are already a form of merged datasets, however, by harmonizing their sources, data collection can potentially be further automated. Furthermore, for example, integration of data from wearable devices into the EHR can add extra perspective, not only for research purposes, but also for healthcare practice. Already some devices have been piloted or implemented in EHR systems. Until now, systems are not always interoperable or cannot connect with wearable devices. In addition, EHR systems have to be ready to store the amount of data that wearables generate [68].

4.3 Artificial intelligence for smart analysis

Next to the efforts to optimizing and combining sources, the method of handling data also can be improved, especially with artificial intelligence. Data language processing techniques have the potential to process and extract information from large amounts of unstructured data, such as captured in the EHR and social media. To transfer EHR data into a form fit for retrospective analysis, manual extraction has been the standard method [69]. However, as this method is time-consuming, it is not very scalable [70]. With use of text mining, including natural language processing (NLP) strategies, the automatization of data extraction from EHR is currently in development [71, 72]. This technique can be divided into rule-based NLP, a system based on a coded set of rules for data extraction, and machine-learning-based NLP, a system based on a statistical learning algorithm that identifies the relevant context based on annotated training data [72]. Automated techniques, including machine learning algorithms, natural language processing software with sentiment analysis, are already used to analyze large social media datasets [38, 39, 73]. And with this, social media studies may develop into a fast method to collect patient experiences and add information to PROs [38]. The challenge in the field of NLP is generating an algorithm that is specific and precise enough to fit the heterogenous sources by rule-based or machine learning [74]. Furthermore, when datasets have sufficient quality and quantity, ultimately, AI can be used for pattern recognition in the rich and complex data to evaluate specific cancer treatments [65].

4.4 Privacy

With these constantly growing sources of patient data, it remains fundamental to protect the privacy of the patients. In the European Union (EU) patient data is protected by the General Data Protection Regulation (GDPR). And, in line, performing research with RWD should always be accompanied with the appropriate consents, anonymization or pseudonymisation of the data and data handling should only be performed by trained and qualified personnel [2]. However, the current privacy- and data regulations also hinder optimal reuse of these valuable real-world data, due to, e.g., hurdles on gaining access to data [75]. Therefore, the individual right of data privacy competes with the populations' right on (cancer) treatment improvements that could be enabled with these RWD [2]. New types of consent from patients, for example, the proposal on data altruism, in the proposed Data Governance Act of the EU, in which patients voluntarily make their data available for scientific and public purposes, could be an outcome for this barrier [75, 76].

5. Conclusion

The amount of RWD available on oncologic patients and their disease and treatment trajectory has significantly grown. This manuscript showed a non-exhaustive overview of the current practice and summarized RWD sources as traditional claims data, EHRs and more recently emerging fields as social media and wearable devices. Data from these sources are already used to gain insight into a wide field of oncologic treatment aspects in daily practice, including, effectiveness, pharmacovigilance, and prescribing patterns and can add relevant data for indication expansion, personalized treatment choice and cost-effectiveness. We conclude that the field of RWD is rapidly developing, and more harmonized and complete sources, smarter connections between data sources, use of artificial intelligence on several levels, and improvement of patient privacy legislation can even further enhance the insights generated with RWD.

References

1. Phillips, C.M., et al., *Assessing the efficacy-effectiveness gap for cancer therapies: A comparison of overall survival and toxicity between clinical trial and population-based, real-world data for contemporary parenteral cancer therapeutics*. Cancer, 2020. **126**(8): p. 1717-1726.
2. Maissenhaelter, B.E., A.L. Woolmore, and P.M. Schlag, *Real-world evidence research based on big data: Motivation-challenges-success factors*. Onkologe (Berl), 2018. **24**(Suppl 2): p. 91-98.
3. 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 Intern Med, 2019. **179**(7): p. 915-921.
4. Makady, A., et al., *What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews*. Value Health, 2017. **20**(7): p. 858-865.
5. Ramamoorthy, A. and S.M. Huang, *What Does It Take to Transform Real-World Data Into Real-World Evidence?* Clin Pharmacol Ther, 2019. **106**(1): p. 10-18.
6. Pisaniello, H.L. and W.G. Dixon, *What does digitalization hold for the creation of real-world evidence?* Rheumatology (Oxford), 2020. **59**(1): p. 39-45.
7. de Lusignan, S. and L. Crawford, *Creating and using real-world evidence to answer questions about clinical effectiveness*. J Innov Health Inform., 2015. **22**(3): p. 368-373.
8. Feinberg, B.A., et al., *Use of Real-World Evidence to Support FDA Approval of Oncology Drugs*. Value in Health, 2020. **23**(10): p. 1358-1365.
9. *PubMed.gov*. 2 February 2022]; Available from: <https://pubmed.ncbi.nlm.nih.gov/?term=real-world+evidence&filter=pubt.review>.
10. Simon, G.E., et al., *When Can We Trust Real-World Data To Evaluate New Medical Treatments?* Clin Pharmacol Ther, 2022. **111**(1): p. 24-29.
11. O'Leary, E., et al., *Data collection in cancer clinical trials: Too much of a good thing?* Clin Trials, 2013. **10**(4): p. 624-32.
12. Bellary, S., B. Krishnankutty, and M.S. Latha, *Basics of case report form designing in clinical research*. Perspect Clin Res, 2014. **5**(4): p. 159-66.
13. Kim, E., et al., *The Evolving Use of Electronic Health Records (EHR) for Research*. Seminars in Radiation Oncology, 2019. **29**(4): p. 354-361.
14. Cowie, M.R., et al., *Electronic health records to facilitate clinical research*. Clin Res Cardiol, 2017. **106**(1): p. 1-9.
15. 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.
16. Fessele, K.L., *The Rise of Big Data in Oncology*. Semin Oncol Nurs, 2018. **34**(2): p. 168-176.
17. Hersh, W.R., et al., *Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research*. Medical Care, 2013. **51**.
18. 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.
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 Medical Informatics and Decision Making, 2019. **19**(3): p. 75.
20. Penberthy, L.T., et al., *An overview of real-world data sources for oncology and considerations for research*. CA Cancer J Clin, 2022. **72**(3): p. 287-300.

21. Gross, M.D., B. Al Hussein Al Awamlh, and J.C. Hu, *Assessing Treatment-Related Toxicity Using Administrative Data, Patient-Reported Outcomes, or Physician-Graded Toxicity: Where Is the Truth?* Seminars in Radiation Oncology, 2019. **29**(4): p. 333-337.
22. Herdman, M., et al., *Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L)*. Quality of Life Research, 2011. **20**(10): p. 1727-1736.
23. Aaronson, N.K., et al., *The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology*. J Natl Cancer Inst, 1993. **85**(5): p. 365-76.
24. Nolte, S., et al., *General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States*. Eur J Cancer, 2019. **107**: p. 153-163.
25. Pérez-Alfonso, K.E. and V. Sánchez-Martínez. *Electronic patient-reported outcome measures evaluating cancer symptoms: A systematic review*. in *Seminars in Oncology Nursing*. 2021. Elsevier.
26. Churrua, K., et al., *Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues*. Health Expectations, 2021. **24**(4): p. 1015-1024.
27. Nguyen, H., et al., *A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs) in routine cancer care*. Journal of Medical Radiation Sciences, 2021. **68**(2): p. 186-195.
28. Zini, M.L.L. and G. Banfi, *A Narrative Literature Review of Bias in Collecting Patient Reported Outcomes Measures (PROMs)*. International Journal of Environmental Research and Public Health, 2021. **18**(23): p. 12445.
29. Garrison, L.P., et al., *Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report*. Value in Health, 2007. **10**(5): p. 326-335.
30. Skovlund, E., H.G.M. Leufkens, and J.F. Smyth, *The use of real-world data in cancer drug development*. European Journal of Cancer, 2018. **101**: p. 69-76.
31. Thong, M.S.Y., et al., *Population-based cancer registries for quality-of-life research*. Cancer, 2013. **119**(S11): p. 2109-2123.
32. Forsea, A.M., *Cancer registries in Europe-going forward is the only option*. Ecancermedalscience, 2016. **10**: p. 641.
33. Hoeijmakers, F., et al., *National quality registries: how to improve the quality of data?* J Thorac Dis, 2018. **10**(Suppl 29): p. S3490-s3499.
34. Parums, D.V., *Editorial: Registries and Population Databases in Clinical Research and Practice*. Med Sci Monit, 2021. **27**: p. e933554.
35. Albino de Queiroz, D., et al., *Internet of Things in active cancer Treatment: A systematic review*. J Biomed Inform, 2021. **118**: p. 103814.
36. Low, C.A., *Harnessing consumer smartphone and wearable sensors for clinical cancer research*. NPJ Digit Med, 2020. **3**: p. 140.
37. Beauchamp, U.L., H. Pappot, and C. Holländer-Mieritz, *The Use of Wearables in Clinical Trials During Cancer Treatment: Systematic Review*. JMIR Mhealth Uhealth, 2020. **8**(11): p. e22006.
38. Kalf, R., et al., *Use of Social Media in the Assessment of Relative Effectiveness: Explorative Review With Examples From Oncology*. JMIR Cancer, 2018. **4**: p. e11.
39. Mazza, M., et al., *Social Media Listening to Understand the Lived Experience of Individuals in Europe With Metastatic Breast Cancer: A Systematic Search and Content Analysis Study*. Frontiers in Oncology, 2022. **12**.

40. Pulini, A.A., et al., *Impact of Real-World Data on Market Authorization, Reimbursement Decision & Price Negotiation*. Ther Innov Regul Sci, 2021. **55**(1): p. 228-238.
41. Kumar, A., et al., *Evaluation of the Use of Cancer Registry Data for Comparative Effectiveness Research*. JAMA Network Open, 2020. **3**(7): p. e2011985-e2011985.
42. Luyendijk, M., et al., *Assessment of Studies Evaluating Incremental Costs, Effectiveness, or Cost-Effectiveness of Systemic Therapies in Breast Cancer Based on Claims Data: A Systematic Review*. Value in Health, 2020. **23**(11): p. 1497-1508.
43. Johal, S., et al., *Real-world treatment patterns and outcomes in small-cell lung cancer: a systematic literature review*. J Thorac Dis, 2021. **13**(6): p. 3692-3707.
44. Patadia, V.K., et al., *Using real-world healthcare data for pharmacovigilance signal detection – the experience of the EU-ADR project*. Expert Review of Clinical Pharmacology, 2015. **8**(1): p. 95-102.
45. Michaeli, D.T., et al., *Initial and supplementary indication approval of new targeted cancer drugs by the FDA, EMA, Health Canada, and TGA*. Invest New Drugs, 2022. **40**(4): p. 798-809.
46. Crestan, D., et al., *Pharmacovigilance of anti-cancer medicines: opportunities and challenges*. Expert Opin Drug Saf, 2020. **19**(7): p. 849-860.
47. 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.
48. Stormoen, D.R., et al., *Patient reported outcomes interfering with daily activities in prostate cancer patients receiving antineoplastic treatment*. Acta Oncologica, 2021. **60**(4): p. 419-425.
49. Shin, G., H.Y. Kwon, and S. Bae, *For Whom the Price Escalates: High Price and Uncertain Value of Cancer Drugs*. Int J Environ Res Public Health, 2022. **19**(7).
50. Vokinger, K.N., et al., *Prices and clinical benefit of cancer drugs in the USA and Europe: a cost-benefit analysis*. Lancet Oncol, 2020. **21**(5): p. 664-670.
51. Bullement, A., et al., *Real-world evidence use in assessments of cancer drugs by NICE*. International Journal of Technology Assessment in Health Care, 2020. **36**(4): p. 388-394.
52. Guggenbickler, A.M., et al. *Rapid Review of Real-World Cost-Effectiveness Analyses of Cancer Interventions in Canada*. Current Oncology, 2022. **29**, 7285-7304 DOI: 10.3390/currencol29100574.
53. Hall, P.S., *Real-world data for efficient health technology assessment*. European Journal of Cancer, 2017. **79**: p. 235-237.
54. Parody-Rúa, E., et al., *Economic Evaluations Informed Exclusively by Real World Data: A Systematic Review*. International Journal of Environmental Research and Public Health, 2020. **17**(4): p. 1171.
55. Shih, Y.T. and L. Liu, *Use of Claims Data for Cost and Cost-Effectiveness Research*. Semin Radiat Oncol, 2019. **29**(4): p. 348-353.
56. van Kampen, R.J.W., et al., *Real-world and trial-based cost-effectiveness analysis of bevacizumab in HER2-negative metastatic breast cancer patients: a study of the Southeast Netherlands Breast Cancer Consortium*. European Journal of Cancer, 2017. **79**: p. 238-246.
57. Flynn, R., et al., *Marketing Authorization Applications Made to the European Medicines Agency in 2018–2019: What was the Contribution of Real-World Evidence?* Clinical Pharmacology & Therapeutics, 2022. **111**(1): p. 90-97.
58. Kraus, A.L., et al., *Real-World Data of Palbociclib in Combination With Endocrine Therapy for the Treatment of Metastatic Breast Cancer in Men*. Clinical Pharmacology & Therapeutics, 2022. **111**(1): p. 302-309.
59. Waser, N., et al., *Real-world treatment patterns in resectable (stages I–III) non-small-cell lung cancer: a systematic literature review*. Future Oncology, 2022. **18**(12): p. 1519-1530.

60. Gan, C.L., et al., *Cabozantinib real-world effectiveness in the first-through fourth-line settings for the treatment of metastatic renal cell carcinoma: Results from the International Metastatic Renal Cell Carcinoma Database Consortium*. *Cancer Med*, 2021. **10**(4): p. 1212-1221.
61. Liu, R., et al., *Evaluating eligibility criteria of oncology trials using real-world data and AI*. *Nature*, 2021. **592**(7855): p. 629-633.
62. Jayakrishnan, T., et al., *Landmark Cancer Clinical Trials and Real-World Patient Populations: Examining Race and Age Reporting*. *Cancers*, 2021. **13**(22): p. 5770.
63. Sarfati, D., B. Koczwara, and C. Jackson, *The impact of comorbidity on cancer and its treatment*. *CA: A Cancer Journal for Clinicians*, 2016. **66**(4): p. 337-350.
64. Beauchemin, M., et al., *Clinical decision support for therapeutic decision-making in cancer: A systematic review*. *Int J Med Inform*, 2019. **130**: p. 103940.
65. Kann, B.H., A. Hosny, and H.J.W.L. Aerts, *Artificial intelligence for clinical oncology*. *Cancer Cell*, 2021. **39**(7): p. 916-927.
66. Rolland, B., et al., *Toward Rigorous Data Harmonization in Cancer Epidemiology Research: One Approach*. *Am J Epidemiol*, 2015. **182**(12): p. 1033-8.
67. 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.
68. 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.
69. Post, A.R., Z. Burningham, and A.S. Halwani, *Electronic Health Record Data in Cancer Learning Health Systems: Challenges and Opportunities*. *JCO Clinical Cancer Informatics*, 2022(6): p. e2100158.
70. Chen, Y., et al., *Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system*. *BMC Medical Research Methodology*, 2022. **22**(1): p. 136.
71. Harpaz, R., et al., *Text Mining for Adverse Drug Events: the Promise, Challenges, and State of the Art*. *Drug Safety*, 2014. **37**(10): p. 777-790.
72. Percha, B., *Modern Clinical Text Mining: A Guide and Review*. *Annu Rev Biomed Data Sci*, 2021. **4**: p. 165-187.
73. Zhang, L., M. Hall, and D. Bastola, *Utilizing Twitter data for analysis of chemotherapy*. *Int J Med Inform*, 2018. **120**: p. 92-100.
74. Sheikhalishahi, S., et al., *Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review*. *JMIR Med Inform*, 2019. **7**(2): p. e12239.
75. Shabani, M., *Will the European Health Data Space change data sharing rules?* *Science*, 2022. **375**(6587): p. 1357-1359.
76. Horgan, D., et al. *European Health Data Space—An Opportunity Now to Grasp the Future of Data-Driven Healthcare*. *Healthcare*, 2022. **10**, DOI: 10.3390/healthcare10091629.

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

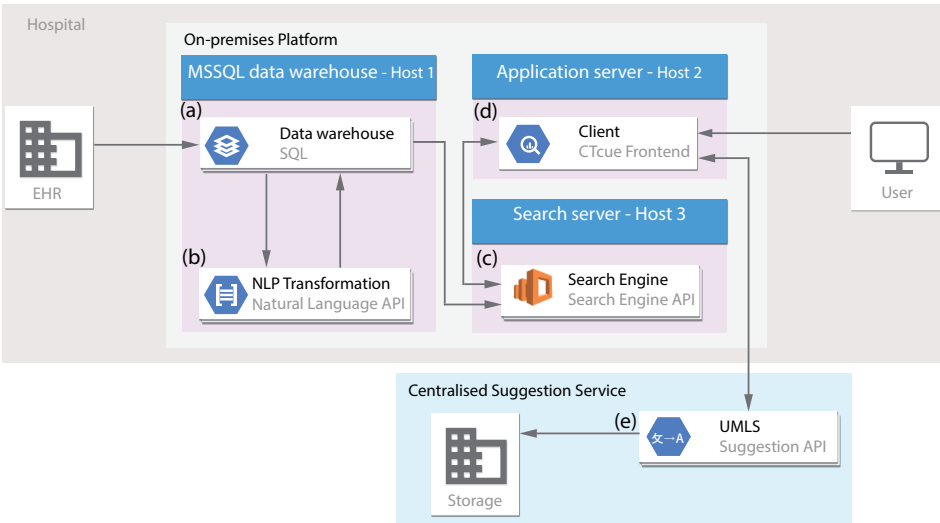


Figure 3.1. Architecture of the Clinical Data Collector on-premises isolation platform. **(a)** Copy of electronic health record (EHR) data transferred, stored, and cleaned in a local MS SQL Server relational database. **(b)** Natural language processing (NLP) transformation application programming interface (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.

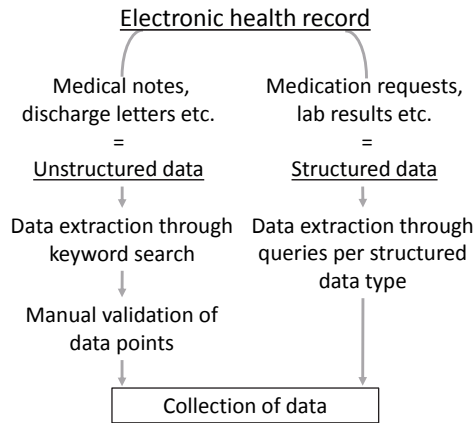


Figure 3.2. Data extraction approach from structured and unstructured data using Clinical Data Collector.

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.

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

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

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

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.

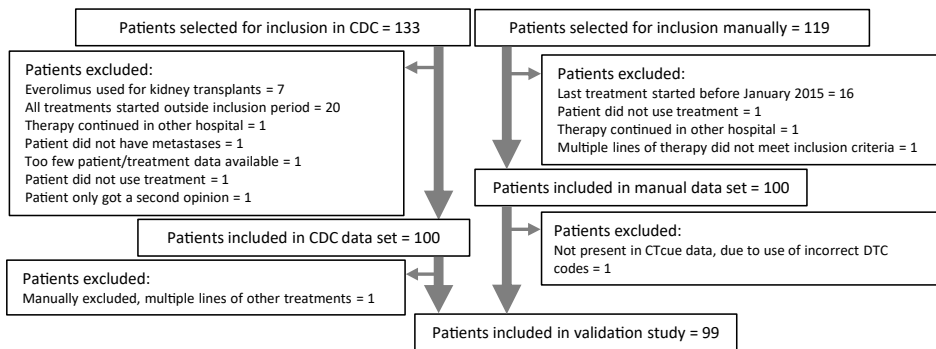


Figure 3.3. Flowchart of patient inclusion.

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/1.73m². 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.

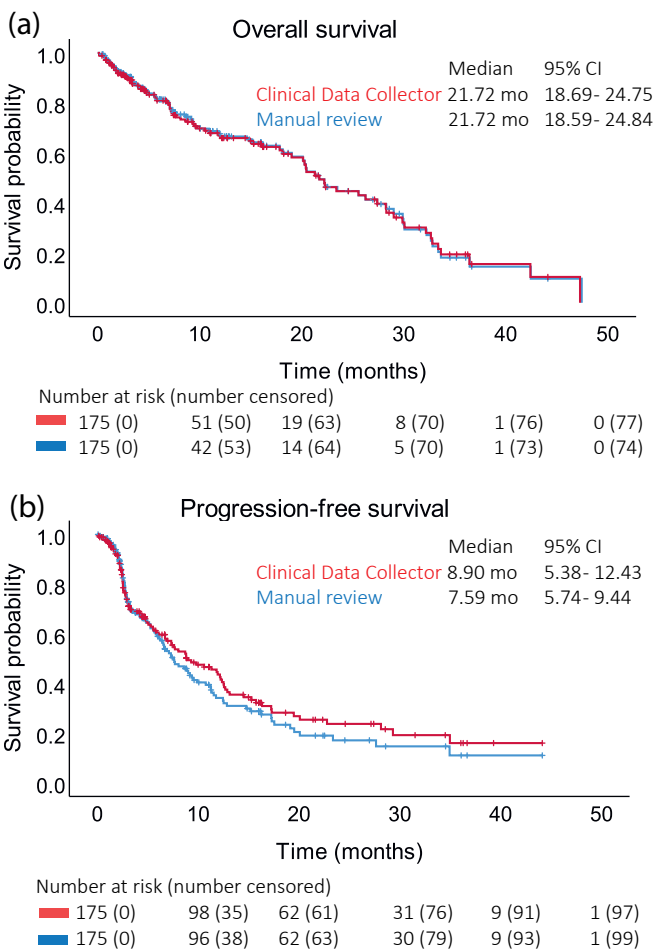


Figure 3.4. Kaplan-Meier plots of overall survival (a) and progression-free survival (b) determined from manual review and Clinical Data Collector data for cabozantinib, everolimus, nivolumab, pazopanib and sunitinib combined.

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

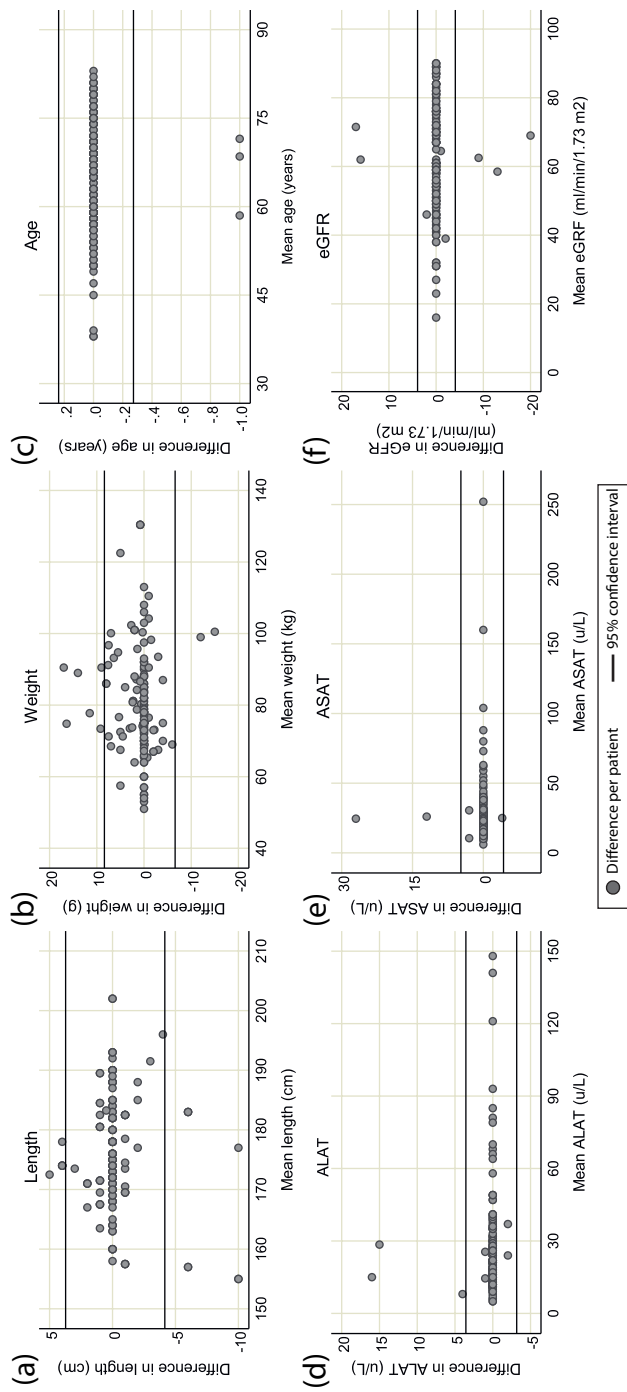


Figure 3.5. Bland-Altman plots of continuous variables collected using CDC versus manual with mean difference and 95% confidence interval. **(a)** Length: -0.21 cm (-4.2 to 4.8), **(b)** weight: 1.1 kg (-6.6 to 8.7), **(c)** Age: -0.17 years (-5.3 to 5.8), **(d)** Estimated glomerular filtration rate (eGFR) 0.22 ml/min/1.73m² (-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 reimbursements 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 (chromophili) renal cell carcinoma Manually added: papillair adenocarcinoom papillair niercarcinoom	Pathology report, radiology report, poli clinical letter or clinical letter		
	Chromofobic	Keyword search	Manually added: Chromofob niercelcarcinoom Chromofob adenocarcinoom Chromofob 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 occurring 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 gradet	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	Length 100 cm - 250 cm			
		Keyword search	Length 100 cm - 250 cm lengthe	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: progressive PD	Radiology report	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				
	Side effects †	Keyword search	Manually added: Stop stoppen staken	Form: consult and Specialism: medical oncology		
	Death †		And per treatment: pazopanib pazo sunitinib suni cabozantinib cabo everolimus eve nivolumab nivo			
	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 occurring 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.	
			From keyword synonym library: Tumorprogressie tumor progressie progressie tumor progressie van tumor progressie van tumoren neoplasma progressie Tumor progression				
			Keyword search				
			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	Measurement 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-EPIegfr 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
Cardiovascular comorbidity	Cardiovascular disease	Keyword search	Manually added: Hypertensie	Within form: consult; question: Anamnese		
			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 occurring 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 afb 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 insufficiëntie cardiaal insufficiëntie 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: Anamnese		* (informal) abbreviation # alternative notation, alternative spelling or frequently occurring typing error STEMI: ST elevation myocardial infarction

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

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 occurring typing error
COPD as comorbidity	COPD	Medication request	Parasympatricolytica sympatricomymetica+parasympatricolytica	Within form: consult; question: Anamnese	Start date should be from 1 year before until the same day as "start date treatment"	* (informal) abbreviation # alternative notation, alternative spelling or frequently occurring 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		Within form: consult; question: Anamnese	
	Keyword search	From keyword synonym library: Chronisch obstructieve longaandoening chronisch obstructieve luchtwegaandoening COPD			
Hypertension as adverse event	Hypertension	Medication request	Antihypertensiva Overige antihypertensiva ACE-remmers Angiotensine-ii-antagonisten Calciumantagonisten Selective 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 occurring 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:	From specialism: medical oncology		
			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			
		Keyword search	From keyword synonym library:	Within form: consult; question: anamnese (anamnesis), lichamelijk onderzoek (physical examination), aanvullend onderzoek (additional examination) of specialism: medical oncology		
			hypertensie hoge bloeddruk verhoogde bloeddruk manually added: PRES			

Liver toxicity as adverse event	Liver toxicity	Keyword search	Manually added:	Exclusion: polliclinical letters from specialist: 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 occurring typing error
			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			

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 occurring typing error
		Keyword search	From keyword synonym library: diarree diarre diaree diarhee 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 erythrodysesthesia syndrome palmar-plantar erythodysthesia palmoplantair erythrodysesthesiesyndroom palmoplantaire erythrodysesthesiesyndroom Hand-foot syndrome Hand and foot syndrome secondary to chemotherapy Hand- en voetsyndroom secundair aan chemotherapie Chemotherapy-induced acral erythema erythem erythemen erythmateuze dermatose erythema rode huid erythmateuze 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 occurring 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			
			bullla			
			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
----------------	---	---

Part II

Real-world treatment effectiveness and treatment patterns

Chapter 4

Real-world metastatic renal cell carcinoma treatment patterns and clinical outcomes in The Netherlands

Sylvia A. van Laar, Kim B. Gombert-Handoko,
Rolf H.H. Groenwold, Tom van der Hulle, Loes E. Visser,
Danny Houtsma, Henk-Jan Guchelaar, Juliëtte Zwaveling

Front Pharmacol. 2022; 13: 803935

Abstract

Introduction: The number of treatment options for patients with metastatic renal cell carcinoma (mRCC) has significantly grown in the last 15 years. Although randomized controlled trials are fundamental in investigating mRCC treatment efficacy, their external validity can be limited. Therefore, the efficacy of the different treatment options should also be evaluated in clinical practice. We performed a chart review using electronic health records (EHRs) text-mining software to study the current treatment patterns and outcomes.

Methods: MRCC patients from two large hospitals in the Netherlands, starting treatment between January 2015 and May 2020, were included. Data were collected from EHRs using a validated text-mining tool. Primary endpoints were progression-free survival (PFS) and overall survival (OS). Statistical analyses were performed using the Kaplan-Meier method.

Results: Most frequent first-line treatments were pazopanib (n=70), sunitinib (n=34), nivolumab with ipilimumab (n=28). The overall median PFS (mPFS) for first-line treatment was 15.7 months (95% confidence interval [(95% CI), 8.8–20.7]; 16.3 months (95% CI, 9.3–not estimable [NE]) for pazopanib, and 6.9 months (95% CI, 4.4–NE) for sunitinib. The overall median OS (mOS) was 33.4 months (95% CI, 28.1–50.9 months); 39.3 months (95% CI, 29.5–NE) for pazopanib, and 28.1 months (95% CI, 7.0–NE) for sunitinib. For nivolumab with ipilimumab, mPFS and mOS were not reached. Of the patients that finished first- and second-line treatment, 64% and 62% received follow-up treatment.

Conclusion: With most patients starting on pazopanib and sunitinib, these real-world treatment outcomes were most likely better than in pivotal trials, which may be due to extensive follow-up treatments.

1. Introduction

Yearly, more than 400,000 patients worldwide are diagnosed with kidney cancer, of which 90% of the tumors are classified as renal cell carcinoma [1, 2]. Metastatic renal cell carcinoma (mRCC) is not susceptible to chemotherapy or hormonal therapy, and until the introduction of the targeted therapy, the first-line treatment of mRCC was cytokine therapy with interleukin-2 or interferon-alpha (IFN) [3, 4]. Sunitinib was the first targeted therapy that obtained marketing authorization for the treatment of mRCC by the European Medicines Agency in 2006. Subsequently, fourteen other targeted therapies, tyrosine kinase inhibitors (TKIs), mammalian target of rapamycin inhibitors (mTORs) and immune checkpoint inhibitors (ICIs), were approved. Recently, five new combination treatments were registered for first-line mRCC patients (Figure 4.1). Patients with mRCC have a median overall survival (mOS) between 22.9 and 29.1 months on TKIs. For immunotherapy (nivolumab with ipilimumab), only in the intermediate/poor risk group mOS is reached and is 48.1 months [5-7].

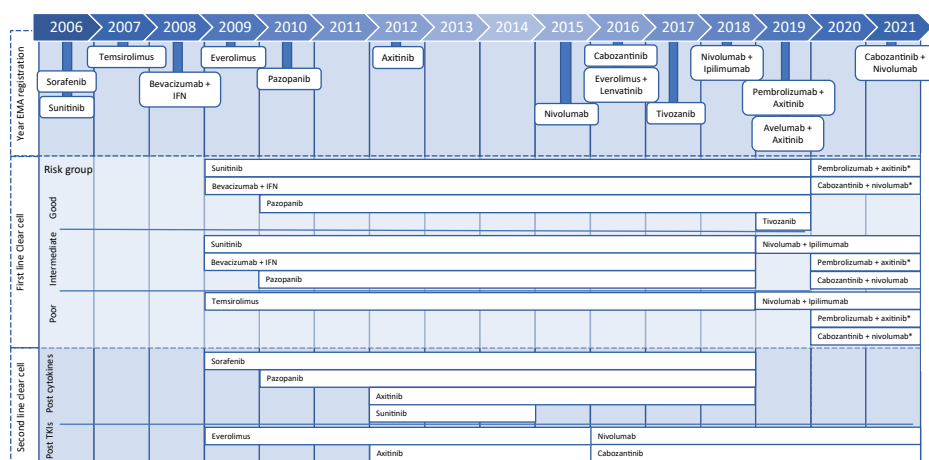


Figure 4.1. Timeline presenting the year of marketing authorization by European Medicines Agency (EMA) of metastatic renal cell carcinoma treatments, and years in which a treatment was included for clear cell renal cell carcinoma treatment in the ESMO clinical Practice Guidelines for diagnosis, treatment and follow-up for first or second line, including the 2020 eUpdate (*).

In 2008, the first European Society for Medical Oncology (ESMO) Clinical Practice Guideline for diagnosis, treatment, and follow-up of RCC has been published and has been updated almost yearly [8]. From 2008 until 2017 the standard of care for patients with clear cell mRCC (ccRCC) with a good or moderate prognosis score, according to the Memorial Sloan-Kettering Cancer Center (MSKCC) or International

mRCC Database Consortium (IMDC) risk model, was first-line treatment with TKIs or bevacizumab with IFN. For patients with a poor prognosis score the standard was treatment with temsirolimus. In 2019, the advised first-line treatment for patients with an intermediate and poor prognosis score changed to nivolumab with ipilimumab. Subsequently, the recommendation for patients with a good prognosis changed in 2020 to pembrolizumab with axitinib or cabozantinib with nivolumab. Both treatment combinations were also added to the intermediate and poor prognosis groups (Figure 4.1). Furthermore, nivolumab and cabozantinib were introduced as second-line in ccRCC treatment options (Figure 4.1) [8, 9].

Randomized controlled trials (RCTs) are the cornerstone for investigating the efficacy of cancer treatments and therefore the pivotal studies for marketing authorization application of mRCC treatments [10-12]. However, due to the structured protocols, ideal setting and inclusion of only a part of the target population, these studies may have limited external validity. Real-world evidence, based on real-world data from registries, databases or observational studies, could be used to close the inferential gap between the RCTs on mRCC treatments and the clinical practice [13-16].

Electronic health records (EHR) contain extensive medical information about patients in clinical practice and therefore are a potentially important resource to assess treatment effectiveness. However, the regular method for EHR data collection, manual chart review, is very time consuming [17-19]. To improve efficiency in real-world data collection in the mRCC population, we investigated and validated a natural language processing (NLP) and text-mining software tool as a more efficient collection method in a cohort of mRCC patients receiving systemic drug treatment [20].

In this study we applied this text-mining tool to investigate how the introduction of new treatment options for mRCC changed clinical practice by evaluating treatment strategies and the efficacy of the different first-line treatment options.

2. Methods

2.1 Study design

A retrospective cohort study was performed to assess the mRCC treatment patterns and treatment outcomes in clinical practice. Patients were included in the Leiden University Medical Center (LUMC), Leiden, and the Haga Hospital, The Hague, the Netherlands.

The study protocol was reviewed and approved by the Medical Ethics Review Committee of the LUMC, Leiden, who waived the need for informed consent.

2.2 Patients and endpoints

All patients aged 18 years or older starting first-line systemic treatment for mRCC between January 2015 and May 2020 (LUMC) or January 2017 and May 2020 (Haga Teaching Hospital) were included. Primary endpoints were first-line progression-free survival (PFS) and overall survival (OS). OS was defined as time from treatment initiation until death from any cause; patients were censored when alive at the moment of data collection. PFS was defined as time from treatment initiation, until date of progression according to the reported PFS by the oncologist which is in line with RECIST 1.1 [21] or death from any cause; patients were censored when still on treatment at moment of data collection, or when treatment ended for other reasons.

The following patient characteristics were collected at start of treatment: age, gender, performance status (PS) (Karnofsky - or Eastern Cooperative Oncology Group (ECOG) PS), tumor histology and prior nephrectomy. Additionally, the IMDC prognostic factors were collected, these include time from diagnosis to systemic therapy within one year, Karnofsky PS below 80%, hemoglobin level below the lower limit of normal and corrected calcium levels, neutrophil levels, and platelet levels above upper limit of normal [22, 23]. All systemic mRCC drug treatments, with concomitant date of progression and date of death as treatment outcome parameters, were collected.

2.3 Data retrieval

Patient inclusion and data collection was performed using a Natural Language Processing (NLP) and text-mining software tool Clinical Data Collector (CDC, CTcue B.V., Amsterdam, The Netherlands). Both the details of the software and the validation process were described previously [24]. In short, for patient selection, characteristics, and outcomes, queries were constructed and validated by comparison to manual review (MR).

2.4 Statistical analysis

Statistical analysis was performed using R (R, 2019, Vienna, Austria). Descriptive statistics were used to describe the total patient cohort and subgroups of first-line treatments with more than 10 patients. Multiple imputation from SPSS (IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.) was used to correct

for missing data on performance status and neutrophil count. Subgroup distribution of variables was compared using the chi-square and ANOVA-tests. Additionally, for the visualization of the treatment patterns, a Sankey plot was created. OS and PFS from first-line treatment were analyzed using Kaplan-Meier method and summarized by medians, 95% confidence intervals and Kaplan-Meier survival plots. All statistical analyses were exploratory.

3. Results

In total 138 patients receiving first-line systemic treatment for metastatic renal cell carcinoma were included in this observational study. Overall, the patient population had a median age of 67 years and was 75% male. Half of the patients had a prior nephrectomy and in 73% of patients a clear cell histology was found. According to the IMDC-criteria, 15% of patients had a good prognosis, 41% of patients an intermediate prognosis, and 44% a poor prognosis (Table 4.1).

3.1 Treatment patterns

From 2015 until 2018, 13 to 24 patients per year started a first-line treatment (Figure 4.2). In this period, approximately 80% of the patients received pazopanib as first-line treatment, the other patients mainly received sunitinib. In 2019, the patients starting with a first-line treatment raised to 43 and approximately half of these patients underwent first-line treatment with nivolumab and ipilimumab.

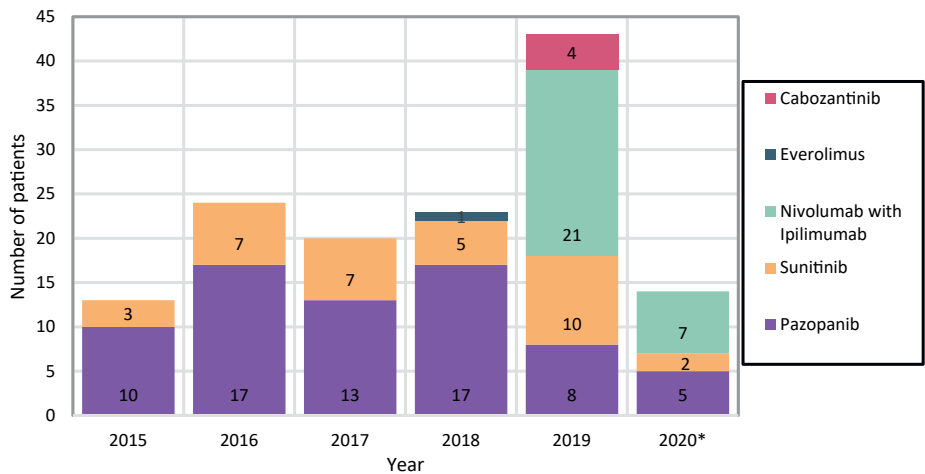


Figure 4.2. Number of first-line treatments started per year. * Patients included until May 2020.

Table 4.1. Patient characteristics of the complete study population, and first-line sunitinib, pazopanib and nivolumab with ipilimumab

	Total n=138, n (%)	First-line sunitinib n=34 (25%), n (%)	First-line pazopanib n=70 (51%), n (%)	First-line nivolumab with ipilimumab n=28 (20%), n (%)	p- value
Sex, male	103 (75)	24 (71)	58 (83)	17 (61)	0.058
Age, median (1st & 3rd qu.)	67 (59–73.75)	64 (58–72.5)	71 (64–76)	63 (57–66.5)	<0.01*
Previous nephrectomy	71 (51)	15 (44)	47 (67)	7 (25)	<0.01*
Histological subtype of renal cell carcinoma					0.034*
Clear cell	101 (73)	19 (56)	56 (80)	23 (82)	
Other					
Papillary	6 (4)	3 (9)	3 (4)	0	
Sarcomatoid	5 (4)	3 (9)	0	0	
Chromofobic carcinoma	1 (1)	0	0	0	
Missing	25 (18)	9 (27)	11 (16)	5 (18)	
IMDC risk score parameters					
Hypercalcemia	54 (38)	10 (29)	24 (34)	16 (57)	0.054
Anemia	88 (64)	25 (74)	40 (57)	17 (61)	0.267
Missing/imputed	3 (2)	1 (3)	0	0	
Neutrophilia	36 (26)	12 (35)	19 (27)	8 (29)	0.69
Thrombocytosis	30 (22)	10 (29)	11 (16)	8 (29)	0.18
Performance status <80% Karnofsky	15 (11)	11 (32)	16 (23)	5 (18)	0.38
Missing/imputed	57 (41)	15 (44)	38 (54)	3 (11)	
Time from diagnosis to systemic therapy <1 year	72 (52)	20 (59)	27 (39)	21 (75)	<0.01*
IMDC risk group					0.175
Favorable risk (0 points)	20 (15)	3 (9)	15 (21)	2 (7)	
Intermediate-risk (1–2 points)	57 (41)	13 (38)	30 (43)	11 (39)	
Poor-risk (>2 points)	61 (44)	18 (53)	25 (36)	15 (54)	

Abbreviations: IMDC: International Metastatic renal cell carcinoma Database Consortium; * p<0.05.

Overall, most patients received pazopanib (n=70, 51%), sunitinib (n=34, 24%), or nivolumab with ipilimumab (n=28, 20%) as first-line treatment. Figure 4.3 shows the treatment patterns of the patients which started on nivolumab with ipilimumab, pazopanib and sunitinib. At moment of data collection, 35 patients were still on first-line treatment and 17 patients and 18 on second- and third-line treatment, respectively. In the 62 patients (64% of the patients that ended first-line treatment) that received

a second-line treatment, nivolumab (n=27, 44%), cabozantinib (n=15, 24%]) and pazopanib (n=10, 16%) were most used. In the third-line of treatment 14 of the 28 patients (62% of the patients that ended second-line treatment) received cabozantinib. Of the eleven patients that ended third-line treatment, eight (73%) received four or more systemic treatments, in which everolimus with lenvatinib was the most used.

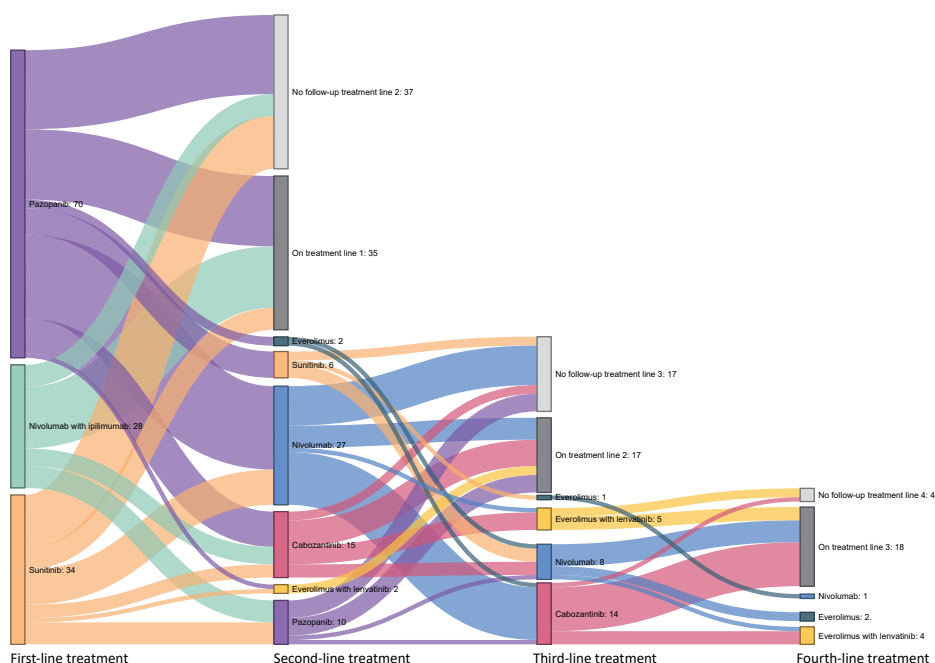


Figure 4.3. Overview of systemic drug treatment patterns following patients which received first-line pazopanib, sunitinib or nivolumab with ipilimumab treatment.

3.2 Treatment outcomes

3.2.1 Patients and patient characteristics

Subgroup comparison of patient characteristics at start of the first-line treatment on pazopanib, sunitinib and nivolumab with ipilimumab showed significant differences in median age (P: 71 years, S: 64 years, N+I: 63 years; $p<0.01$), previous nephrectomy (P: 67%, S: 44%, N+I: 25%; $p<0.01$), known histological subtypes (P: 95% clear cell; 5% papillary, S: 76% clear cell; 12% papillary; 12% sarcomatoid, N+I 100% clear cell; $p=0.034$), and time from diagnosis to systemic therapy within a year (P: 39%, S: 59%, N+I: 75%; $p<0.01$).

3.2.2 Progression-free survival and overall survival

Figure 4.4 shows the progression-free survival and overall survival Kaplan Meier plots stratified for first-line treatment. The overall median PFS after first-line treatment was 15.7 months (95% confidence interval [CI], 8.8–20.7). Patients with first-line pazopanib had a median PFS (mPFS) of 16.3 months (95% CI, 9.3–not estimable [NE]) (Figure 4.4A), and for patients receiving sunitinib a mPFS of 6.9 months (95% CI, 4.4–NE) was observed (Figure 4.4C). No mPFS could be estimated for treatment of nivolumab with ipilimumab due to the limited follow-up time (Figure 4.4E). The overall median OS (mOS) for all patients was 33.4 months (95% CI, 28.1–50.9 months). The mOS per treatment was 39.3 months (95% CI, 29.5–NE) for first-line treatment with pazopanib (Figure 4.4B), and 28.1 months (95% CI, 7.0–NE) after first-line sunitinib (Figure 4.4D). For nivolumab with ipilimumab, mOS could not be established, as the median is not yet reached (Figure 4.4F).

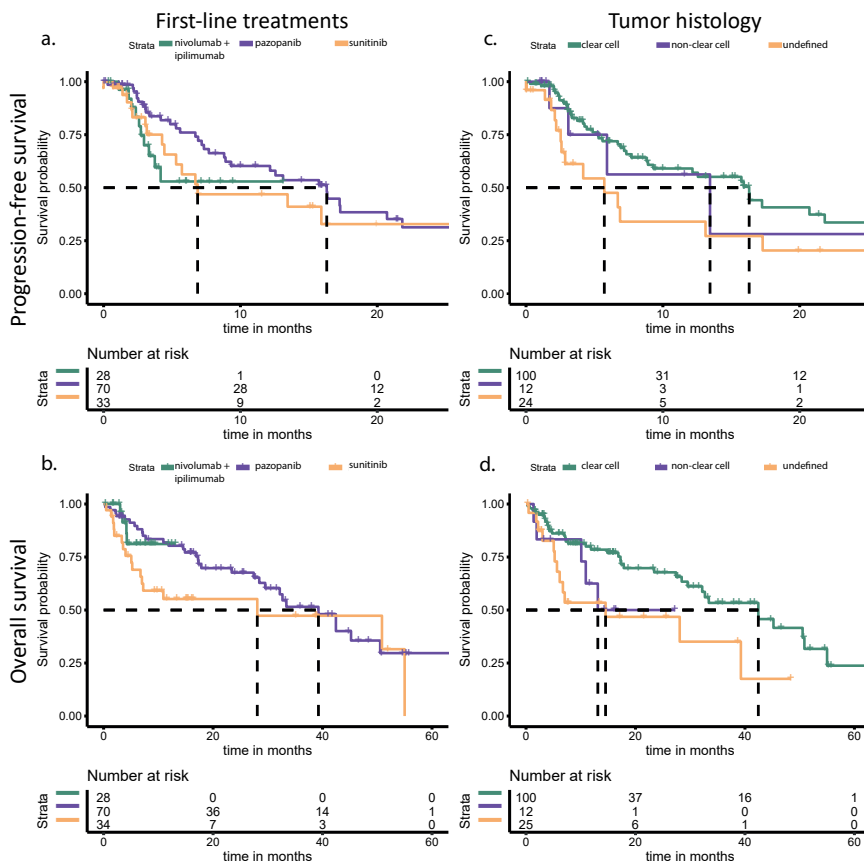


Figure 4.4. Progression-free (PFS) and overall survival (OS) of first-line treatments pazopanib (A: PFS, B: OS), sunitinib (C: PFS, D: OS) and nivolumab with ipilimumab (E: PFS, F: OS).

The stratification to clear-cell and non-clear histology showed a mPFS of 16.3 months (95% CI, 9.3–27.3) and 13.5 months (5.9–NE) (Figure 4.5A), and mOS of 42.4 months (95% CI, 32.2–NE) and 13.1 months (95% CI, 10.9–NE) (Figure 4.5B), respectively.

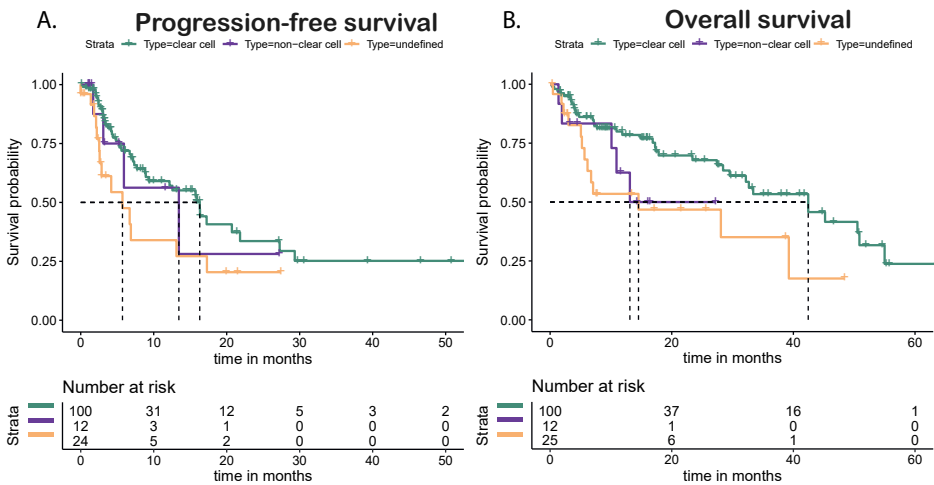


Figure 4.5. Progression-free (A.) and overall (B.) survival of clear-cell versus non-clear cell histology.

4. Discussion

In this study, by using an electronic health record text-mining tool, we were able to evaluate treatment strategies and estimate treatment effectiveness in patients with mRCC in daily practice. After first-line treatment with pazopanib (n=70), sunitinib (n=34) the mPFS was 16.3 months, 6.9 months, and mOS was 39.3 months, 28.1 months, respectively. For nivolumab with ipilimumab (n=28) the mPFS en mOS could not be determined yet.

The pivotal trial of pazopanib by Sternberg et al. showed a mPFS of 11.1 months for the treatment-naïve population and mOS of 22.9 months [25, 26]. Motzer et al. showed a mPFS of 11 months and mOS 26.4 months after first-line sunitinib treatment [27, 28]. Additionally, the COMPARZ-trial, comparing first-line pazopanib to sunitinib prospectively, showed a mPFS of 10.5 months vs. 10.2 months and mOS of 28.4 months vs. 29.3 months, demonstrating non-inferiority of pazopanib to sunitinib [29]. All these RCTs only included patients with (predominant) clear cell histology. The overall mOS in our study (33.4 months) was considerably higher compared to those found in the RCTs, which may be attributed to a longer mOS in the pazopanib treated patients.

However, the mPFS of both treatments and the mOS after treatment with sunitinib seem to be comparable to those found in the pivotal trials.

Although sunitinib and pazopanib are both TKIs and often-used first-line treatments, these subgroups should not be used for a head-to-head comparison as the populations in our study differ on several patient characteristics (age, previous nephrectomies, tumor histology, and number of patients that started systemic treatment within a year). The sunitinib-population included more patients with worse scores for all characteristics, except age. Pazopanib was of preference in the higher age population since the COMPARZ-trial showed better safety quality-of-life profiles than sunitinib.

The patients in our real-world study differed from the RCTs concerning several patient characteristics. Patients treated with pazopanib, e.g., had a higher median age (71 vs 59), fewer (partial) nephrectomies (67.1% vs 89%) and more patients (27.3%) were assigned to a IMDC poor risk group than in the pivotal study which included only 3% of patients with a poor Memorial Sloan-Kettering Cancer Center (MSKCC) risk [25]. For sunitinib treated patients we observed that, in contrast to the pivotal study, a significant part of our patients treated with sunitinib had a non-clear cell histology. Furthermore, our population had less (partial) nephrectomies (44.1% vs. 91%) and at least 32% of the patients had a confirmed Karnofsky performance status <80% in contrast to 0% in the RCT [28]. Since all these differences in patient characteristics in the real-world compared to RCTs are related with worse prognosis [22, 30, 31], they do not explain the promising effectiveness of sunitinib and pazopanib we found in our real-world population in comparison to RCTs.

A longer mOS, but comparable mPFS, of pazopanib, as compared to the pivotal studies, is also seen in recent real-world studies [32-34]. The earlier systematic review and meta-analysis of Climent et al. showed that real-world pazopanib studies published until 31 December 2016 resulted in mPFS ranging from 8.1–15.9 months and mOS ranging from 16.1 until 31.0 months, with a pooled mPFS of 10.0 months and mOS of 22.7 months [35]. Also, for sunitinib an improved mOS is found in recent studies [33, 34]. However, our findings for sunitinib are more in accordance with the systematic review and meta-analysis of Moran et al., which showed a pooled mPFS of 9.4 months and mOS of 20.8 months for sunitinib with individual results for mPFS and mOS ranging from 7.5–11.0 months and 6.8–33.2 months [36]. Even though mOS of sunitinib in our real-world situation is similar to the RCTs, this was unexpected given the patient characteristics, which are indicative for poor outcome.

More and better follow-up treatments may be the most obvious explanation for prolonged OS. In this study, we show that 64% of the patients who finish a first-line treatment receive a second line of treatment, and congruent 62% a third line, among which the relative new treatments nivolumab and cabozantinib are the most used. Both cabozantinib and nivolumab have shown to significantly improve mOS as second-line treatment, even after recurrent progression, when compared to treatment with everolimus [37, 38]. A real-world unselected patient group showed a median duration of cabozantinib treatment of 7.6 months [39]. In the pivotal trial for first-line pazopanib, only 30% of the patients received any post-progression systemic therapy, most often sunitinib and sorafenib [26], which is a considerable lower percentage of the patients than in the real-world situation. Additionally, Schmidinger et al. showed, evaluating real-world data of an Austrian population, that 67.7% of the patients received second-line treatment and had improved OS when compared to the RCTs. They argue that an improved OS is related to the quality of the healthcare system in a country and reimbursement of subsequent treatments [34]. For example, in Brazil and The United Kingdom only 20% and 15.8% of the mRCC patients are treated with second-line systemic therapy, and in these countries, limited availability and costs are expected to be barriers to optimal treatment [40, 41].

Moreover, toxicity-related dose reduction was linked to better outcomes in the study of Wagstaff et al. [42]. Therefore, clinicians having more experience with treatments, such as sunitinib and pazopanib, may be more alert to apply dose-reductions when necessary and in this way contribute to the overall better prognosis of these treatments in the real-world with respect to earlier stage RCTs.

Twenty-eight patients in our cohort started first-line treatment with nivolumab and ipilimumab since 2019. Since one of the hospitals was designated for nivolumab in combination with ipilimumab treatment, this resulted in an influx of new mRCC patients. In 92% of the patients of which the IMDC prognosis was known, the prognosis was intermediate or poor, according to the ESMO guidelines. Therefore, this study shows that the expected paradigm shift from TKIs to immunotherapy in mRCC treatment has started in the group of patients with intermediate and poor IMDC prognosis [43]. However, no median survival data were reached in this study. Therefore, we were not able to verify the mPFS of 11.2 months and mOS of 48.1 months reached in the CheckMate-214 study in the intermediate and poor risk patient group [44].

As far as we know, this is the first study in which a text-mining tool is used for real-world data extraction from electronic health records to study treatment patterns and

outcomes of metastatic renal cell carcinoma treatments. By using the text-mining tool, this study could be performed more efficient than by manual review (a mean of 12 minutes vs 86 minutes per patient). Also, most outcome measurements could be extracted with the desired high accuracy of at least 90% [24]. Since this method enables faster data extraction compared to manual review, in the future tools as these could be repeatedly used to evaluate treatments. These data can illustrate who real-world patients are and how new treatments influence their survival. The data can be used on several levels [45]: on the patient level, as additional information to share with a patient in the process of shared decision making [46], on hospital level, or, since these queries are transferrable to other hospitals, even multicenter level, for treatment evaluation and pursuit of value-based healthcare and patient-centered care [47].

This study has some limitations. First, although the EHR is an information rich source containing longitudinal patient data, it is a secondary source [17]. Therefore the data quality is dependent on the healthcare professionals documentation and not all data desired for research may be documented. In this study the performance status was the least well documented and missing in 41.3% of the patients, which is often seen in chart reviews [48, 49]. Also, patients in this study could have participated in other post-registration trials, e.g. the DIET-study [50]. In addition, we were limited in the extraction of prognostic factors, e.g., tumor load and location of metastases, as the EHR documentation was still too complex for the text-mining tool.

4.1 Future perspectives

This paper shows that the real-world population differs from the pivotal trial populations and that multiple factors may influence treatment outcomes. Using a text-mining tool, a quick evaluation of the effectiveness of the treatments for mRCC carcinoma in real-world patients is possible. Therefore, repeated use of the queries from the EHR text-mining tool can provide recurring information for physicians, useful for decision-making for the treatment of individual patients, now and in the future in the field of mRCC carcinoma. This is especially relevant in present field, since more treatment combinations are already there and are expected to come [51].

Conclusion

The aim of this study was to evaluate the treatment patterns and outcomes of mRCC patients. With most patients starting on a treatment with pazopanib and sunitinib, the outcomes of these real-world patients most probably were better than expected from pivotal trials. The extensive follow-up treatments patients received, may have contributed to the improved outcomes. The used EHR text-mining method can be easily applied for evaluation of other treatments in clinical practice, which may be useful in the rapidly evolving field of renal cell carcinoma treatments.

References

1. Ferlay, J., et al., *Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods*. Int J Cancer, 2019. **144**(8): p. 1941-1953.
2. Janzen, N.K., et al., *Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease*. Urol Clin North Am, 2003. **30**(4): p. 843-52.
3. Hutson, T.E., *Targeted therapies for the treatment of metastatic renal cell carcinoma: clinical evidence*. The oncologist, 2011. **16 Suppl 2**(Suppl 2): p. 14-22.
4. Oudard, S., et al., *Treatment options in renal cell carcinoma: past, present and future*. Annals of Oncology, 2007. **18**: p. x25-x31.
5. Nocera, L., et al., *Clinical Outcomes and Adverse Events after First-Line Treatment in Metastatic Renal Cell Carcinoma: A Systematic Review and Network Meta-Analysis*. J Urol, 2022. **207**(1): p. 16-24.
6. Albiges, L., et al., *Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial*. ESMO Open, 2020. **5**(6): p. e001079.
7. Choueiri, T.K. and R.J. Motzer, *Systemic Therapy for Metastatic Renal-Cell Carcinoma*. N Engl J Med, 2017. **376**(4): p. 354-366.
8. Escudier, B., et al., *Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†*. Annals of Oncology, 2019. **30**(5): p. 706-720.
9. Committee, E.G. *eUpdate – Renal Cell Carcinoma Treatment Recommendations*. 2020 9 July 2021].
10. 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.
11. Bothwell, L.E. and S.H. Podolsky, *The Emergence of the Randomized, Controlled Trial*. 2016. **375**(6): p. 501-504.
12. Verweij, J., et al., *Innovation in oncology clinical trial design*. Cancer Treatment Reviews, 2019. **74**: p. 15-20.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. van Laar, S.A., et al., *An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma*. Clin Pharmacol Ther, 2020. **108**(3): p. 644-652.

21. 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.
22. Heng, D.Y., et al., *External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study*. Lancet Oncol, 2013. **14**(2): p. 141-8.
23. Heng, D.Y., et al., *Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study*. J Clin Oncol, 2009. **27**(34): p. 5794-9.
24. van Laar, S.A., et al., *An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma*. Clin Pharmacol Ther, 2020.
25. Sternberg, C.N., et al., *Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial*. J Clin Oncol, 2010. **28**(6): p. 1061-8.
26. 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.
27. Motzer, R.J., et al., *Sunitinib versus interferon alfa in metastatic renal-cell carcinoma*. N Engl J Med, 2007. **356**(2): p. 115-24.
28. Motzer, R.J., et al., *Sunitinib in patients with metastatic renal cell carcinoma*. JAMA, 2006. **295**(21): p. 2516-24.
29. Motzer, R.J., et al., *Pazopanib versus sunitinib in metastatic renal-cell carcinoma*. N Engl J Med, 2013. **369**(8): p. 722-31.
30. Bhindi, B., et al., *Comparative Survival following Initial Cyto-reductive Nephrectomy versus Initial Targeted Therapy for Metastatic Renal Cell Carcinoma*. J Urol, 2018. **200**(3): p. 528-534.
31. Padala, S.A., et al., *Epidemiology of Renal Cell Carcinoma*. World journal of oncology, 2020. **11**(3): p. 79-87.
32. Schmidinger, M., et al., *Prospective Observational Study of Pazopanib in Patients with Advanced Renal Cell Carcinoma (PRINCIPAL Study)*. Oncologist, 2019. **24**(4): p. 491-497.
33. Kim, M.S., et al., *Efficacy of First-Line Targeted Therapy in Real-World Korean Patients with Metastatic Renal Cell Carcinoma: Focus on Sunitinib and Pazopanib*. J Korean Med Sci, 2018. **33**(51): p. e325.
34. Schmidinger, M., et al., *Real-World Evidence Data on Metastatic Renal-Cell Carcinoma Treatment in Austria: The RELACS Study*. Clin Genitourin Cancer, 2019. **17**(5): p. e957-e967.
35. Climent, M.A., et al., *Systematic review and survival meta-analysis of real world evidence on first-line pazopanib for metastatic renal cell carcinoma*. Crit Rev Oncol Hematol, 2018. **121**: p. 45-50.
36. Moran, M., et al., *Sunitinib for Metastatic Renal Cell Carcinoma: A Systematic Review and Meta-Analysis of Real-World and Clinical Trials Data*. Target Oncol, 2019. **14**(4): p. 405-416.
37. Choueiri, T.K., et al., *Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial*. Lancet Oncol, 2016. **17**(7): p. 917-927.
38. Motzer, R.J., et al., *Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma*. N Engl J Med, 2015. **373**(19): p. 1803-13.
39. Albiges, L., et al., *Real-world evidence of cabozantinib in patients with metastatic renal cell carcinoma: Results from the CABOREAL Early Access Program*. European Journal of Cancer, 2021. **142**: p. 102-111.

40. Bergerot, P.G., et al., *Assessment of Treatment Patterns for Metastatic Renal Cell Carcinoma in Brazil*. Journal of Global Oncology, 2018(4): p. 1-8.
41. Wagstaff, J., et al., *Treatment patterns and clinical outcomes in patients with renal cell carcinoma in the UK: insights from the RECCORD registry*. Ann Oncol, 2016. **27**(1): p. 159-65.
42. Escudier, B., et al., *Treatment Beyond Progression in Patients with Advanced Renal Cell Carcinoma Treated with Nivolumab in CheckMate 025*. Eur Urol, 2017. **72**(3): p. 368-376.
43. Calvo, E., et al., *The Current and Evolving Landscape of First-Line Treatments for Advanced Renal Cell Carcinoma*. Oncologist, 2019. **24**(3): p. 338-348.
44. Albiges, L., et al., *Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial*. ESMO Open, 2020. **5**(6): p. e001079.
45. Venkatakrishnan, K., P.H. van der Graaf, and S.A. Holstein, *The Changing Face of Oncology Research, Drug Development, and Clinical Practice: Toward Patient-Focused Precision Therapeutics*. Clinical Pharmacology & Therapeutics, 2020. **108**(3): p. 399-404.
46. Bomhof-Roordink, H., et al., *Shared decision making in oncology: A model based on patients', health care professionals', and researchers' views*. Psycho-Oncology, 2019. **28**(1): p. 139-146.
47. Tseng, E.K. and L.K. Hicks, *Value Based Care and Patient-Centered Care: Divergent or Complementary?* Current Hematologic Malignancy Reports, 2016. **11**(4): p. 303-310.
48. Noize, P., et al., *Real-life patterns of use, safety and effectiveness of sunitinib in first-line therapy of metastatic renal cell carcinoma: the SANTORIN cohort study*. Pharmacoepidemiol Drug Saf, 2017. **26**(12): p. 1561-1569.
49. Day, D., et al., *Patterns of care for metastatic renal cell carcinoma in Australia*. BJU Int, 2015. **116** Suppl 3: p. 36-41.
50. Lubberman, F.J.E., et al., *The Effect of Using Pazopanib With Food vs. Fasted on Pharmacokinetics, Patient Safety, and Preference (DIET Study)*. Clinical Pharmacology & Therapeutics, 2019. **106**(5): p. 1076-1082.
51. Committee, E.G. *eUpdate - Renal Cell Carcinoma Treatment Recommendations 2*. 2020 [cited 2020 October 8, 2020]; Available from: <https://www.esmo.org/guidelines/genitourinary-cancers/renal-cell-carcinoma/eupdate-renal-cell-carcinoma-treatment-recommendations-2>.

Chapter 5

Application of an electronic health record text mining tool: Treatment patterns and outcomes for Dutch patients with hepatocellular carcinoma

Sylvia A. van Laar, Willem van der Weg,
Kim B. Gombert-Handoko, Tom van der Hulle,
Henk-Jan Guchelaar, Juliëtte Zwaveling

Manuscript in preparation

Abstract

Introduction: Sorafenib was the first systemic treatment registered for hepatocellular carcinoma (HCC) in 2007, followed by others after 2017. Due to the high variety in etiology, and risk factors for HCC, the randomized controlled trial (RCT) results may not be representative for all populations. In this study we used a text-mining tool to collect real-world data (RWD) from electronic health records (EHRs) to review baseline characteristics, including trial eligibility criteria, and treatment outcomes.

Methods: Adult HCC patients receiving systemic treatment between January 2015 and 15 October 2021 at the Leiden University Medical Center, The Netherlands were included. Baseline characteristics, progression-free survival (PFS) and overall survival (OS) were collected. The patient inclusion and data collection from EHR were performed with CTcue[®] text-mining software.

Results: In total, 94 patients were included, 88% received sorafenib as first-line treatment, 7% received atezolizumab plus bevacizumab, and 4% lenvatinib; 16% started second-line treatment, mostly with regorafenib. Of patients treated with first-line sorafenib, at least 28% did not meet eligibility criteria. Furthermore, median PFS was 4.1 months (95% confidence interval 2.8–6.1) and the median OS was 10.5 months (95% confidence interval 7.8–16.9). Child-Pugh class B-C ($p=0.014$) was significantly associated with poor survival outcomes.

Conclusion: Most patients in this retrospective EHR text-mining study were treated with sorafenib. The median first-line sorafenib treatment outcomes were comparable to the RCTs and other real-world studies. In our hospital patients with Child-Pugh B-C also received treatment, but did not clinically benefit. This study confirms text-mining can optimize the evaluation of the current and upcoming HCC treatments.

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most lethal cancers, with a 1-year survival of 45% and 5-year survival of 19% in the last decade (2011–2019) [1]. Yearly, 4.27 per 100,000 persons in The Netherlands are diagnosed with HCC [2]. In 2007, sorafenib – a tyrosine kinase inhibitor – was registered by the European Medicines Agency (EMA) as first effective systemic treatment of advanced HCC [3]. This was followed by the introduction of several other first- and second-line treatments, including tyrosine kinase inhibitors (lenvatinib, cabozantinib, regorafenib) and immunotherapy (atezolizumab plus bevacizumab). Table 5.1 shows an overview of the current systemic treatments of HCC registered by the EMA, including the related phase III randomized controlled trial (RCT) results and the region, population, inclusion criteria, and survival outcomes.

All RCTs for HCC treatments had inclusion criteria with limitations on Child-Pugh class, performance status, life expectancy, and organ functions (Table 5.1). It is expected that in clinical practice patients do not always meet these eligibility criteria, e.g., other studies showed that patients with Child-Pugh class B and C and patients undergoing dialysis also received sorafenib treatment [4, 5]. Therefore, trial results might not always be representative for HCC populations in clinical practice, and since results from RCTs show limited improvement of outcomes with respect to placebo (Table 5.1), it raises the question whether treatment in daily practice really adds to patient outcomes [6].

Next to inclusion criteria, the trial populations differ between East Asia, Sub-Saharan Africa, and Western countries. Chronic and viral hepatitis infections (hepatitis B [HBV] and hepatitis C [HCV]) have traditionally been the most important risk factors and are mainly represented in the RCTs [7, 8]. However, fatty liver disease (FDL) is the risk factor emerging in the last decade as leading etiology, which can be further specified in non-alcoholic FLD (NAFLD), associated with multiple metabolic disorders originating from a changed western dietary pattern, and lifestyle, and alcoholic FDL (AFLD), originating from an excess of alcohol consumption [9].

The electronic health record (EHR) is a relevant source of real-world data, containing longitudinal data on patients and their treatment trajectories [10]. However, manual data extraction, the standard data collection method, is known to be labor intensive and time-consuming. Previously, we validated and applied a text-mining tool as a more efficient method to collect real-world data from EHR regarding renal cell carcinoma patients and their treatments [11, 12].

Table 5.1. Systemic treatments for hepatocellular carcinoma registered at the EMA including their phase III trial results

Treatment	Registration year EMA	Phase III trial	Compara- tor	Region	Eligibility criteria	Line of treatment	Cause of disease	Median PFS (months)	Median OS (months)
Sorafenib	2007	SHARP [13]	Placebo	Inter- continental	ECOG PS ≤2, Child-Pugh class A, life expectancy > 12 weeks, and adequate hematologic-, hepatic and renal function.	First	HCV (29%), alcohol (26%), HBV (19%)	S: 5.5 HR: 0.58 p<0.001	S: 10.7 P: 7.9 HR: 0.69 p=0.001
Regorafenib	2017 [15]	RESORCE [16]	placebo	Asia & rest of the world	ECOG PS ≤2, Child-Pugh class A, life expectancy > 12 weeks, and adequate hematologic-, hepatic and renal function.	First	HBV (71%), HCV (11%)	TTP S: 2.8 P: 1.4 HR: 0.57 p=0.0005	S: 6.5 P: 4.2 HR: 0.68 p=0.014
Lenvatinib	2018 [17]	REFLECT [18]	Sorafenib	Asia-pacific, Europe, North America	ECOG PS 0 or 1, Child- Pugh class A, adequate, hematologic, controlled blood pressure, adequate liver, bone marrow, blood, renal and pancreatic function	First	HBV (38%), Alcohol (24%), HCV (21%)	R: 3.1 P: 1.5 HR: 0.46 p<0.001	R: 10.6 P: 7.8 HR: 0.63 p<0.0001
Lenvatinib	2018 [17]	REFLECT [18]	Sorafenib	Asia-pacific, Europe, North America	ECOG PS 0 or 1, Child- Pugh class A, adequate, hematologic, controlled blood pressure, adequate liver, bone marrow, blood, renal and pancreatic function	First	HBV (53%), HCV (19%)	L: 7.4 S: 3.7 HR: 0.66 p<0.0001	L: 13.6 S: 12.3 HR: 0.92 p>0.05

Cabozantinib	2018 [19]	CELESTIAL [20]	Placebo	Asia, Europe, North America	ECOG PS 0 or 1, Child-Pugh class A, previous treatment with sorafenib, adequate hematologic measures and renal function	Second	HBV (38), HCV (24%), Alcohol (24%)	C: 5.2 P: 1.9 HR: 0.44 p<0.001	C: 10.2 P: 8.0 HR: 0.76 p=0.005
Ramucirumab	2019 [21]	REACH-2 [22]	Placebo	America's, Europe, Australia, Israel, Asia, Japan	ECOG PS 0 or 1, serum alfa-fetoprotein concentration>400 ng/ml, adequate hematological and biochemical parameters, previous treatment with sorafenib	Second	HBV (36%), HCV (24%), Alcohol (24%)	R: 2.8 P: 1.6 HR: 0.452 p<0.0001	R: 8.5 P: 7.3 HR: 0.71 p=0.0199
Atezolizumab + bevacizumab	2020 [23]	IMbrave150 [24]	Sorafenib	Asia (excl. Japan), rest of the world	ECOG PS 0 or 1, Child-Pugh class A, adequate hematologic and organ function	First	HBV (49%), HCV (21%)	AB: 6.8 S: 4.3 HR: 0.59 p<0.001	AB: NE S: 13.2 HR: 0.58 p<0.001

EMA: European Medicines Agency, HBV: hepatitis B virus, HCV: hepatitis C virus, PFS: progression-free survival, OS: overall survival, TTP: time to radiological progression.

In this study, we applied this text-mining tool to review baseline characteristics, including eligibility criteria and disease etiology, treatment patterns and treatment outcomes in a Dutch population of HCC patients.

2. Methods

We performed a retrospective cohort study by reviewing EHRs using text mining. The study protocol was reviewed and approved by the Medical Ethics Review Committee of the Leiden University Medical Center (LUMC), Leiden, The Netherlands, who waived the need for informed consent.

2.1 Patient population

Patients ≥ 18 years diagnosed with HCC were included if they started any systemic treatment (atezolizumab plus bevacizumab, cabozantinib, lenvatinib, ramucirumab, regorafenib, sorafenib) between 1 January 2015 and 15 October 2021 in the LUMC, The Netherlands.

2.2 Collected data

The following patient characteristics were collected at start of treatment for patients: sex, age, length, weight, cause of the disease, performance status (Karnofsky and WHO/ECOG), Barcelona Clinic Liver cancer (BCLC) stage, presence of extrahepatic metastasis, esophageal varices, liver cirrhosis, macrovascular invasion of the tumor, ascites, and presence of hepatic encephalopathy, Child-Pugh score, previously received treatments (hepatectomy, ablation, and embolization), time from diagnosis to systemic treatment < 1 year, the levels of aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin, albumin and alpha-fetoprotein (AFP), and the prothrombin time. If the Child-Pugh score was not mentioned in the text, this was calculated based on known risk factors [25]. We determined whether patients met the eligibility criteria of the sorafenib RCTs as shown in Table 5.1, excluding the life expectancy, as this could not retrospectively be assessed.

The effectiveness was estimated by estimating the median progression-free survival (PFS) and overall survival (OS) after initiation of sorafenib treatment. PFS was determined as the time from start of treatment until tumor progression or death by any cause, with the moment of progression reported by the oncologist, who evaluate

progression by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Patients were censored if they discontinued without tumor progression or if they were on treatment at moment of data extraction. OS was defined as time from start of treatment until death by any cause. Patients were censored when they were alive at moment of data extraction.

2.3 Data collection method

Both patient inclusion and data collection were performed using an EHR text-mining tool CTcue (CTcue B.V., Amsterdam, The Netherlands). This tool enables medical researchers to search and extract both structured (e.g., laboratory values and prescribed medication) and unstructured (e.g., medical notes) data from EHR (HiX, Chipsoft B.V., Amsterdam, The Netherlands) by rule-based text mining, including, and was previously validated and applied for treatment evaluation of a population of renal cell carcinoma patients [11, 12].

Patients were selected based on the diagnosis treatment code for hepatocellular carcinoma, combined with at least one prescription of atezolizumab plus bevacizumab, cabozantinib, lenvatinib, ramucirumab, regorafenib, sorafenib between January 2015 and October 2021 and at least one mention of HCC (or synonyms) in free text notes. Context analysis of the free text notes was performed to finally include patients in the study population. Furthermore, all data collected from unstructured text (cause of disease, performance status, stage, presence of extrahepatic metastasis, esophageal varices, liver cirrhosis, macrovascular invasion of the tumor, ascites, hepatic encephalopathy, Child-Pugh score, previous treatments) was confirmed with context analysis in the software tool. Applied queries are available in Supplementary File S5.1.

2.4 Analysis

Statistical analyses were performed using SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp) and R (R core team, 2021. R foundation for statistical computing, Version 4.1.1. Vienna, Austria). Descriptive statistics were used to summarize baseline characteristics from the total cohort and subgroups of first-line treatments with more than ten patients. Subsequent analyses also were limited to first-line treatment groups larger than ten patients. The PFS and OS were analyzed with the Kaplan-Meier method. Cox regression was used to estimate the hazard ratio of prognostic factors on OS. All statistical analyses were exploratory.

3. Results

In total, 107 patients were identified with the text-mining tool, of which 13 were excluded after manual confirmation mainly since they had another indication, did not start treatment, or started treatment outside of the inclusion period, and therefore leaving 94 patients in the analysis (Figure 5.1).

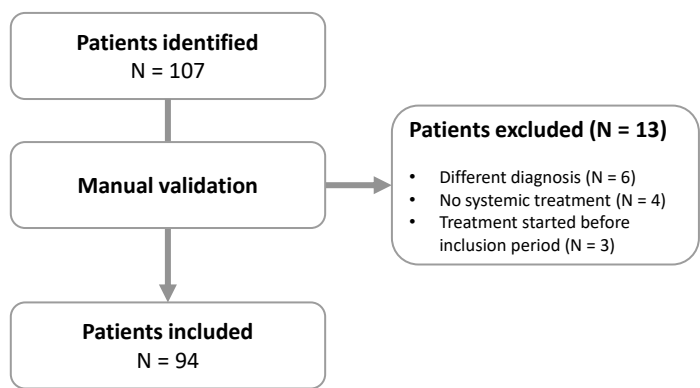


Figure 5.1. Patient inclusion flowchart.

Table 5.2. Patient characteristics

Variable	Total n=94	Sorafenib n=83 (88%)
Age, mean (years, $\pm$ SD)	66.7 ($\pm$ 10.0)	66.3 ($\pm$ 10.0)
Male (n, %)	64 (68.1)	56 (67.5)
BMI, mean (kg/m ² , $\pm$ SD)	27.2 ($\pm$ 5.8)	27.2 ($\pm$ 5.8)
Unknown	12	10
ECOG performance status (n, %)		
0	22 (23.4)	20 (24.1)
1	14 (14.9)	12 (14.5)
2	5 (5.3)	4 (4.8)
3	1 (1.1)	1 (1.2)
Unknown	52 (55.3)	46 (55.4)
Cause of disease (n, %)		
Alcohol use	43 (45.7)	40 (48.2)
Hepatitis B	14 (14.9)	13 (15.7)
Hepatitis C	16 (17.0)	16 (19.3)
Metabolic syndrome	11 (11.7)	11 (13.3)
Hemochromatosis	4 (4.3)	4 (4.8)
Unknown	27 (28.7)	15 (18.1)

Table 5.2 continues on next page.

Table 5.2. *Continued*

Variable	Total n=94	Sorafenib n=83 (88%)
Child-Pugh score (n, %)		
A	67 (71.2)	56 (67.5)
B	16 (17.0)	16 (19.3)
C	1 (1.1)	1 (1.2)
Unknown	10 (10.6)	10 (12.0)
BCLC stage (n, %)		
A (early stage)	1 (1.1)	1 (1.2)
B (intermediate)	10 (10.6)	10 (12.0)
C (advanced)	51 (54.3)	45 (54.2)
Unknown	32 (34.0)	27 (32.5)
Biochemical analysis		
Albumin, median (g/dL, IQR)	40.0 (36–44)	40.0 (36–44)
Unknown	1	1
Alpha-fetoprotein, median (µg/L, IQR)	59.5 (6–1896))	84.5 (6.3–2196
Unknown	4	3
ALT, median (U/L, IQR)	42.5 (25.3–58.8)	42.0 (25.5–56.5)
Unknown	2	2
AST, median (U/L, IQR)	58.0 (35–78)	58.0 (35–78)
Unknown	2	2
Bilirubin, median (µmol/L, IQR)	12.0 (7–20)	12.0 (7–20)
Unknown	1	1
Creatinine, median (µmol/L, IQR)	74 (60–88)	75 (60–87.25)
Unknown	1	1
Hemoglobin, median (mmol/L, IQR)	7.9 (7.0–8.6)	7.8 (6.8–8.6)
Unknown	3	3
PT-INR, median (ratio, IQR)	1.1 (1.0–1.1)	1.1 (1.0–1.2)
Unknown	5	1
Platelet count, median (x10 ⁹ /L, IQR)	195 (120–275)	186.0 (118.8–263.8)
Unknown	1	
Previous therapy (n, %)	51 (54.3)	46 (55.4)
Hepatectomy	17 (18.1)	17 (20.5)
Embolization (TACE/TARE)	37 (39.4)	32 (38.6)
Ablation (RFA/MWA)	28 (29.8)	24 (33.7)

ALT: alanine aminotransferase, AST: aspartate aminotransferase, BCLC: Barcelona Clinic Liver Cancer staging system, BMI: body mass index, ECOG: Eastern Cooperative Oncology Group, TACE/TARE: Transarterial chemo-embolization / transarterial radioembolization.

Table 5.2 shows an overview of all baseline patient- and disease characteristics. Overall, patients had a mean age of 66.7 years, 68% was male, and had a mean BMI of 27.2 kg/m². Of the patients it was known for, most had a ECOG performance status of 0

(23%), BCLC stage C (54%), Child-Pugh score A (71%) and alcohol as cause of disease (46%). Furthermore, most patients received sorafenib as first-line treatment (n=83, 88%), and a small group started with atezolizumab plus bevacizumab (n=7, 7%) or lenvatinib (n=4, 4%). Most patients only received one line of treatment (n=78, 83%), but regorafenib (n=11, 12%) was the most prevalent second-line treatment (Figure 5.2).

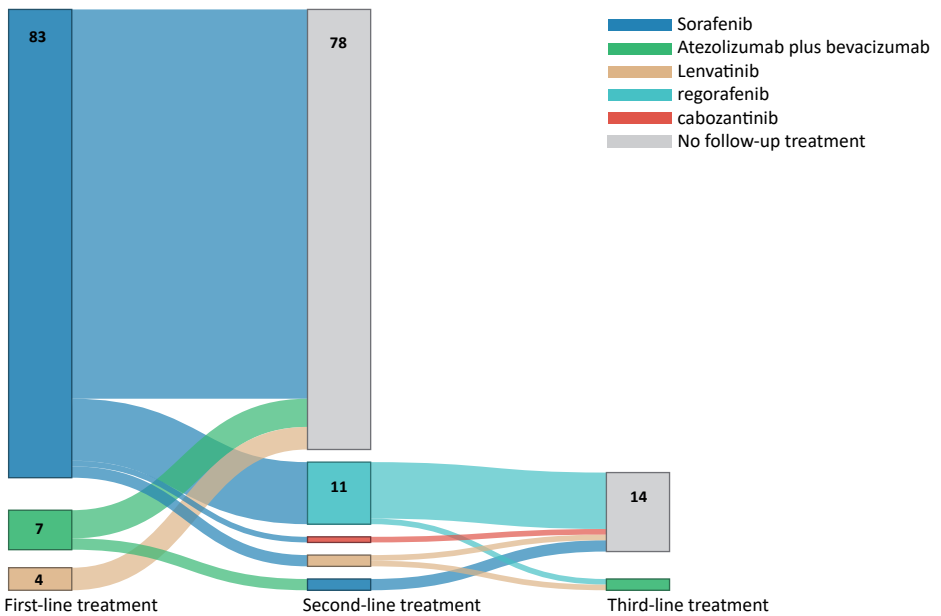


Figure 5.2. Treatment patterns.

3.1 First-line sorafenib

Of the 83 patients that started with sorafenib, at least 28% did not meet the trial eligibility criteria (Table 5.3), mostly due to not having Child Pugh Class A, but B or C. Furthermore, Figure 5.2 shows Kaplan-Meier plots summarizing the progression-free survival and overall survival results. The median PFS was 4.1 months (95% CI 2.8–6.1), and the median OS was 10.5 months (95% CI 7.8–16.9). Furthermore, Child-Pugh B-C at baseline was significantly associated with worse survival ($p=0.014$), but not meeting eligibility criteria was not ($p=0.059$, Table 5.4).

Table 5.3. Patients treated with sorafenib not meeting eligibility criteria

Eligibility criterium	n (%)
ECOG performance status ≤ 2	1 (1.2)
Child-Pugh class A	17 (20.5)
Hematologic function	
Platelet count $\geq 60 \times 10^9/l$	2 (2.4)
Hemoglobin ≥ 8.5 g/dl	0 (0)
PT-INR ≤ 2.3	2 (2.4)
Hepatic function	
Albumin ≥ 2.8 g/dl	1 (1.2)
Total bilirubin ≤ 3 mg/dl	0 (0)
ALT ≤ 5 x ULN	0 (0)
AST ≤ 5 x ULN	5 (6.0)
Renal function	
Creatinine ≤ 1.5 x ULN	2 (2.4)
Total not meeting ≥ 1 eligibility criterium	23 (27.7)

ALT: Alanine aminotransferase; AST: Aspartate aminotransferase, ECOG: Eastern Cooperative Oncology Group, PT-INR: Prothrombin time – international normalized ration, ULN: upper limit of normal.

Table 5.4. Prognostic factors for overall survival of first-line sorafenib treatment (univariate)

Variable	n (%)	HR (95% CI)	p-value
Sex			
Male	23	1	
Female	14	0.77 (0.44–1.33)	0.34
Age			
<65 years	30	1	
≥ 65 years	53	0.84 (0.50–1.41)	0.51
Etiology: Alcoholic liver disease			
No	39	1	
Yes	40	1.27 (0.76–2.14)	0.37
Child-Pugh score			
A	56	1	
B-C	17	2.19 (1.18–4.08)	0.014*
Meets eligibility criteria			
Yes / unknown	58	1	
No	25	1.72 (0.98–3.04)	0.059

HR: Hazard-ratio, *: Meeting statistical significance of $p < 0.05$.

4. Discussion

In this study, we successfully applied an EHR text-mining tool to collect HCC patient characteristics, treatments, and outcomes in Dutch daily clinical practice. In the period from 2015 to 2021, most of the 94 patients received sorafenib (88%) as first-line treatment, with a median PFS of 4.1 months and a median OS of 10.5 months. 28% of these patients did not meet trial eligibility criteria, of which baseline Child-Pugh score B-C was prognostic for worse survival in this population.

4.1 Sorafenib

The mPFS of sorafenib in our study was slightly lower than the mPFS of 5.5 months in the SHARP trial, but the mOS of 10.7 months was comparable [13]. However, the Asia-Pacific trial reported a time to progression of 2.8 months and a mOS of 6.5 months (95% CI 5.56–7.56) [14], which is significantly lower than we reported. Our study population differed from the trial populations, as expected, on certain characteristics; we included more women (32% vs. 13% and 15%) and alcohol use was the predominant cause of the disease (48%), which was associated with poorer outcomes, compared to HCV (29%) and HBV (71%) in the SHARP and Asia-Pacific trial, respectively [26]. In our study, these factors were not significantly associated with poor survival.

Furthermore, at least 28% of the patients would not have been included in the trials as they did not meet one or more eligibility criteria, mostly due to worse Child-Pugh classification. Specifically, this patient characteristic was associated with poor overall survival. Due to the small number of patients not meeting the other eligibility criteria, we were not able to determine their individual association with poor survival. However, the meta-analysis of McNamara et al. [27], also showing a negative association between OS and patients with poor performance status (ECOG PS >2) and Child-Pugh score B (vs A) in an Asian HCC patient population. They concluded that sorafenib treatment unlikely would be clinically meaningful for these patients, and our study result is in line with that conclusion.

Other real-world studies with sorafenib have been performed in Europe. The GIDEON study showed overall survival of 13.6 months for Child-Pugh A patients and 5.2 for Child-Pugh B patients, and this study also highlighted variations in HCC management and patient outcomes based on their region [28]. Furthermore, the observational INSIGHT study showed a median survival of 15.1 months in German and Austrian

HCC patients [29] and a multicenter prospective study in Italy showed a median survival of 10.5 months [30]. Our study adds the Dutch perspective on the treatment of HCC and insight in the patient population.

4.2 Other treatments

For a decade, sorafenib has been the only targeted therapy available, until the introduction of regorafenib in 2017 for second-line, and lenvatinib in 2018, for first-line treatment [31], and ultimately until the change of the ESMO guidelines to atezolizumab plus bevacizumab in 2021 for the first-line treatment [32]. These developments are reflected in the treatment patterns of our population, reflecting high use of first-line sorafenib in this population, and limited use of atezolizumab plus bevacizumab and lenvatinib. This study further shows that only 16% of the patients received second-line treatment, mostly regorafenib.

Limited other real-world studies are available reviewing the treatment patterns for HCC. Bonafede et al. [33] showed that between 2008 and 2015 in the USA, 77% of 1,459 patients were treated with sorafenib in the first-line, and 11% received a second-line treatment, mostly systemic chemotherapy, or other targeted therapies. A more recent Korean study from Novick et al. [33] showed that between 2016 and 2018 approximately 50% of the 127 patients who first received sorafenib had a second-line treatment, mostly regorafenib. In addition, Chen et al. [34] showed that between 2018 and 2021 in Taiwan 41 patients who were treated with atezolizumab plus bevacizumab, 73% received second-line systemic therapy, mostly sorafenib or lenvatinib. The last two studies show a significant rise in second-line treatment recently. This might indicate that the new treatment options enable more second- and third-line treatment use.

4.3 Limitations

In this study, data is extracted from EHR with a text-mining tool, with the advantage of a faster and more standardized data collection compared to manual extraction. However, the amount of data collected in this study is limited by the data stored in the EHR, and as the EHR is primarily used for healthcare purposes, the occurrence of missing data in observational retrospective research is common [35]. This is reflected by missing performance status, BCLC stage, and causes of disease. Furthermore, even though this study shows that selection of patients based on structured data is feasible, context analysis to ensure the indication and treatment use was still necessary, making

this a semi-automatic method. And, in this population, not enough patients were available to reflect on the population and outcomes of all treatments.

4.4 Future perspectives

With the evolving landscape of HCC treatment options, the need for real-world data and real-world evidence also rises, especially due to the complexity of the disease and potential regional differences [36]. In this study, we show that it is possible to collect relevant data on patients with HCC receiving systemic treatments from EHR with text mining and evaluate patterns, outcomes, and what patients might benefit. Therefore, comprehensive implementation of text-mining tools can benefit timely treatment evaluation and be used to further optimize HCC patient care.

5. Conclusion

This is the first real-world EHR text-mining study reviewing the population characteristics and outcomes of systemic treatments of HCC patients. It showed that in the period between 2015 and 2021, most patients received a first-line treatment of sorafenib. Furthermore, our population of first-line sorafenib users differed from the trial populations which overall did not result in different treatment outcomes. However, patients not meeting eligibility criteria, specifically Child-Pugh score B-C we associated with worse outcomes, and might not clinically benefit of treatment. At last, this study shows that the collection of real-world data for the evaluation of the current, but also upcoming HCC treatments can be performed more efficiently with text mining.

References

1. Integraal Kankercentrum Nederland. *NKR cijfers*. Available from: <https://iknl.nl/nkr-cijfers>.
2. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA: A Cancer Journal for Clinicians, 2021. **71**(3): p. 209-249.
3. Peck-Radosavljevic, M., et al., *Consensus on the current use of sorafenib for the treatment of hepatocellular carcinoma*. European Journal of Gastroenterology & Hepatology, 2010. **22**(4).
4. Díaz-González, Á., et al., *International and multicenter real-world study of sorafenib-treated patients with hepatocellular carcinoma under dialysis*. Liver International, 2020. **40**(6): p. 1467-1476.
5. Keating, G.M., *Sorafenib: A Review in Hepatocellular Carcinoma*. Target Oncol, 2017. **12**(2): p. 243-253.
6. Petracci, F., et al., *Use of real-world evidence for oncology clinical decision making in emerging economies*. Future Oncology, 2021. **17**(22): p. 2951-2960.
7. Suresh, D., A.N. Srinivas, and D.P. Kumar, *Etiology of Hepatocellular Carcinoma: Special Focus on Fatty Liver Disease*. Frontiers in Oncology, 2020. **10**.
8. Konyn, P., A. Ahmed, and D. Kim, *Current epidemiology in hepatocellular carcinoma*. Expert Rev Gastroenterol Hepatol, 2021. **15**(11): p. 1295-1307.
9. Kulik, L. and H.B. El-Serag, *Epidemiology and Management of Hepatocellular Carcinoma*. Gastroenterology, 2019. **156**(2): p. 477-491.e1.
10. 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.
11. van Laar, S.A., et al., *An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma*. Clin Pharmacol Ther, 2020. **108**(3): p. 644-652.
12. van Laar, S.A., et al., *Real-World Metastatic Renal Cell Carcinoma Treatment Patterns and Clinical Outcomes in The Netherlands*. Frontiers in Pharmacology, 2022. **13**.
13. Llovet, J.M., et al., *Sorafenib in advanced hepatocellular carcinoma*. N Engl J Med, 2008. **359**(4): p. 378-90.
14. Cheng, A.L., et al., *Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial*. Lancet Oncol, 2009. **10**(1): p. 25-34.
15. European Medicines Agency. *CHMP post-authorisation summary of positive opinion of stivarga*. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-stivarga-ii-20_en.pdf.
16. Bruix, J., et al., *Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial*. The Lancet, 2017. **389**(10064): p. 56-66.
17. European Medicines Agency. *CHMP post-authorisation summary of positive opinion of lenvima*. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-lenvima_en.pdf.
18. Kudo, M., et al., *Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial*. The Lancet, 2018. **391**(10126): p. 1163-1173.

19. European Medicines Agency. *CHMP post-authorisation summary of positive opinion of Cabometyx*. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-cabometyx_en.pdf.
20. Abou-Alfa, G.K., et al., *Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma*. *New England Journal of Medicine*, 2018. **379**(1): p. 54-63.
21. European Medicines Agency. *CHMP post-authorisation summary of positive opinion of cyramza*. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-cyramza-ii-27_en.pdf.
22. Zhu, A.X., et al., *Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α -fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial*. *The Lancet Oncology*, 2019. **20**(2): p. 282-296.
23. European Medicines Agency. *CHMP post-authorisation summary of positive opinion of Tecentriq*. Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-tecentriq-ii-39_en.pdf.
24. Finn, R.S., et al., *Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma*. *New England Journal of Medicine*, 2020. **382**(20): p. 1894-1905.
25. Tsores, A. and C.A. Marlar, *Use Of The Child Pugh Score In Liver Disease*, in *StatPearls*. 2022, StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.: Treasure Island (FL).
26. Marot, A., et al., *Alcoholic liver disease confers a worse prognosis than HCV infection and non-alcoholic fatty liver disease among patients with cirrhosis: An observational study*. *PLOS ONE*, 2017. **12**(10): p. e0186715.
27. McNamara, M.G., et al., *Sorafenib as first-line therapy in patients with advanced Child-Pugh B hepatocellular carcinoma—a meta-analysis*. *European Journal of Cancer*, 2018. **105**: p. 1-9.
28. Kudo, M., et al., *Regional differences in sorafenib-treated patients with hepatocellular carcinoma: GIDEON observational study*. *Liver Int*, 2016. **36**(8): p. 1196-205.
29. Ganten, T.M., et al., *Sorafenib in Patients with Hepatocellular Carcinoma-Results of the Observational INSIGHT Study*. *Clin Cancer Res*, 2017. **23**(19): p. 5720-5728.
30. Iavarone, M., et al., *Field-practice study of sorafenib therapy for hepatocellular carcinoma: a prospective multicenter study in Italy*. *Hepatology*, 2011. **54**(6): p. 2055-63.
31. Huang, A., et al., *Targeted therapy for hepatocellular carcinoma*. *Signal Transduction and Targeted Therapy*, 2020. **5**(1): p. 146.
32. Vogel, A., et al., *Updated treatment recommendations for hepatocellular carcinoma (HCC) from the ESMO Clinical Practice Guidelines*. *Annals of Oncology*, 2021. **32**(6): p. 801-805.
33. Bonafede, M.M., et al., *Treatment Patterns and Economic Burden by Lines of Therapy Among Patients with Advanced Hepatocellular Carcinoma Treated with Systemic Cancer Therapy*. *J Gastrointest Cancer*, 2020. **51**(1): p. 217-226.
34. Chen, C.-T., et al., *Prognosis and treatment pattern of advanced hepatocellular carcinoma after failure of first-line atezolizumab and bevacizumab treatment*. *Hepatology International*, 2022.
35. Jazayeri, A., O.S. Liang, and C.C. Yang, *Imputation of Missing Data in Electronic Health Records Based on Patients' Similarities*. *Journal of Healthcare Informatics Research*, 2020. **4**(3): p. 295-307.
36. Liu, L., S. Qin, and Y. Zhang, *The Evolving Landscape of Checkpoint Inhibitor Combination Therapy in the Treatment of Advanced Hepatocellular Carcinoma*. *Targeted Oncology*, 2021. **16**(2): p. 153-163.

Supplementary data

Supplementary Table S5.1A. Patient inclusion

Inclusion criterium	type of data	Searched terms	Time window	Comments
DBC-code*	DBC	specialism: 0313 (internal medicine) with diagnosis: 955 (Liver tumor) specialism: 0303 (surgery) with diagnosis: 955 (malignant neoplasm liver) specialism: 0318 (gastro-enterology) with diagnosis: 955 (Liver malignancy) specialism: 0318 (gastro-enterology) with diagnosis: 955 (recipient liver transplant)		Diagnose behandel combinatie = diagnosis treatment combination, a code used for reimbursements in the Netherlands
Systemic treatment use	medication request	prescription of atezolizumab, bevacizumab, cabozantinib, lenvatinib, regorafenib or sorafenib	>01-01-2015	
Verification hepatocellular carcinoma	Free-text search	liver carcinoma, hepatocellular carcinoma, hepatocarcinoma, hepatoma, carcinoma epatocellular, lcc, hcc, carcinoom hepatocellulair, hepatoom, carcinom lever, levercarcinoom, carcinoom van de lever, hepatocellulair carcinom, levercelcarcinoom, leverkanker, levercelca, lever kanker, levercelkanker, hepatocellulair ca, leverca		

Supplementary Table S5. 1B. Data collection

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to
Start date treatment	Medication request		Start date of eldest prescription		
Stop date treatment	Medication request		Start date of most recent prescription		
Reason stop	Text search	<u>Multiple choice options</u> - Progression - Adverse events - Deceased - Other - Unclear - Did not end treatment	Multiple choice option + date of selected text source	General: Behandeling staken OR Sorafenib OR Stoppen OR Staken OR Discontinueren OR Niet continueren OR Gestaakt OR Gestopt OR Staak OR Stop OR Nexavar OR Tumorprogressie OR Neoplasma OR Progressie OR tumorgroei OR Groei tumor OR Nieuwe laesie OR Nieuwe laesies OR groei tumoren OR Groei van tumor OR Groei van tumoren OR toename tumor OR toename tumorweefsel OR PD OR Progressive Disease OR Progressieve ziekte	Consult OR consulten
Radiology report	Text search		Date of selected report	Tumorprogressie OR Neoplasma Progressie OR tumorgroei OR Groei tumor OR Nieuwe laesie OR Nieuwe laesies OR groei tumoren OR Groei van tumor OR Groei van tumoren OR toename tumor OR toename tumorweefsel OR PD OR Progressive Disease OR Progressieve ziekte	Radiology reports
BMI - weight	Measurement		Numeric value	Gewicht	
BMI - Length	Measurement		Numeric value	Lengte	

Performance status	Text search	<u>Multiple choice options</u> - 0 - 1 - 2 - 3	Multiple choice option	Performance OR Karnofsky performance status OR karnofsky index OR karnofsky scale OR karnofsky OR karnofsky score OR karnofskyindex OR KPS OR WHO performance OR WHO OR who klasse OR who score OR exercise tolerance OR inspanningstolerantie OR ecog performance status OR ecog score OR ecog
Child-Pugh score	Text search	<u>Multiple choice options</u> - A - B - C	Multiple choice option	Child-pugh Score Class A OR CP-A OR Child-turcotte-pugh Class A OR Child Pugh A OR CP A OR cpt A OR Child-Pugh A OR Child-Pugh score A OR Child Pugh score A OR Child Pugh klasse A OR Child-turcotte-pugh Klasse A OR Child-Pugh klasse A Childpugh A OR CTP A OR CTP-A OR CPA OR Child A Child-pugh Score Class B OR CP-A OR Child-turcotte-pugh Class B OR Child Pugh B OR CP B OR cpt B OR Child-Pugh B OR Child-Pugh score B OR Child Pugh score B OR Child Pugh klasse B OR Child-turcotte-pugh Klasse B OR Child-Pugh klasse B Childpugh B OR CTP B OR CTP-A OR CPA OR Child B Child-pugh Score Class C OR CP-A OR Child-turcotte-pugh Class C OR Child Pugh C OR CP C OR cpt C OR Child-Pugh C OR Child-Pugh score C OR Child Pugh score C OR Child Pugh klasse C OR Child-turcotte-pugh Klasse C OR Child-Pugh klasse C Childpugh C OR CTP C OR CTP-A OR CPA OR Child C
Ascites	Text search	<u>Multiple choice options</u> - Yes - No - Unclear	Multiple choice option	Ascites OR abdomen Ascites OR Ascitesvocht OR peritoneal fluid OR peritoneal effusion OR ascitic fluid OR peritoneaal vocht OR drainage of ascites OR Ascitesdrainage OR ascitesdrainages OR ascites drainage OR drainage ascites OR Ascitis OR Buikvocht OR Vochttophopping

Supplementary Table S5. 1B continues on next page.

Supplementary Table S5.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to
Encephalopathy	Text search	<u>Multiple choice options</u> - Yes - No - Unclear	Multiple choice option	Hepatic encephalopathy OR portosystemic encephalopathy OR hepatisch encefalopathie OR hepatische encefalopathie OR encefalopathtie hepatisch OR encefalopathie hepatische OR HE OR Encephalopathies OR encephalopathy OR encefalopathieen OR encefalopathie OR Delirium OR Delirious OR ABS OR delirant OR acuut delier OR delier acuut OR delirante OR sufheid OR comatose OR coma OR comateus OR lethargy OR lethargic OR lethargisch OR letargie OR disorientation OR disorientated OR gedesoriënteerd OR desorientatie OR gestoorde orientatie OR confusion OR confused OR mental confusion OR confusional state OR delirium confusional state OR verwardheid OR verward OR verwarde OR verwardheidstoestand OR verwarring OR verwarde OR fuzzy OR warrig hoofd OR asterixis OR flapping tremor OR liver flap OR gedaald bewustzijn OR ataxie OR ataxia OR dyssynergia OR ataxy OR ataxie OR dyssynergie OR coordinatiestoornis OR dysarthria OR dysartrie OR somnolence OR somnolentie	
BCLC stage	Text search	<u>Multiple choice options</u> - Early - Intermediate - Advanced - Unclear	Multiple choice option	BCLC OR stage OR intermediate OR advanced	
Albumin	Measurement		Numeric value	Albumine	
Alpha-feto protein	Measurement		Numeric value	Alfa-foetoproteïne OR AFP	
ALT	Measurement		Numeric value	ALAT GPT OR ALAT	
AST	Measurement		Numeric value	ASAT GOT OR ASAT	

Bilirubin	Measurement	Numeric value	Bilbubine totaal OR totaal bilirubine
Creatinine	Measurement	Numeric value	Kreatinine
Hemoglobin	Measurement	Numeric value	Hb OR hemoglobine
PT-INR	Measurement	Numeric value	INR
Platelet count	Measurement	Numeric value	Thrombocyten
Hepatectomy	Text search	Multiple choice options - Yes - No - Unclear	Hepatectomy OR leverresectie OR hepatectomie OR resectie lever OR lever resectie OR resection of liver OR liver excision OR resectie OR excisie OR verwijdering OR ectomie OR lobectomy OR segmentectomy OR hemihepatectomy OR hemihepatectomie OR hemi hepatectomie OR liver transplant OR levertransplantaat OR levertransplantatie OR levertransplantaties OR lever transplantatie OR transplantaat lever
Embolization	Text search & Medication request	Multiple choice options - Yes - No - Unclear	Transarterial chemoembolization OR hepatic artery chemoembolization OR TACE OR HACE OR transkatheter arteriele chemo-embolisatie OR chemo-embolisatie OR chemoembolisatie OR TARE OR radio-embolisatie OR radioembolisatie Mitomycine OR cisplatine OR doxorubicine OR yttrium y 90 verbindingen
Ablation	Text search	Multiple choice options - Yes - No - Unclear	Radiofrequency ablation OR radiofrequente ablatie OR RFA OR ablatie OR thermoablatie OR microwave ablation OR MWA or microwave ablatie

Supplementary Table S5.1B continues on next page.

Supplementary Table S5.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to
Alcoholic liver disease	Text search & Diagnosis code	<u>Multiple choice options</u> - Yes - No - Unclear	Multiple choice option	Alcoholic liver disease OR ALD OR alcohol lever OR alcoholische leverziekte OR alcoholische leverziekten OR lever alcohol OR chronic alcoholic hepatitis OR chornische alcoholhepatitis OR hepatitis alcoholic or alcoholic hepatitis OR alcoholhepatitis OR alcoholische hepatitis OR hepatitis door alcoholgebruik OR hepatitis alcoholische OR hepatitis alcoholisch OR alcohol hepatitis OR hepatitis ethanol OR alcoholisch misbruik alcohol OR alcoholabuses OR alcoholisme OR alcohol-abusus OR probleem alcoholgebruik OR alcohol misbruik OR alcohol gebruik OR alcoholgebruik OR gebruik alcohol OR alcohol OR alcohol use disorder OR liver cirrhosis alcoholic OR alcoholic cirrhosis OR cirrose door alcoholgebruik OR levercirrose alcoholische OR alcoholische levercirrose OR cirrose alcoholisch OR alcohol cirrose OR alcoholische cirrose OR alcoholic fatty liver disease OR alcoholic fatty liver OR alcoholische vetlever OR alcoholische leververvetting Code for "Alcoholische levercirrose"	
Hepatitis B	Text search & Diagnosis code	<u>Multiple choice options</u> - Yes - No - Unclear	Multiple choice option	Hepatitis B OR hepatitis B virus OR hepatitis B infection OR serum hepatitis OR viral hepatitis B OR hepatitis B virus infection OR hepatitis serum OR virale hepatitis B OR hep B OR hepb OR hepatitisb OR B hepatitis OR HBV OR hepa B Code for: hepatitis B chronisch OR hepatitis B OR virus hepatitis B	
Hepatitis C	Text search & Diagnosis code	<u>Multiple choice options</u> - Yes - No - Unclear	Multiple choice option	Hepatitis C OR viral hepatitis C OR hepatitis C infection Hepatitis non-a non-b OR virale hepatitis C OR non-a-non-b-hepatitis OR hep C OR hepc OR C hepatitis OR HCV Code for: chronische hepatitis C OR hepatitis C OR virus hepatitis C OR hepatitis C chronisch	

Metabolic syndrome	Text search & diagnosis code	<u>Multiple choice options</u> - Yes - No - Unclear	Nonalcoholic fatty liver disease OR non alcoholic fatty liver disease OR non-alcoholic fatty liver OR nonalcoholic fatty liver OR NAFLD OR niet-alcoholische leververvetting OR nonalcoholic steatohepatitis OR non-alcoholic steatohepatitis OR NASH OR niet alcoholische steatohepatitis Code for: niet-alcoholische vetlever NAFLD OR steatosis hepatis/ NASH OR chron niet-alcoholische leverziekte overige OR chron.niet-alcoholische leveraandoening/overig
Hemochromatosis	Text search	<u>Multiple choice options</u> - Yes - No - Unclear	Hemochromatosis OR diabetes bronze OR bronze diabetes OR hemochromatose OR alfa 1-antitrypsine deficientie OR serum antitrypsinedeficientie

Part III

Real-world treatment safety

Chapter 6

Application of electronic health record text mining: Real-world tolerability, safety, and efficacy of adjuvant melanoma treatments

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

Cancers. 2022; 14(21): 5426

Abstract

Introduction: Nivolumab (N), pembrolizumab (P), and dabrafenib plus trametinib (D+T) are registered as adjuvant treatments for resected stage III and IV melanoma since 2018. Electronic health records (EHR) are a real-world data source that can be used to review treatments in clinical practice. In this study, we applied EHR text-mining software to evaluate the real-world tolerability, safety, and efficacy of adjuvant melanoma treatments.

Methods: Adult melanoma patients receiving adjuvant treatment between January 2019 and October 2021 at the Leiden University Medical Center, The Netherlands, were included. CTcue text-mining software was used to construct rule-based queries and perform context analysis for patient inclusion and data collection from structured and unstructured EHR data.

Results: In total, 122 patients were included: 54 patients treated with nivolumab, 48 with pembrolizumab, and 20 with D+T. Significantly more patients discontinued treatment due to toxicity on D+T (N: 16%, P: 6%, D+T: 40%), $X^2(6, n=122) = 14.6$, $p=0.024$. Immune checkpoint inhibitors (ICI) mainly showed immune-related treatment-limiting adverse events (AEs), and chronic thyroid-related AE occurred frequent (hyperthyroidism: N: 15%, P: 13%, hypothyroidism: N: 20%, P: 19%). Treatment-limiting toxicity from D+T was primarily a combination of reversible AEs, including pyrexia and fatigue. The 1-year recurrence-free survival was 70.3% after nivolumab, 72.4% after pembrolizumab, and 83.0% after D+T.

Conclusions: Text mining EHR is a valuable method to collect real-world data to evaluate adjuvant melanoma treatments. ICI were better tolerated than D+T, in line with RCT-results. For BRAF+-patients, physicians must weigh the higher risk of reversible treatment-limiting AEs of D+T against the risk of long-term immune-related AEs.

1. Introduction

In 2020, approximately 300,000 patients worldwide were diagnosed with melanoma of the skin, accounting for 1.7% of all cancer diagnoses [1]. In the Netherlands, the incidence of melanoma has more than doubled in the last 20 years from 18/100,000 people in 2001 to 43/100,000 in 2021 with a mortality rate of 4–5/100,000 [2]. The introduction of immune checkpoint inhibitors (ICI) (e.g., nivolumab, pembrolizumab and ipilimumab) and inhibitors of the mitogen-activated protein kinase pathway (e.g. BRAF inhibitors and MEK inhibitors) have improved the treatment of metastatic melanoma in the past years [3]. In 2018, nivolumab (N), pembrolizumab (P), and the combination of dabrafenib plus trametinib (D+T) were registered by the European Medicines Agency as adjuvant treatment in resected stage III and IV of melanoma [4–7]. Patients receive adjuvant treatment after surgical resection, for a maximum of 12 months or until treatment-limiting toxicity or recurrence of the disease.

The results of the phase III trials that were the basis for the indication expansion are summarized in Table 6.1. Eligibility criteria differed between trials, e.g., the inclusion of resected stage IV in the CheckMate 238 trial, and inclusion only of patients with a BRAF V600E or V600K mutation in the COMBI-AD trial [8]. Furthermore, all treatments were superior to their comparator regarding recurrence-free survival but concluding overall survival results are yet unknown.

As these patients are, in principle, cured after surgery, the safety profile of adjuvant therapy may be even more relevant than in the setting of palliative treatment. Regarding safety, nivolumab and pembrolizumab are comparable, but differ from dabrafenib plus trametinib treatment. Grade 3 or 4 adverse events (AEs) occurred in 25.4% and 31.6% of the study populations of nivolumab and pembrolizumab versus 41% in dabrafenib plus trametinib treated patients, and 9.7% and 13.8% of the patients discontinued treatment due to toxicity, compared to 26% in the dabrafenib and trametinib treated. However, most reported AEs on ICI were skin reactions and fatigue. All immune-related (ir)AEs combined also have a high incidence, of which hypothyroidism and hyperthyroidism – both manifestations of thyroiditis – were most frequent, but also include, e.g., diabetes mellitus type I [9, 11]. Of these immune-related endocrine toxicities it is presumed they result in permanent and irreversible dysfunction, resulting in lifelong hormone supplementation [14]. This contrasts with the AEs of dabrafenib plus trametinib, of which pyrexia, fatigue and nausea are most common and easily reversible after treatment interruption [12, 15].

Table 6.1. Phase III trial results for adjuvant treatments of melanoma

	Nivolumab	Pembrolizumab	Dabrafenib plus trametinib
Phase III trial	Checkmate 238 [8, 9]	EORTC/KEYNOTE-054 [10, 11]	COMBI-AD [12, 13]
Comparator	Ipilimumab	Placebo	Placebo
Eligibility criteria	Resected stage IIIb, IIIc and IV melanoma, ECOG PS: 0 or 1	Resected stage IIIa, IIIb, and IIIc melanoma, ECOG PS: 0 or 1	Resected stage IIIa, IIIb, or IIIc melanoma with a BRAF V600E or V600K mutation, ECOG PS: 0 or 1
Recurrence-free survival	1-year NIV: 71% IPI: 61% 4-year NIV: 51.7% IPI: 41.2% HR: 0.71, 95% CI = 0.60–0.86	1-year PEM: 75% PLA: 61% 3.5-year PEM: 60% PLA: 41% HR: 0.59, 95% CI = 0.49–0.70	1-year D+T: 88% PLA: 56% 5-year D+T: 52% PLA: 36% HR: 0.51, 95% CI = 0.42–0.61
Overall survival	4-year NIV: 77.9% IPI: 76.6% HR: 0.87 95% CI = 0.66–1.14	-	3-year D+T: 86% PLA: 77% HR: 0.57 95% CI = 0.42–0.79*
≥ Grade 3 AE	NIV: 25.4% IPI: 55.2%	PEM: 31.6% PLA: 18.5%	D+T: 41% PLA: 14%
Most common AE	Skin reactions: 44.5% Fatigue: 34.5% Gastrointestinal: 25.2%	All immune-related AE: 37.3% Fatigue or asthenia: 37.1% Skin reactions: 28.3%	Pyrexia: 63% Fatigue: 47% Nausea: 40%
AE leading to discontinuation	NIV: 9.7% IPI: 42.6%	PEM: 13.8% PLA: 2.2%	D+T: 26% PLA: 3%

Abbreviations: AE: adverse event, D+T: dabrafenib plus trametinib, ECOG PS: Eastern Cooperative Oncology Group Performance Status, NIV: nivolumab, PEM: pembrolizumab, PLA: placebo, 95% CI: 95% confidence interval.

* did not meet the prespecified interim analysis boundary of $p=0.000019$.

Since RCT results may not represent the benefits and risks of treatments in clinical practice, real-world data can add insightful information and support decision making [16]. The electronic health record (EHR) is one of the sources that contains relevant real-world data for cancer treatment evaluation, since it includes, e.g., hospital visits, patient demographics, medication orders, laboratory data, vital signs, and imaging results [17]. However, as most data is captured in free-text notes, manual chart review is still the standard method for data extraction from EHR, which is very labor intensive and time-consuming. Previously we validated a text-mining tool to extract data

from EHR for the evaluation of metastatic renal cell carcinoma treatments (mRCC) and showed that this method is accurate and seven times faster than manual data extraction [18]. This tool has already been used in other real-world studies, e.g., to review mRCC treatment patterns and outcomes in two hospitals, to review the use of granulocyte-colony stimulating factor and incidence of febrile neutropenia in breast cancer patients, and the identification of treosulfan-induced myalgia in a pediatric hematopoietic stem cell transplantation patients [19-21].

The aim of this study was to apply a text-mining tool to retrospectively review the tolerability, safety, and efficacy of new adjuvant treatments for melanoma in a Dutch clinical hospital setting.

2. Methods

We performed a retrospective EHR review by collecting data from EHR through text mining. The study protocol was reviewed and approved by the Medical Ethics Review Committee of the Leiden University Medical Center (LUMC), Leiden, The Netherlands, which waived the need for informed consent.

2.1 Electronic health record text mining

We retrospectively identified patients and collected all data from the electronic health record with text-mining software (CTcue B.V., Amsterdam, The Netherlands). This software program enables rule-based text mining of the EHR, which can extract both data from structured data (e.g., medical prescriptions and laboratory values) and unstructured, free-text notes (e.g., medical notes and correspondence). Previously, we validated this tool and method to collected real-world data of renal cell carcinoma treatments in clinical practice [18].

Figure 6.1 shows the four steps taken within the program, per tab, from start until the export of the final dataset. First, rule-based queries were constructed in the criteria tab, to select the correct patient population. Secondly, the cohort tab gives a brief overview of the selected patients and the criteria on which they were included. The context of a datapoint can be further expanded by selecting the datapoint; in case of free-text notes, the system shows the complete form or report with the specified key word(combination)s. An overview of all cases that matched the search results, e.g., all

free-text forms with a hit, added to all structured data points, e.g., lab values between a specific period, is provided. The patient cohort is confirmed by reviewing all hits in unstructured text. In case of missing data necessary for patient inclusion, a patient is not included. Similarly, rule-based queries were constructed for the collection of data in the questions tab and all data from unstructured text was confirmed per patient by context analysis.

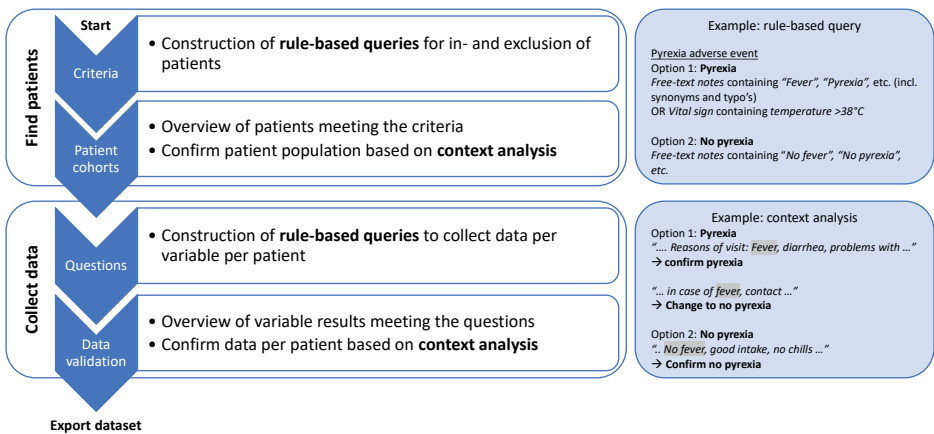


Figure 6.1. Steps taken to find patients and collect data with the text-mining tool, including the construction of rule-based queries and context analysis.

2.2 Patient population

Patients were included in a university medical center, the LUMC, Leiden, The Netherlands. We aimed to include all patients aged 18 years and older with melanoma if they started adjuvant treatment of nivolumab or pembrolizumab, and dabrafenib plus trametinib between January 2019 and October 2021. Therefore, the search criteria included: 1. Patients who received nivolumab, pembrolizumab or dabrafenib plus trametinib, 2. Patients with a reimbursement code (diagnosis treatment code) specific for melanoma, 3. A free-text note that confirmed use of the treatment (e.g., search for terms as “nivolumab”, “nivo”, “Opdivo®”), 4. A free-text note that confirmed the adjuvant aspect of the treatment (e.g., “stage III”, “adjuvant”). The complete list of criteria is available in Supplementary File S6.1.

2.3 Data collection

The following patient characteristics were collected at start of treatment: age, sex, performance status, disease stage (according to the 7th edition of the American Joint Committee on Cancer [AJCC]), primary tumor location, subtype, brain metastases, lactate dehydrogenase (LDH), ulceration status, and BRAF-, NRAS- and KIT mutation status. Also, the following outcomes were collected: the time on treatment, the reason to end the treatment, most common AEs as reported in the RCTs (abdominal pain, arthralgia, asthenia, chills, cough, diarrhea, dyspnea, fatigue, headache, nausea, pruritus, pyrexia, and rash, with additionally irAEs (colitis, diabetes mellitus type 1, hepatitis, hyperthyroidism, hypothyroidism, pneumonitis)), all treatment-limiting AEs, and recurrence of melanoma. The common terminology criteria for adverse events (CTCAE) v5.0 were used to grade the severity of adverse events in the context analysis.

2.4 Statistical analysis

Data management and analysis were performed using R Statistical Software (v4.1.1; R CoreTeam 2021). We used descriptive statistics to describe the patient- and disease characteristics. The time on treatment and time until recurrence per patient were visualized in a swimmer plot and the median time on treatment per reason to end treatment was calculated. Chi-square test was performed to test for differences in tolerability (treatment-limiting toxicity) between treatments. Per treatment, the RFS was analyzed with the Kaplan-Meier method and visualized in a survival plot. All statistical analyses were exploratory.

3. Results

By text mining, 217 patients with melanoma were identified in the EHR. Patients who received their systemic treatment as palliative treatment (n=93) were excluded after context analysis (Figure 6.2).

In total, 122 melanoma patients were included in the study, 54 patients started with nivolumab, 48 with pembrolizumab and 20 patients with dabrafenib plus trametinib between January 2019 and October 2021. The median age of the patients was 59 years (range: 21–84), and more than half of the patients was male (61.5%). All baseline patient- and disease characteristics are shown in Table 6.2.

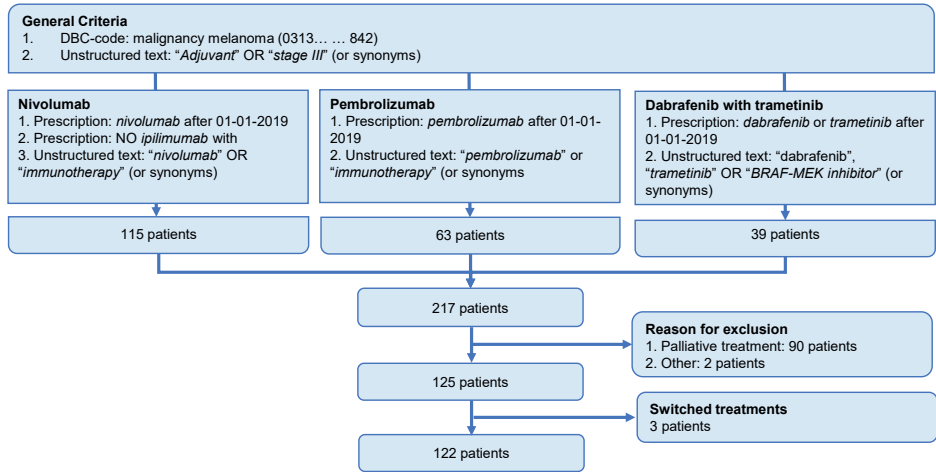


Figure 6.2. Patient flow chart. DBC-code: diagnosis-treatment combination code.

3.1 Time on treatment

The adjuvant treatment was ended for 97 (79%) of the patients, 25 patients were still on treatment at the time of data extraction (Figure 6.3). The mean time on treatment of patients, who ended treatments was 10.2 months for nivolumab (interquartile range [IQR]: 6.4–12.0), 11 months for pembrolizumab (IQR: 4.2–12.4), and 8.4 months for dabrafenib plus trametinib (IQR: 2.2–11.6).

3.2 Adverse events

The reported AEs are summarized in Table 6.3. Fatigue was the most reported AE in all groups (N: 77.8%, P: 75.0%, D+T: 80.0%), followed by diarrhea (44.5%) and headache (29.6%) for patients treated with nivolumab; diarrhea (23.0%), and nausea, pruritus, and hypothyroidism (all 18.8%) for patients treated with pembrolizumab, and pyrexia (65.0%) and nausea (55.0%) for dabrafenib plus trametinib treatment. The adverse events include the following severe AEs: hepatitis (3.7%), asthenia (1.9%) and diarrhea (1.9%) on nivolumab, diarrhea (4.2%) and colitis (2.1%) on pembrolizumab, and nausea, incl. vomiting (5.0%) on dabrafenib plus trametinib.

3.3 Treatment-limiting toxicity

In total, in twenty (16.4%) patients, treatment was ended early due to toxicity. The reason to end the adjuvant treatment due to toxicity was significantly different between the treatment groups, with more patients ending treatment in the dab-

rafenib plus trametinib group (8/15), than patients treated with nivolumab (9/46), and pembrolizumab (3/36), X^2 (6, $n=122$) =14.6, $p=0.024$. Eight patients stopped within the first three months (N: 2, P: 1, D+T: 5) (Figure 6.3), and in this group, in total, 28 treatment-limiting AEs were reported (Table 6.4). IrAEs were most reported as treatment-limiting in patients treated with nivolumab (8/9) and pembrolizumab (2/3), including hepatitis, colitis, pneumonitis, and thyroiditis. Pyrexia (5/8), skin disorders (4/8), chills and nausea (both 3/8) were most reported in patients treated with dabrafenib plus trametinib ($n=8$). An overview of treatment-limiting AEs per patient is available in Supplementary File S6.2.

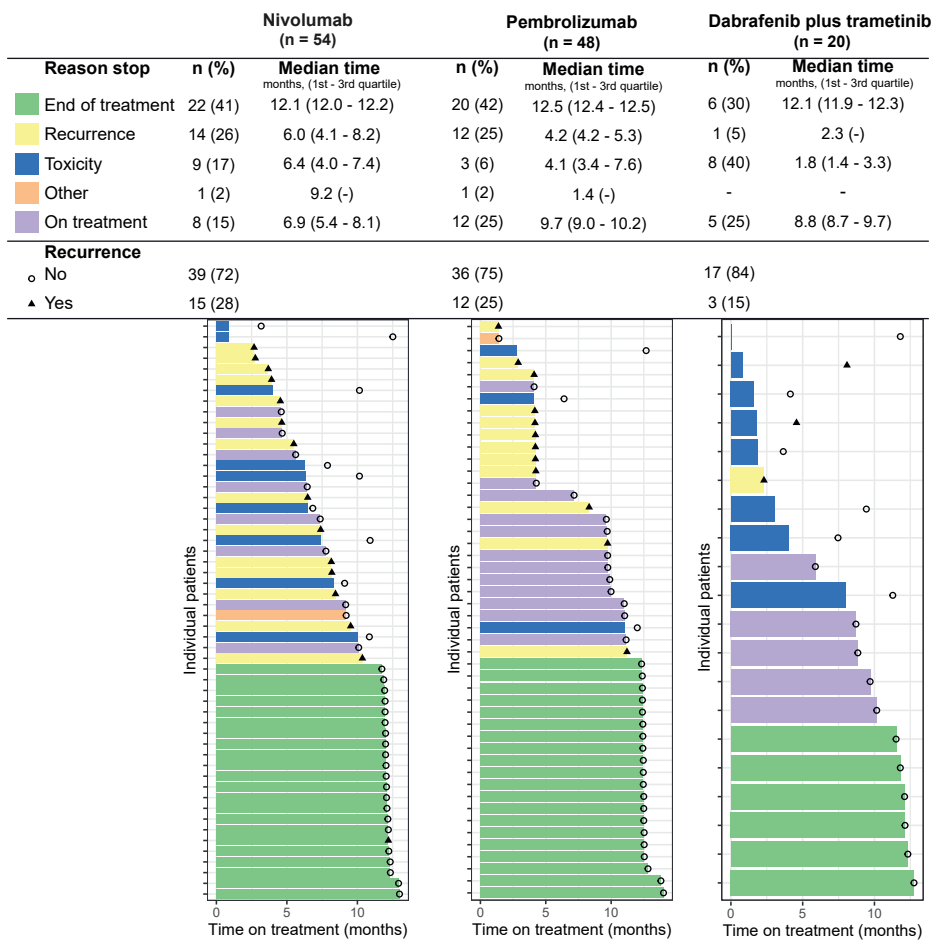


Figure 6.3. Patients' time on treatment, causes of treatment discontinuation, and recurrence per treatment.

Table 6.2. Patient characteristics

Characteristics	Nivolumab (n=54)	Pembrolizumab (n=48)	Dabrafenib plus trametinib (n=20)	All patients (n=122)
Median age (range) – year	61 (21–84)	57.5 (23–80)	57.5 (31–73)	59 (21–84)
Sex, no. (%)				
Male	35 (64.8)	31 (64.6)	9 (45.0)	75 (61.5)
Female	19 (35.2)	17 (35.4)	11 (55.0)	47 (38.5)
ECOG performance status, no. (%)				
0	32 (59.3)	24 (50.0)	9 (45.0)	65 (53.3)
1	7 (13.0)	5 (10.4)	1 (5.0)	13 (10.7)
Unknown	15 (27.8)	18 (37.5)	10 (50.0)	43 (35.2)
Disease stage, no. (%) AJCC 7				
III	34 (63.0)	48 (100)	20 (100)	102 (83.6)
Unspecified	3 (5.6)	1 (2.1)	1 (5.0)	5 (4.1)
IIIa	1 (1.9)	20 (41.7)	5 (25.0)	26 (21.3)
IIIb	10 (18.5)	19 (39.6)	9 (45.0)	38 (31.1)
IIIc	20 (37.0)	8 (16.7)	5 (25.0)	33 (27.0)
Resected IV	20 (37.0)	-	-	20 (16.4)
Primary tumor location, no. (%)				
Head or neck	6 (11.1)	6 (12.5)	2 (10.0)	14 (11.5)
Body	19 (35.2)	22 (45.8)	8 (40.0)	49 (40.2)
Extremities	17 (31.5)	13 (27.1)	6 (30.0)	36 (29.5)
Acral	1 (1.9)	3 (6.3)	1 (5.0)	5 (4.1)
Mucosal	1 (1.9)	-	-	1 (0.8)
Unknown primary	9 (16.7)	2 (4.2)	3 (15.0)	14 (11.5)
Unknown	1 (1.9)	2 (4.2)	-	3 (2.5)
Subtype, no. (%)				
Superficial spreading	16 (29.6)	25 (52.1)	11 (55.0)	52 (42.6)
Nodular	9 (16.7)	8 (16.7)	5 (25.0)	22 (18.0)
Acral lentiginous	-	2 (4.2)	1 (5.0)	3 (2.5)
Spindle cell	-	3 (6.3)	-	3 (2.5)
Unclear	29 (53.7)	10 (20.8)	3 (15.0)	42 (34.4)
LDH above ULN, no. (%)				
Yes	5 (9.3)	-	-	5 (4.1)
No	43 (79.6)	44 (91.7)	20 (100.0)	107 (87.7)
Unknown	6 (11.1)	4 (8.3)	-	10 (8.2)
Ulceration, no. (%)				
Yes	-	15 (31.3)	1 (5.0)	16 (13.1)
No	14 (25.0)	25 (52.1)	5 (25.0)	44 (36.1)
Unknown	40 (74.1)	8 (16.7)	14 (70.0)	62 (50.8)

Table 6.2 continues on next page.

Table 6.2. *Continued*

Characteristics	Nivolumab (n=54)	Pembrolizumab (n=48)	Dabrafenib plus trametinib (n=20)	All patients (n=122)
BRAF mutation , no. (%)				
Yes	28 (51.9)	17 (35.4)	20 (100.0)	65 (53.3)
No	21 (38.9)	20 (41.7)	-	41 (33.6)
Unknown	5 (9.3)	11 (22.9)	-	16 (13.1)
NRAS mutation , no. (%)				
Yes	18 (33.3)	12 (25.0)	1 (5.0)	31 (25.4)
No	28 (51.9)	13 (27.1)	15 (75.0)	56 (45.9)
Unknown	8 (14.8)	23 (47.9)	4 (20.0)	35 (28.7)
KIT mutation , no. (%)				
Yes	-	1 (2.1)	-	1 (0.8)
No	45 (83.3)	31 (64.6)	17 (85.0)	93 (76.2)
Unknown	9 (16.7)	16 (33.3)	3 (15.0)	28 (23.0)
Previous systemic therapy for melanoma , no. (%)				
Adjuvant	1 (1.9)	1 (2.1)	1 (5.0)	5 (4.1)
Neo-adjuvant	1 (1.9)	-	3 (15.0)	3 (2.5)

3.4 Recurrence

In total 28 patients (23.0%) ended treatment due to recurrence, 15 of the patients were treated with nivolumab, 12 with pembrolizumab and 1 with dabrafenib plus trametinib (Figure 6.3). The recurrence probability at 1 year was 70.3% (95% CI: 58.1–85.0) for nivolumab, 72.4% (95% CI: 60–87.3) for pembrolizumab, and 83.0% (95% CI: 67.1–1) for dabrafenib plus trametinib (Figure 6.4). The No median RFS was reached for any of these treatments within the maximum follow-up period of 13 months.

4. Discussion

In an effort to optimize data extraction from EHRs and to evaluate the tolerability, safety and efficacy of nivolumab, pembrolizumab and dabrafenib plus trametinib as adjuvant treatments for resected stage III and IV melanoma, we performed a retrospective study by text mining the EHR in a university hospital. This is the first real-world study on adjuvant treatment in resected stage III and IV melanoma with retrieval of clinical data by text mining. By collecting a sufficient amount of data, this study shows that text mining EHR is a valuable and effective new method. The majority of the 122 included patients received ICI (N: 45%, P: 39%), and 16% of our patients received dabrafenib

Table 6.3. Adverse events

Adverse events, no. (%)	Nivolumab (n=54)		Pembrolizumab (n=48)		Dabrafenib plus trametinib (n=20)	
	Grade 1–2	Grade 3–4	Grade 1–2	Grade 3–4	Grade 1–2	Grade 3–4
Abdominal pain	2 (3.7)	-	4 (8.3)	-	1 (5.0)	-
Arthralgia	2 (3.7)	-	2 (4.2)	-	2 (10.0)	-
Chills	2 (3.7)	-	-	-	9 (45.0)	-
Cough	12 (22.2)	-	7 (14.6)	-	4 (20.0)	-
Diarrhea	22 (42.6)	1 (1.9)	9 (18.8)	2 (4.2)	1 (5.0)	-
Dyspnea	6 (11.1)	-	4 (8.3)	-	-	-
Fatigue or asthenia	43 (79.6)	-	-	-	16 (80.0)	-
Fatigue	42 (77.8)	-	36 (75.0)	-	16 (80.0)	-
Asthenia	1 (1.9)	1 (1.9)	1 (2.1)	-	1 (5.0)	-
Headache	16 (29.6)	-	8 (16.7)	-	10 (50.0)	-
Nausea, incl. vomiting	13 (24.1)	-	9 (18.8)	-	10 (50.0)	1 (5.0)
Pyrexia	15 (27.8)	-	4 (8.3)	-	13 (65.0)	-
Skin reaction	17 (31.5)	-	9 (18.8)	-	6 (30.0)	-
Pruritus	13 (24.1)	-	9 (18.8)	-	3 (15.0)	-
Rash	9 (16.7)	-	2 (4.2)	-	5 (25.0)	-
Immune-related adverse events						
Colitis	3 (5.6)	-	1 (2.1)	1 (2.1)	-	-
Diabetes Mellitus type 1	-	-	-	-	-	-
Hepatitis	-	2 (3.7)	-	-	-	-
Hyperthyroidism	8 (14.8)	-	6 (12.5)	-	-	-
Hypothyroidism	11 (20.4)	-	9 (18.8)	-	-	-
Pneumonitis	-	-	2 (4.2)	-	-	-

plus trametinib, and we found that adjuvant ICI was better tolerated than dabrafenib plus trametinib, which had a higher risk of treatment-limiting AEs, although reversible.

Even though included patients were treated in the adjuvant setting for melanoma, patient populations slightly differ since the studied treatments are applied for specific indications within adjuvant treatment (e.g., D+T for BRAF positive patients only, and nivolumab also for resected stage IV). Furthermore, dabrafenib plus trametinib was only available through an expanded access program until November 2020, potentially influencing treatment choice. Therefore, our results should be interpreted with caution when comparing treatments head-to-head.

Table 6.4. Treatment-limiting adverse events in patients who ended treatment due to toxicity. Patients could have one, or combinations of adverse events.

Adverse events, no.	Nivolumab n=9	Pembrolizumab n=3	Dabrafenib plus trametinib n=8
Allergic reaction	-	-	1
Arthralgia	1	1	1
Chills	-	-	3
Decreased appetite	-	-	1
Fatigue	-	-	2
Fever	-	-	5
Headache	-	-	1
Liver function disorders	-	-	1
Malaise	-	-	2
Myalgia	-	1	2
Nausea	-	-	3
Skin disorder	-	-	4
Syncope	-	-	1
Tachycardia	-	-	1
Immune-related adverse events	8	2	-
Adrenalitis	1	-	-
Colitis	2	1	-
Hepatitis	2	-	-
Hypocortisolism	1	-	-
Meningitis	1	-	-
Myocarditis, no.	1	-	-
Myositis, no.	1	-	-
Polymyalgia rheumatica, no.	1	-	-
Pneumonitis, no.	2	1	-
Thyroiditis, no.	2	-	-

4.1 Tolerability

As the investigated treatments are administered adjuvant after surgery, the accepted toxicity, and therefore tolerability, might be valued differently than during palliative treatment. In this small study, significantly more patients treated with dabrafenib plus trametinib ended their treatment due to treatment-limiting toxicity (53%), than after treatment with ICI (N: 20%, P: 8%). This is not completely unexpected, as the rate for treatment-limiting toxicity in COMBI-AD trial of dabrafenib plus trametinib was also higher (26%), than in the CheckMate 238-trial for nivolumab (8%), and the EORTC

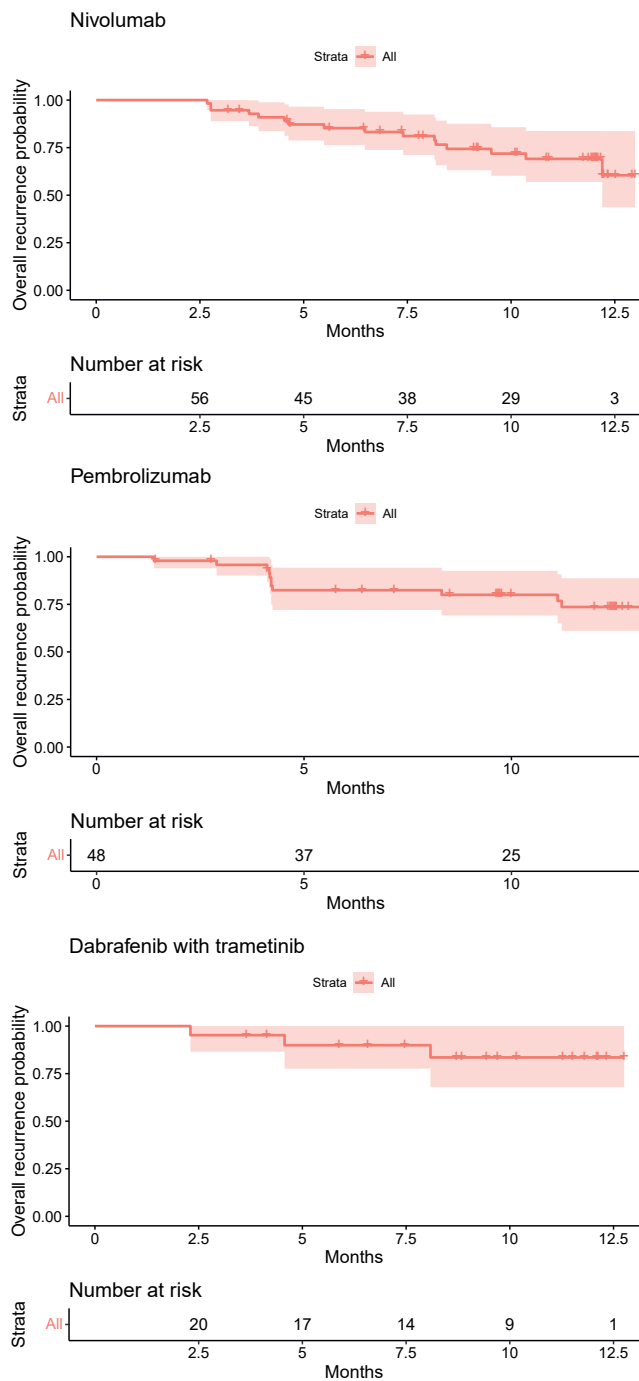


Figure 6.4. Kaplan–Meier curves representing the recurrence-free survival (RFS) of nivolumab, pembrolizumab, and dabrafenib plus trametinib 1-year RFS was 70.3% (95% confidence interval (CI): 58.1–85.0) for nivolumab, 72.4% (95% CI: 60–87.3) for pembrolizumab, and 83.0% (95% CI: 67.1–100) for dabrafenib plus trametinib.

1325/keynote-045 trial for pembrolizumab (14%) [9, 11, 12]. However, both for dabrafenib plus trametinib and nivolumab, the relative number of patients who ended treatment due to toxicity was higher than in the trials. This was also observed in the studies of de Meza et al. and Hoffmann et al., which showed higher treatment-limiting AEs on ICI in clinical practice [22, 23]. As far as we know, no other real-world data is yet published with treatment limiting toxicity rates for dabrafenib plus trametinib.

4.2 Safety

4.2.1 Immune checkpoint inhibitors

In this study, not all severe ($\geq$ grade 3) AEs were treatment-limiting, and vice versa. However, the incidence of one or two concurring irAEs was mostly treatment-limiting. IrAEs are characterized by auto-reactive T-cells, and can affect many organs, as was demonstrated in this study by the variety of treatment-limiting AEs [24-26]. Furthermore, we reported severe hepatitis (n=2), asthenia (n=1), and diarrhea (n=1) during treatment with nivolumab, and diarrhea (n=2), and colitis (n=1) during pembrolizumab. De Meza et al. did report higher rates of severe AEs in clinical practice compared to the RCTs [22]. Due to the low incidence of the individual severe AEs, we could not determine if this was the case for our study population.

Besides, manifestations of thyroiditis, hyperthyroidism (N: 15%, P: 13%) and hypothyroidism (N: 20%, P: 19%) seemed to occur more often in this study than in the trials (hyperthyroidism, N: 8%, P: 10%; hypothyroidism: N: 10%, 14%) [9, 11]. Even though a thyroid-related AE not necessarily leads to treatment termination, patients often need lifelong hormone replacement [27], and therefore the occurrence of this type of AE might have a severe impact on an individual patient's quality of life.

The chronic aspect and the recurrence of irAEs can play a role in the treatment decision. In our study, we did not determine the duration of AEs, but Patrinely et al. also showed that in clinical practice irAEs appeared more often and were frequently persistent, than was shown in clinical trials [28]. Furthermore, restarting therapy after a significant irAE was shown to result in another irAE in 50% of the patients in both a population with renal cell carcinoma and lung cancer, even though mostly with a lower severity rate [29, 30].

4.2.2 Dabrafenib plus trametinib

For dabrafenib plus trametinib treatment, the only reported severe AE was nausea, even though a total of 28 AEs contributed to treatment discontinuation in eight patients. In contrast to ICI, it were always combinations of AEs (two up to six) that were treatment-limiting. Pyrexia (5/8), was the most frequent reported, which is not surprising, as pyrexia specifically is known to occur during treatment with dabrafenib plus trametinib [31]. In three occurrences, it was reported with, at least, chills and nausea. In total, pyrexia occurred in total in 13/20 patients, comparable to the RCT [12]. And, even though in the RCT pyrexia was also identified as the AE most often (9%) resulting in treatment discontinuation [32], the rate of five out of twenty patients in our population seemed higher.

In this study, five of the eight patients on dabrafenib plus trametinib with treatment-limiting toxicity ended treatment within three months. Early treatment-limiting toxicity (≤ 3 months) is especially undesirable since these patients might not benefit at all from adjuvant treatment. Remarkably, four out of the five patients were female. However, as these patient numbers are small, we were not able to substantiate this with statistical analysis. Atkinson et al. showed that the incidence of AEs on dabrafenib plus trametinib is the highest during the first three months [32], which is in accordance with the pattern we observed. To encourage patients to remain on treatment, patient education and empowerment could be useful. Mansfield et al. showed that the maximum acceptable risk for pyrexia in melanoma patients receiving dabrafenib plus trametinib was higher when the awareness of the benefits of the therapy is higher. Therefore longer follow-up results on efficacy of the COMBI-AD trial might be beneficial [33]. Furthermore, the COMBI-APlus trial showed that treatment interruption of both dabrafenib and trametinib in case of pyrexia at a temperature of $\geq 38^\circ\text{C}$, instead of discontinuation of only dabrafenib in case of pyrexia of $\geq 38.5^\circ\text{C}$, helped patients in the long term to remain on treatment [32].

4.3 Efficacy

The preliminary 1-year recurrence-free survival rates overlapped with the results of the RCTs [9, 11, 12]. Even though this is indicative for comparable efficacy, and RFS seems to be a valid surrogate endpoint for overall survival for adjuvant melanoma therapy, it is too soon for conclusions on real-world effectiveness [34]. De Meza et al. showed comparable 1-year survival rates of 87.0% for the stage IIIA, and 76.5% for

stage IIIB, and 60.3% for stage IIIC in a population treated with ICI in Dutch clinical practice [22]. Hoffmann et al. showed a RFS at 1-year of 77.1% after nivolumab and 63.5% after pembrolizumab in a Swiss population [23]. Furthermore, even though Koelblinger et al. showed a 1-year RFS rate of 64.8% in an Austrian population, they showed comparable distant metastasis free-survival compared to the Checkmate-238 trial (77.4% vs. 80%), and concluded treatment cost-effectiveness for the Austrian population [35]. All real-world studies showed comparable effectiveness to the RCTs, which is in line with our premature results. Hoffmann et al. is to our knowledge the only published study reviewing dabrafenib plus trametinib in clinical practice, however they only included three patients who finished their treatment [23].

4.4 Eligibility criteria

A Dutch nationwide registry showed that 40% of the patients who received treatment for advanced melanoma, were ineligible for phase III trials [36]. However, our study population with patients who received adjuvant treatment, in general, met the criteria from the trials. Key exclusion criteria in the phase III trials for adjuvant treatment were ECOG performance status ≥ 1 , ocular (N, D+T), uveal (N) or mucosal (D+T) melanoma, previous systemic therapy for melanoma (all), auto-immune disease (N, P), uncontrolled infection (P) or use of systemic glucocorticoids (N, P) [9, 11, 12]. Even though the performance status is unknown for 35% of our patients – comparable to other real-world studies [37, 38] – there was no reason to believe this would be significantly different from phase III trials. Only seven (5.7%) patients that had previous systemic therapy for melanoma did not meet the eligibility criteria, including neo-adjuvant treatment in the PRADO trial and patients who switched adjuvant treatments after early AEs [39].

4.5 Need for real-world data

None of the treatments in this study, and in general, is clearly superior to the others regarding the tolerability, safety and efficacy combined, and proving superiority was not the aim of this study. However, for patients with a BRAF-mutation, a choice between adjuvant treatment with ICI or dabrafenib plus trametinib must be made. Overall, this study reflects the safety profiles shown in the RCTs, however incidence of thyroid-related irAEs during ICI, often chronic, and treatment-limiting AEs during both nivolumab and dabrafenib plus trametinib seemed higher. Furthermore, effectiveness, estimated by the 1-year RFS, was comparable with the trials. Weilandt et al. showed

that in a discrete choice experiment that overall response rate is the most important parameter for patients when choosing a treatment, followed by the 2-year survival rate, type of adverse events and the probability of treatment-limiting AEs. However, individual (e.g., age, sex, partnership) and disease-related (e.g., tumor burden, experience with ICI or BRAF-MEK inhibitors) resulted in variation of the preferences [40]. Livingstone et al. showed that these values overlap with factors Australian physicians and nurses consider in the recommendation of ICI to patients. [41].

To be able to inform melanoma patients on the risks and benefits related to their personal characteristics, more data is needed, which underlines the need for real-world data [42]. This study shows that these data can be collected with text mining from EHR. Even though this study represents only the population of a single university hospital, one of the fourteen melanoma centers in The Netherlands. Currently, data on this population are also collected from EHR by data managers for the Dutch Melanoma Treatment Registry since 2011. However, this is done manually by data managers, is updated once every year and does not include < grade 3 adverse events [22]. This study showed that data extraction using text mining is feasible, and in the future, can potentially be extended to other hospitals and registries, for efficient inclusion of larger populations.

4.6 Strengths and limitations

We used rule-based text mining to detect patients, and collect characteristics, outcomes, and adverse events from electronic health records. The advantage of this method is the faster and more standardized data extraction. Furthermore, the set of queries can repeatedly be reused to review the status of adjuvant melanoma treatments in this hospital in the future, for example on yearly basis, and has the potential to be implemented in other hospitals treating melanoma patients.

However, due to its retrospective design, the extracted data is both limited by the information stored in the EHR, which is prone to have missing data [43], and by the terms included in the queries [44]. In this study we aimed for a high sensitivity, by using extensive lists of keywords for the searches in unstructured text, and additionally performed visual context analysis to confirm outcomes, resulting in limited missing data. However, certain information can be underreported.

5. Conclusions

This was the first real-world study on adjuvant treatment in resected stage III and IV melanoma with retrieval of clinical data by text mining. In this study we showed that text mining EHR data valuable method to evaluate the tolerability, safety, and efficacy of adjuvant melanoma treatments. In this population we found a higher tolerability of ICI, as compared to dabrafenib plus trametinib, primarily due to pyrexia and fatigue. Regarding safety this study shows results comparable to the clinical trial data, except for a higher incidence of treatment-limiting AEs for both nivolumab and dabrafenib plus trametinib, thyroid-related irAEs for nivolumab and pembrolizumab, and some mild adverse events. Furthermore, all 1-year RFS rates were comparable. Implementation of text mining of EHR in multiple hospitals can further improve efficiency of capturing real-world data.

References

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA: A Cancer Journal for Clinicians, 2021. **71**(3): p. 209-249.
2. Integraal Kankercentrum Nederland. *NKR-cijfers. Incidentie, Melanoom van de huid en externe geslachtsorganen*, CR. Available from: https://iknl.nl/nkr-cijfers?fs%7Cepidemiologie_id=526&fs%7Ctumor_id=243&fs%7Cregio_id=550&fs%7Cperiode_id=564%2C565%2C566%2C567%2C568%2C569%2C570%2C571%2C572%2C573%2C574%2C575%2C576%2C577%2C578%2C579%2C580%2C581%2C582%2C583%2C584%2C585%2C586%2C587%2C588%2C589%2C590%2C591%2C592%2C593%2C563%2C562%2C561&fs%7Cgeslacht_id=644&fs%7Cleeftijdsgroep_id=677&fs%7Cjaren_na_diagnose_id=687&fs%7Ceenheid_id=702&cs%7Ctype=line&cs%7CcxAxis=periode_id&cs%7Cseries=epidemiologie_id&ts%7CcrowDimensions=&ts%7CcolumnDimensions=periode_id&lang%7Clanguage=nl.
3. Curti, B.D. and M.B. Faries, *Recent Advances in the Treatment of Melanoma*. New England Journal of Medicine, 2021. **384**(23): p. 2229-2240.
4. European Medicines Agency. *CHMP post-authorisation summary of positive opinion for Keytruda*. 2018; Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-keytruda-ii-47_en.pdf.
5. European Medicines Agency. *CHMP post-authorisation summary of positive opinion for Opdivo*. 2018; Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-opdivo_en-1.pdf.
6. European Medicines Agency. *CHMP post-authorisation summary of positive opinion for Mekinist*. 2018; Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-mekinist-ws/1274_en.pdf.
7. European Medicines Agency. *CHMP post-authorisation summary of positive opinion for Tafinlar*. 2018; Available from: https://www.ema.europa.eu/en/documents/smop/chmp-post-authorisation-summary-positive-opinion-tafinlar-ws-1274_en.pdf.
8. Ascierto, P.A., et al., *Adjuvant nivolumab versus ipilimumab in resected stage IIIB-C and stage IV melanoma (CheckMate 238): 4-year results from a multicentre, double-blind, randomised, controlled, phase 3 trial*. Lancet Oncol, 2020. **21**(11): p. 1465-1477.
9. Weber, J., et al., *Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma*. New England Journal of Medicine, 2017. **377**(19): p. 1824-1835.
10. Eggermont, A.M.M., et al., *Adjuvant pembrolizumab versus placebo in resected stage III melanoma (EORTC 1325-MG/KEYNOTE-054): distant metastasis-free survival results from a double-blind, randomised, controlled, phase 3 trial*. Lancet Oncol, 2021. **22**(5): p. 643-654.
11. Eggermont, A.M.M., et al., *Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma*. New England Journal of Medicine, 2018. **378**(19): p. 1789-1801.
12. Long, G.V., et al., *Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma*. New England Journal of Medicine, 2017. **377**(19): p. 1813-1823.
13. Dummer, R., et al., *Five-Year Analysis of Adjuvant Dabrafenib plus Trametinib in Stage III Melanoma*. New England Journal of Medicine, 2020. **383**(12): p. 1139-1148.
14. Wright, J.J., A.C. Powers, and D.B. Johnson, *Endocrine toxicities of immune checkpoint inhibitors*. Nature Reviews Endocrinology, 2021. **17**(7): p. 389-399.
15. Dimitriou, F., G.V. Long, and A.M. Menzies, *Novel adjuvant options for cutaneous melanoma*. Annals of Oncology, 2021. **32**(7): p. 854-865.

16. Wang, S.V., et al., *Using Real-World Data to Extrapolate Evidence From Randomized Controlled Trials*. Clinical Pharmacology & Therapeutics, 2019. **105**(5): p. 1156-1163.
17. 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.
18. van Laar, S.A., et al., *An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma*. Clinical Pharmacology & Therapeutics, 2020. **108**(3): p. 644-652.
19. van Laar, S.A., et al., *Real-world evaluation of supportive care using an electronic health record text-mining tool: G-CSF use in breast cancer patients*. Supportive Care in Cancer, 2022.
20. van Laar, S.A., et al., *Real-World Metastatic Renal Cell Carcinoma Treatment Patterns and Clinical Outcomes in The Netherlands*. Frontiers in Pharmacology, 2022. **13**.
21. van der Stoep, M.Y.E.C., et al., *Treosulfan-induced myalgia in pediatric hematopoietic stem cell transplantation identified by an electronic health record text mining tool*. Scientific Reports, 2021. **11**(1): p. 19084.
22. 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.
23. Hoffmann, M., S. Hayoz, and B.C. Özdemir, *Prescription Patterns, Recurrence, and Toxicity Rates of Adjuvant Treatment for Stage III/IV Melanoma-A Real World Single-Center Analysis*. Biology (Basel), 2022. **11**(3).
24. 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.
25. 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.
26. 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.
27. Darnell, E.P., et al., *Immune-Related Adverse Events (irAEs): Diagnosis, Management, and Clinical Pearls*. Current Oncology Reports, 2020. **22**(4): p. 39.
28. Patrinely, J.R., Jr, et al., *Chronic Immune-Related Adverse Events Following Adjuvant Anti-PD-1 Therapy for High-risk Resected Melanoma*. JAMA Oncology, 2021. **7**(5): p. 744-748.
29. Santini, F.C., et al., *Safety of retreatment with immunotherapy after immune-related toxicity in patients with lung cancers treated with anti-PD(L)-1 therapy*. Journal of Clinical Oncology, 2017. **35**(15_suppl): p. 9012-9012.
30. Abou Alaiwi, S., et al., *Safety and efficacy of restarting immune checkpoint inhibitors after clinically significant immune-related adverse events in metastatic renal cell carcinoma*. Journal for immunotherapy of cancer, 2020. **8**(1): p. e000144.
31. Heinzerling, L., et al., *Tolerability of BRAF/MEK inhibitor combinations: adverse event evaluation and management*. ESMO Open, 2019. **4**(3): p. e000491.
32. 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.
33. Mansfield, C., et al., *Risk tolerance in adjuvant and metastatic melanoma settings: a patient perspective study using the threshold technique*. Future Oncology, 2021. **17**(17): p. 2151-2167.
34. Suci, S., et al., *Relapse-Free Survival as a Surrogate for Overall Survival in the Evaluation of Stage II-III Melanoma Adjuvant Therapy*. J Natl Cancer Inst, 2018. **110**(1).

35. Koelblinger, P., et al., *Adjuvant anti-PD-1 antibody treatment in stage III/IV melanoma: real-world experience and health economic considerations*. JDDG: Journal der Deutschen Dermatologischen Gesellschaft, 2021. **19**(8): p. 1186-1198.
36. van Zeijl, M.C.T., et al., *Real-world outcomes of advanced melanoma patients not represented in phase III trials*. Int J Cancer, 2020. **147**(12): p. 3461-3470.
37. 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.
38. 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.
39. Reijers, I.L.M., et al., *Personalized response-directed surgery and adjuvant therapy after neoadjuvant ipilimumab and nivolumab in high-risk stage III melanoma: the PRADO trial*. Nat Med, 2022.
40. Weilandt, J., et al., *Patient Preferences in Adjuvant and Palliative Treatment of Advanced Melanoma: A Discrete Choice Experiment*. Acta Derm Venereol, 2020. **100**(6): p. adv00083.
41. Livingstone, A., et al., *Adjuvant immunotherapy recommendations for stage III melanoma: physician and nurse interviews*. BMC Cancer, 2021. **21**(1): p. 1014.
42. Bica, I., et al., *From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges*. Clinical Pharmacology & Therapeutics, 2021. **109**(1): p. 87-100.
43. Haneuse, S., et al., *Learning About Missing Data Mechanisms in Electronic Health Records-based Research: A Survey-based Approach*. Epidemiology, 2016. **27**(1).
44. Percha, B., *Modern Clinical Text Mining: A Guide and Review*. Annual Review of Biomedical Data Science, 2021. **4**(1): p. 165-187.

Supplementary data

Supplementary File S6.1A. Patient inclusion

A patient is included if it meets criteria 1, 2 and at least one of 3.						
Inclusion criterion	Description	Type of data	Searched terms	Used for inclusion or exclusion	Time window	Comments
1.	DBC-code*	DBC	Specialism: 0313 (internal medicine) with diagnosis: 842 (melanoma/malignancy skin)	Inclusion		Diagnose behandel combinatie = diagnosis treatment combination, a code used for reimbursements in the Netherlands
2.	Mention of adjuvant treatment or stage III	Verification adjuvant treatment or stage III	Text search	Inclusion		

Supplementary Table S6.1A continues on next page.

Supplementary File S6.1A. *Continued*

Inclusion criterion	Description	Type of data	Searched terms	Used for inclusion or exclusion	Time window	Comments
3a. Nivolumab use	Nivolumab prescription	Medication request	Nivolumab	Inclusion	> 1 January 2019	
	Verification nivolumab use	Text search	Nivolumab OR nivo OR opdivo	Inclusion		
	Not together with ipilimumab (=palliative indication)	Medication request	Ipilimumab	Exclusion	Start is between start date nivolumab AND start date nivolumab + 3 weeks	
3b. Pembrolizumab use	Pembrolizumab prescription	Medication request	Pembrolizumab	Inclusion	> 1 January 2019	
	Verification pembrolizumab use	Text search	Pembrolizumab OR pembro OR keytruda	Inclusion		
3c. Dabrafenib plus trametinib use	Dabrafenib	Medication request	Dabrafenib	Inclusion	> 1 January 2019	
	Trametinib	Medication request	Trametinib	Inclusion	Start is between 1 week before start dabrafenib AND 1 week after dabrafenib	
	Verification dabrafenib plus trametinib use	Text search	Dabrafenib OR dab OR tafinlar OR trametinib OR tram OR mekinist	Inclusion		

Supplementary File S6.1B. Data collection

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Start date	Medication request		Start date	Oldest prescription date of dabrafenib OR trametinib			
Stop	Medication request		Start date of the oldest prescription	Newest prescription date of dabrafenib OR trametinib			
Stop date	Text search		Date of selected text source	Dabrafenib OR dab OR dabrafenib OR tafinlar OR trametinib OR tram OR trametinib OR mekinist OR TKI OR braf met remmer OR braf met remmers OR braf mek inhibitor OR braf mek inhibitors OR behandelings staken OR stoppen OR staken OR discontinueren OR niet continueren OR gestaakt OR gestopt OR staak OR stop OR adjuvante behandeling OR adjuvant OR adjuvante neo-adjuvante	Consult, consulten, reports	Answer section "0-1 month after stop": between start date dab tram AND stop date dab tram+ 1 month Answer section "1-2 month after stop": between stop date dab tram+ 1 month AND stop date dab tram + 2 months Answer section "2-3 month after stop": between stop date dab tram+ 2 month AND stop date dab tram + 3 months Answer section "3-6 month after stop": between stop date dab tram+ 3 month AND stop date dab tram + 6 months Answer section "6-9 month after stop": between stop date dab tram+ 6 month AND stop date dab tram + 9 months Answer section "9-12 month after stop": between stop date dab tram+ 9 month AND stop date dab tram + 12 months	

For treatment with dabrafenib and trametinib

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Start date	Medication request		Start date	Oldest prescription date of nivolumab			
Stop	Medication request		Start date of the oldest prescription	Newest prescription date of nivolumab			
Stop date	Text search		Date of selected text source	Nivolumab OR nivo OR opdivo OR immunotherapy OR immunotherapie OR behandelings staken OR stoppen OR staken OR niet continueren OR gestaakt OR gestopt OR staak OR stop OR adjuvante behandeling OR adjuvant OR adjuvante OR neo-adjuvante	Consult, consulten, reports	Answer section "0-1 month after stop": between start date nivolumab AND stop date nivolumab+ 1 month Answer section "1-2 month after stop": between stop date nivolumab+ 1 month AND stop date nivolumab + 2 months Answer section "2-3 month after stop": between stop date nivolumab+ 2 month AND stop date nivolumab + 3 months Answer section "3-6 month after stop": between stop date nivolumab+ 3 month AND stop date nivolumab + 6 months Answer section "6-9 month after stop": between stop date nivolumab+ 6 month AND stop date nivolumab + 9 months Answer section "9-12 month after stop": between stop date nivolumab+ 9 month AND stop date nivolumab + 12 months	

For treatment with nivolumab

Start date	Medication request	Start date	Oldest prescription date of pembrolizumab	
Stop	Medication request	Start date of the oldest prescription	Newest prescription date of pembrolizumab	
Stop date	Text search	Date of selected text source	Pembrolizumab OR pembro OR keytruda OR pem OR immunotherapy OR immuno therapie OR behandelng staken OR stoppen OR staken OR niet discontinueren OR gestopt OR staak OR stop OR adjuvante behandeling OR adjuvant OR adjuvante OR neo-adjuvante	Consult, consulten, reports Answer section "0-1 month after stop": between start date pembrolizumab AND stop date pembrolizumab+ 1 month Answer section "1-2 month after stop": between stop date pembrolizumab+ 1 month AND stop date pembrolizumab + 2 months Answer section "2-3 month after stop": between stop date pembrolizumab+ 2 month AND stop date pembrolizumab + 3 months Answer section "3-6 month after stop": between stop date pembrolizumab+ 3 month AND stop date pembrolizumab + 6 months Answer section "6-9 month after stop": between stop date pembrolizumab+ 6 month AND stop date pembrolizumab + 9 months Answer section "9-12 month after stop": between stop date pembrolizumab+ 9 month AND stop date pembrolizumab + 12 months

For treatment with pembrolizumab

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Treatment outcomes							
Reason stop	Text search	Multiple choice options* - Toxicity - Recurrence - End or treatment - Did not end treatment - Other - Unclear	Multiple choice option + All treatment-limiting adverse events date of recurrence	Adverse event OR bijwerking OR bijwerkingen OR tox OR toxiciteit OR graad 3 OR graad 4 OR ernstig OR klachten OR last Terug OR recurrence OR metastasis OR metastasize OR gemetastaseerd OR uitgezaaid OR gemetastaseerd OR uitgezaaide OR gemetast OR uitzaaiingen OR gemet OR meta's OR gemetastaseerd OR tumormetastase OR metastase OR metastasen OR metastasen lever OR tumor metastase	Consult, consulten, reports consult, consulten ALL radiology reports	Between start date treatment AND stop date treatment+ 1 month	*Reason of stop is based on the query stop date. In case of toxicity, all dose-limiting adverse events are extracted from text. In case of recurrence, the date of recurrence is selected by identifying the moment in consultation reports or radiology report.

Recurrence after tox	Text search	Multiple choice op- tions* - recur- rence - no re- currence within 12 months	Multiple choice option + date of selected text source	Terug OR recurrence OR metastasis OR metastasize OR gemetastaseerd OR uitgezaaid OR gemetaseerd OR metastases OR uitgezaaide OR gemetast OR uitzaaligen OR gemet OR meta's OR gemetastaseerd OR tumormetastase OR metastase OR metastasen OR metastasen lever OR tumor metastase	Consult, consulten ALL radiology reports	Between Start date treatment AND start date treatment + 13 months	*Both multiple choice answers have the same search criteria, chosen answer is based on content
-------------------------	-------------	--	--	--	--	--	--

Supplementary Table S6. 1B continues on next page.

Supplementary File S6.1B. Continued

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Patient and disease characteristics							
Performance score	Measurement		Most recent value between 0 and 5 OR Most recent value between 0 and n	WHO performance OR WHO graad OR WHO klasse OR Definitieve WHO classificatie OR performance WHO OR inspanningstolerantie WHO-klasse OR WHO performance score OR WHO performance status OR WHO-klasse OR WHO classificatie OR performance status WHO OR WHO-score OR WHO PS OR PS WHO Karnofsky OR Karnofsky score OR Karnofsky performance scale OR Karnofsky index OR Karnofsky performance OR Karnofsky score OR Karnofsky performance score OR who karnofsky OR Karnofsky performance status OR Karnofsky score OR performance karnofsky OR Karnofsky score OR Karnofsky score OR Karnofsky landsky score OR kps OR kps score OR kps geschat		Before start date treatment	

Perfor- mance score text	Text search	Multiple choice options -WHO 0 (incl karnofsky 100) - WHO 1 (incl karnofsky 90-80) - WHO 2 (incl karnofsky 70-60) - WHO 3 (incl carnofsky 50-0)	n = WHO/ECOG performance status or karnofsky performance status according to multiple choice label WHO performance n OR WHO n OR WHO klasse n OR WHO score n OR ecog performance status n OR ecog score n OR ecog n OR performance score n OR karnofsky index n OR karnofsky n OR karnosky score n OR karnoskyindex n OR karnofsky-index n OR kanofsky performance status n OR karnosky scale n OR karnofsky performance status scale n OR KPS n	
LDH	Measurement	Most recent value	LDH	Before start date treatment
Previous systemic treatments	Medication request	Medication name	All treatments starting with ATC code L01, excluding topical applications	Before start date treatment

Supplementary Table S6.1B continues on next page.

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Stage	Text search	<u>Multiple choice options:</u> - III - IIIa - IIIb - IIIC - IIIC resectable - IIIC irresectable (excluded) - IV resected - unknown	Multiple choice option	Stadium III OR stadium 3 OR stage III or stage 3 Stadium IIIa OR tadium 3a OR stadium III a OR stadium 3 a OR stage IIIa OR stage 3a OR stage III a OR stage 3 a Stadium IIIb OR tadium 3b OR stadium III b OR stadium 3 b OR stage IIIb OR stage 3b OR stage III b OR stage 3 b Stadium IIIC OR tadium 3c OR stadium III c OR stadium 3 c OR stage IIIC OR stage 3c OR stage III c OR stage 3 c (Stadium IIIC OR tadium 3c OR stadium III c OR stadium 3 c OR stage IIIC OR stage 3c OR stage III c OR stage 3 c) AND (resectable OR resectabel) Stadium IIIC OR tadium 3c OR stadium III c OR stadium 3 c OR stage IIIC OR stage 3c OR stage III c OR stage 3 c AND (niet resectabel OR niet resectable OR irresectable) stadium IV OR stadium 3 OR stage IV OR stage 4 stadium OR adjuvant OR adjuvante		Before start date treatment	

Primary tumor location	Text search	<u>Multiple choice options:</u>	Multiple choice option	(Eye OR eyeball) OR optic OR ocular O oog OR ogen OR oculair OR oculaire) AND (melanoom OR melanoma OR melanomen) AND (primair OR primaire) Acraal OR acral lentiginous melanoma OR acraal melanoom OR acral lentiginou malinant melanoma OR ALM OR acrolentiginous melanoom (mucous membrane OR mucosa OR mucosal OR slijmvliesweefsel OR slijmvlies OR slijm vlies OR mucosaal) AND (melanoom OR melanoma OR melanomen) AND (primair OR primaire) (Leg or lower leg OR been OR benen OR onderbeen OR bovenbeen OR kuit OR dij OR arm OR armen OR bovenarm OR onderarm OR extremititeit OR ledemaat OR ledematen OR extremiteiten OR bovenste extremititeit OR onderste extremititeit OR scheen OR scheenbeen OR hand OR vinger OR teen) AND (melanoom OR	Before start date treatment
------------------------	-------------	---------------------------------	------------------------	--	-----------------------------

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
				Melanoma OR melanomen) AND (primair OR primaire) (Romp OR bulk OR rug OR onderrug OR flank OR flanken OR bovenbulk OR boventrug OR onderbulk) AND (melanoom OR melanoma OR melanomen) AND (primair OR primaire) (primair OR primaire) AND (tumor OR melanoom OR melanomen OR tumorlocatie) AND (onbekend OR niet bekend OR onduidelijk OR onduidelijk OR zonder OR occult OR occulte) (Melanoom OR melanomen OR melanom)			

Histology	Text search	Multiple choice options: - superficial spreading - Nodular - Acral Lentiginous - Spindle cell - Unclear	Multiple choice option	Superficieel spreidend nodulair ALM OR acrolentigineus melanoom OR acrolentigineus Spindle cell carcinoma OR sacomatoid carcinoma OR OR spoelcelcarcinoom OR spoelceica OR spoelcel OR spoelcelmelanoom Melanoom OR melanoma OR melanomen	Before start date treatment
BRAF mutation	Text search	Multiple choice options: - Positive - Negative - Unclear	Multiple choice option	BRAF positief OR braf v600e-mutatie positief OR braf v600e mutation positiveOR braf v600e positief OR braf analyse positief OR braf gemuteerd OR braf POS BRAF negatief OR braf v600e-mutatie negatief OR braf v600e negatief OR braf analyse negatief OR braf neg OR geen braf mutatie BRAF	Before start date treatment
NRAS mutation	Text search	Multiple choice options: - Positive - Negative - Unclear	Multiple choice option	NRAS positief OR NRAS analyse positief OR NRAS gemuteerd OR klasse 5 pathogene variatie in NRAS OR NRAS pos NRAS negatief OR geen NRAS OR NRAS neg NRAS	Before start date treatment

Supplementary Table S6. 1B continues on next page.

Supplementary File S6.1B. Continued

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
KIT mutation	Text search	Multiple choice options: - Positive - Negative - Unclear	Multiple choice option	KIT*		Before start date treatment	*Since KIT mutations are rare, we started with sorting all patients with a "KIT" mutation as mention as "negative", and verified/corrected if this was (not) the case.
Ulceration	Text search	Multiple choice options: - Yes - No - Unclear	Multiple choice option	- Ulceratie - Geen ulceratie		Before start date treatment	

Adverse events			
Pyrexia	Text search Vital signs	Multiple choice options: - Any grade - Grade 3-4 - no	Multiple choice option
			Fever OR febrile OR high body temperature OR pyrexia OR hyperthermia OR temperature elevation OR feverish OR high temperature OR pyrexia OR hyperthermie OR lichaamstemperatuur verhoging OR hoge temperatuur OR temperatuursverhoging OR koortsig OR koortsachtig OR koorts OR temperatuur hoog OR verhoging lichaamstemperatuur OR hoog temperatuur OR koorst OR lichaamstemperatuur verhoogd OR temperatuur stijging OR koortsige OR koortsen OR verhoogde lichaamstemperatuur OR febriel OR fever with chills OR koortsrillingen OR koorts met koude rillingen OR koorts met koude rilling temperature > 38 degrees
			Between start date treatment AND stop date treatment + 1 month
			Geen koorts

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Artralgia	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Artralgia OR joint pain OR aching joints OR articular pain OR painful joint OR pijnlijk gewricht OR pijn in gewricht OR artralgie OR pijn gewricht OR pijnlijke gewrichten OR gewrichtspijn OR gewricht pijnlijk OR pijnlijk gewrichten OR pijnlijke gewricht OR gewrichten pijnlijk OR gewrichten doen pijn OR aralgie OR arhalgie		Between start date treatment AND stop date treatment + 1 month	

Colitis	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Colitis OR ontsteking van het colong OR colitis NOT "besproken zoals colitis, huisafwijkingen, hepatitis" OR "bijwerkingen, met name diarree/colitis, schildklier OR het immuunsysteem, te weten colitis, hepatitis." *	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded, as these were often used in EHR files indicating that a patient received the AE information prior to treatment.
Diarrhea	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Diarrhea OR diarrhee OR diarree OR diarree OR defecatie OR diarree? (onverklaard) diarree? OR geen diarree OR defecatie: vast patroon * Geen diarree OR defecatie: vast patroon	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded from this multiple choice option as these were indicative for not having diarrhea

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. Continued

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Hypothyroidism	Text search Measurement Text search Medication request	Multiple choice options: - Any grade - Grade 3-4 - No - Unclear	Multiple choice option	Hypothyroidism OR hypothyreosis OR hypothyroid OR hypothyroidie OR deficientie schildklier OR hypothyroidie OR hypothyreose OR schildklier deficientie OR insufficientie schildklier OR schildklier insufficientie OR hypothyroidie Free T4 <9		Between start date treatment AND stop date treatment + 1 month	
				Levothyroxine OR thyroxine OR L-thyroxine OR schildklier levothyroxine			
Hyperthyroidism	Text search Measurement Text search Medication request	Multiple choice options: - Any grade - Grade 3-4 - No - Unclear	Multiple choice option	Hyperthyroidism OR hyperthyroid OR hyperthyreose OR hyperthyroidie OR hyperthyrodie OR hyperthyroidie OR basedow Free T4>24		Between start date treatment AND stop date treatment + 1 month	
				Levothyroxine OR thyroxine OR L-thyroxine OR schildklier levothyroxine			

Pneumonitis	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiople choice option	Pneumonitis OR pulmonary inflammation OR long ontsteking OR ontsteking long OR longontstekingen OR longonsteking OR pneumonia OR pulmonitis OR longontsteking OR inflammatie van de long OR lage luchtweginfectie OR longinfect OR pneumonie	Between start date treatment AND stop date treatment + 1 month
Diabetes Mellitus I	Text search Medication request	Multiple choice options: - Any grade - Grade 3-4 - No	Multiople choice option	Diabetes mellitus OR diabetes OR DM OR NIDDM OR IDDM OR diabets OR diabeet OR suikerziekte OR dmt 1 OR dm 1 OR mellitus diabetes NOT "let op: DM-patienten hebben een" OR "bv. Lichte DM, behandelde HT" OR DM schema: NVT* Insulins OR insulin-analogs	Between start date treatment AND stop date treatment + 1 month *EHR files with these keywords were excluded, as they are often used sentences including "DM"but not indicating DM

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. Continued

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Fatigue	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Fatigue OR loss of energy OR lack of energy OR energie afgenomen OR futloosheid OR lusteloosheid OR verlies van energie OR gevoel van energieloosheid OR lusteloos OR moeheid OR vermoeid OR uitzonderlijk moe OR futloos OR vermoeidheid OR vermoeide OR vermoeiden OR vermoeit OR vermoei OR afgenomen energie OR vermoeien OR moe OR malaise OR suf OR duf OR geen energie OR hagnerig OR zwak OR slap OR uitgeput OR verzwakt OR loom OR moeie OR extreme vermoeidheid?: ja NOT vermoeidheid?: nee*		Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded, indicating no fatigue
				Vermoeidheid?: nee			

Nausea	Text search Medication request	Multiple choice options: - Any grade - Grade 3-4 - No	Multiople choice option	Nausea OR nauseous OR nauseated OR misselijkheid OR onpasselijk OR onpasselijk gevoel OR misselijk OR misselijke OR vomiting OR emesis OR gebrakt OR regurgitatie maaginhoud OR braken OR braakt OR kots OR kotsen OR overgeven OR braaksel OR braakte OR vomeren OR spugen OR oprisping NOT "u misselijk bent of moet overgeven" OR "- misselijkheid / braken - tandbeschadiging" Aprepitant OR fosaprepitant OR netupitant OR granisetron OR ondansetron OR palonosetron	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded, it only indicated that a patient received AE information
--------	--------------------------------------	--	-------------------------------	---	---	--

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Headache	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Headache OR cephalgia OR cephalalgia OR head pain OR hoofd pijn OR cefalgie OR pijn hoofd OR hoofdpijn OR hoofdpijnen OR cefalalgie OR migraine OR headache OR migraine OR headache OR migrainous headache OR migraine type headaches OR migraine headache OR migrainehoofdpijn OR migraineuze hoofdpijn OR syndroom migraine OR hoofdpijn migraine OR migrainehoofdpijnen OR migraine hoofdpijn OR migraine syndroom OR migraineuze hoofdpijnen OR migraines NOT "geen hoofdpijn Geen hoofdpijn		Between start date treatment AND stop date treatment + 1 month	
Chills	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Shivering OR rillingen OR rillen OR rilling OR bibber OR bibberen OR beven OR rilde OR bibberde OR beefde OR chills OR chill OR koude rillingen OR koude rilling		Between start date treatment AND stop date treatment + 1 month	

Hepatitis	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Hepatitis OR ontsteking van de lever OR ontstekingsproces van lever OR niet-specifiek hepatitis OR ontsteking lever OR lever ontsteking NOT "een colitis, hepatitis, schildklierafwijkingen, vitiligo, etc."*	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded since it only indicated that a patient received AE information
Dyspnea	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Dyspneu OR shortness of breath OR breathlessness OR breathless OR difficulty breathing OR DIB OR SOB OR kortademigheid OR kortademigheid adem OR dyspnoea OR dyspneu OR ademnood OR moeilijk ademen OR adem kortademigheid NOT "geen dyspneu" OR "heeft u last van kortademigheid in rust of bij lichte inspanning?: nee"	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded, indicating no dyspnea

Supplementary Table S6.1B continues on next page.

Supplementary File S6.1B. *Continued*

Name data point	Type of data	Multiple choice labels	Collected answer	Searched terms/keywords	Keyword search restricted to	Time window	Comments
Cough	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Coughing OR cough OR hoest OR hoesten OR gehoest OR hoeste OR hoeste OR kuchen OR kuch NOT geen hoest OR "heeft u last van hoesten of keelpijn?: nee" OR geen hoesten OR geen dyspneu of hoesten*		Between start date treatment AND stop date treatment + 1 month	*EHR files with this sentence were excluded, indicating no cough
Pruritus	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Pruritus OR pruritis OR itching OR itchy OR jeukende huid OR jeuk OR jeukend OR jeuken Geen jeuk OR geen pruritus OR geen pruritis		Between start date treatment AND stop date treatment + 1 month	
Asthenia	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Asthenia OR zwakte OR asthenie OR krachteloosheid OR zwakheid OR zwak NOT geen zwakte OR geen asthenie Gene asthenie OR geen zwakte		Between start date treatment AND stop date treatment + 1 month	

Rash	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Exanthema OR skin eruption OR cutaneous eruption OR eruption OR skin rash OR exanthem OR rash OR huiduitslag OR vluchtig exantheem en overige niet-specifieke erupties OR effluorescentie OR huidruptie OR exantheem OR uitslag huid OR huid uitslag NOT geen uitslag OR geen rash*	Between start date treatment AND stop date treatment + 1 month	*EHR files with this sentence were excluded, indicating no rash
Abdominal pain	Text search	Multiple choice options: - Any grade - Grade 3-4 - No	Multiple choice option	Abdominal pain OR darmpijn OR abdominale pijn OR buikpijn OR pijn abdominaal OR abdominaal pijn OR pijn buik OR buik pijn OR pijn in buik OR pijn in onderbuik NOT geen buikpijn*	Between start date treatment AND stop date treatment + 1 month	*EHR files with these keywords were excluded, indicating no abdominal pain

Supplementary File S6.2a.

Treatment-limiting adverse event per individual patient	Nivolumab n=9									Total
	1	2	3	4	5	6	7	8	9	
Colitis	1	1								2
Pneumonitis		1	1							2
Hepatitis				1	1					2
Thyroiditis				1	1					2
Myalgia				1						1
Myocarditis				1						1
Arthralgia						1				1
Hypocortisolism							1			1
Polymyalgia rheumatica								1		1
Meningitis									1	1
Adrenalitis									1	1
Total	1	2	1	4	2	1	1	1	2	

Supplementary File S6.2b.

Treatment-limiting adverse event per individual patient	Pembrolizumab n=3			Total
Colitis	1			1
Arthralgia		1		1
Myalgia		1		1
Pneumonitis			1	1
Total	1	2	1	

Supplementary File S6.2c.

Treatment-limiting adverse event per individual patient	Dabrafenib plus trametinib n=8								Total
	1	2	3	4	5	6	7	8	
Pyrexia	1	1	1	1	-	-	1	-	5
Chills	1	1	1	-	-	-	-	-	3
Nausea	1	1	1	-	-	-	-	-	3
Myalgia	1	-	-	-	-	1	-	-	2
Headache	1	-	-	-	-	-	-	-	1
Liver function disorders	1	-	-	-	-	-	-	-	1
Fatigue	-	1	-	-	-	-	-	1	2
Syncope	-	1	-	-	-	-	-	-	1
Skin disorder	-	-	-	1	1	1	1	-	4
Allergic reaction	-	-	-	-	1	-	-	-	1
Tachycardia	-	-	-	-	1	-	-	-	1
Decreased appetite	-	-	1	-	-	-	-	-	1
Arthralgia	-	-	-	-	-	1	-	-	1
Malaise	-	-	-	1	-	-	-	1	2
Total	6	5	4	3	3	3	2	2	

Chapter 7

Real-world evaluation of supportive care using an electronic health record text mining tool: G-CSF use in breast cancer patients

Sylvia A. van Laar, Kim B. Gombert-Handoko, Sophie Wassenaar,
Judith R. Kroep, Henk-Jan Guchelaar, Juliëtte Zwaveling

Support Care Cancer. 2022; 30: 9181-9189

Abstract

Purpose: Chemotherapy-induced febrile neutropenia (FN) is a life-threatening and chemotherapy dose-limiting adverse event. FN can be prevented with granulocyte-colony stimulating factors (G-CSFs). Guidelines recommend primary G-CSF use for patients receiving either high (>20%) FN risk (HR) chemotherapy, or intermediate (10–20%) FN risk (IR) chemotherapy if the overall risk with additional patient-related risk factors exceeds 20%. In this study we applied an electronic health record (EHR) text-mining tool for real-world G-CSF treatment evaluation in breast cancer patients.

Methods: Breast cancer patients receiving IR or HR chemotherapy treatments between January 2015 and February 2021 at LUMC, The Netherlands, were included. We retrospectively collected data from EHR with a text-mining tool and assessed G-CSF use, risk factors and the FN and neutropenia (grade 3–4) incidence.

Results: 190 female patients were included, who received 77 HR and 113 IR treatments. In 88.3% of the HR regimens G-CSF was administered, 7.3% of these patients developed FN vs. 33.3% without G-CSF. Although most IR regimen patients had ≥ 2 risk factors, only 4% received G-CSF, of which none developed neutropenia. However, without G-CSF, 11.9% developed FN and 31.2% severe neutropenia.

Conclusions: Our text-mining study shows high G-CSF use among HR regimen patients, and low use among IR regimen patients, although most had ≥ 2 risk factors. Therefore, current practice is not completely in accordance with the guidelines. This shows the need for increased awareness and clarity regarding risk factors. Also, text mining can effectively be implemented for the evaluation of patient care.

1. Introduction

Quantifying healthcare outcomes in clinical practice is a crucial step towards the improvement of cancer patients' care [1]. The electronic health record (EHR) is a valuable source of real-world medical data, including, e.g., demographics, vital signs, laboratory data, and medication orders, which can be used for treatment evaluation [2, 3]. However, most of the information is stored in unstructured text, specifically regarding treatment outcomes, e.g., in pathology reports, and detailed adverse drug events in narrative notes [4, 5]. Manual data extraction has been the standard extraction method for EHR data, which is known to be time consuming and error-prone [3, 6]. Novel natural language processing and text-mining techniques facilitate automatized data extraction from EHR [6-8], and therefore enable the evaluation of treatments and guidelines in clinical practice.

For years, granulocyte colony-stimulating factors (G-CSF) are used to prevent chemotherapy-induced neutropenia in breast cancer patients [9]. Neutropenia is one of the most serious and common adverse events of myelosuppressive chemotherapy [10, 11]. Complications of neutropenia are fever, or febrile neutropenia (FN), due to opportunistic infections, which often require intravenous antibiotic treatment and hospitalization [12-14]. Moreover, patients developing severe FN often receive chemotherapy dose reductions or treatment delays, which is associated with worse survival outcomes [13].

Prophylactic use of granulocyte colony-stimulating factors (G-CSF), e.g., filgrastim, has shown to reduce the severity and duration of neutropenia and the incidence of FN by 50–90% [10, 14, 15]. G-CSF use can also result in adverse events, mainly mild to moderate bone pain (25–36%), but also potentially secondary myeloid neoplasms [16]. Furthermore, broad use of G-CSF was previously assumed to be a significant financial burden to the healthcare system; the cost-effectiveness of G-CSF is highly related with both the FN risk and the G-CSF costs in clinical practice [17, 18].

FN incidence is primarily related to the type and intensity of the chemotherapy regimen. Therefore, European Organisation for Research and Treatment of Cancer (EORTC) and National Comprehensive Cancer Network (NCCN) guidelines indicate that primary prevention of FN with G-CSF (PP G-CSF) should be applied if patients receive treatment with a high-risk (HR) (>20%) for FN. For patients receiving a chemotherapy regimen with an intermediate FN risk (IR) of 10–20%, other patient-related risk factors should be

considered to define if the overall FN risk exceeds 20%, and thus PP G-CSF is indicated [10, 19, 20]. Although multiple patient-related risk factors are related with an increased FN risk, e.g., age ≥ 65 years, advanced disease, and female gender, the exact risk attribution of these factors to FN is not yet defined [10]. This may complicate decision-making as to whether or not to administer PP G-CSF. It is already shown that in clinical practice, not all patients receive PP G-CSF when recommended [11, 21]. However, real-world evidence on PP G-CSF utilization linked to risk factors for FN is limited.

Even though several guidelines on G-CSF use for clinical practice are present, it is not clear to what extent these are followed. Therefore, the aim of this study is to retrospectively review guideline adherence for HR and IR chemotherapy regimens in a breast cancer population by collecting data from the EHR with text mining.

2. Methods

2.1 Patient population

All patients aged 18 years and older with breast cancer were included if they started a HR or IR chemotherapy regimen between January 2015 and February 2021 at Leiden University Medical Center (LUMC), Leiden, The Netherlands. Patients participating in the DIRECT study were excluded, since these patients by study design were not allowed to receive PP G-CSF treatment and therefore might bias results [22]. The study protocol was reviewed and approved by the Medical Ethics Review Committee of the LUMC, Leiden, who waived the need for informed consent.

2.2 Data collection method

We performed the data collection from the EHR with rule-based text-mining software (CTcue B.V., Amsterdam, The Netherlands). This tool enables extraction of structured (e.g., laboratory results and medical prescriptions) and unstructured (free-text notes) data and the immediate conversion of results into a dataset. For unstructured notes, it enables to search for (combinations of) keywords, and shows all notes matching with these results. We validated the software previously for data extraction to evaluate metastatic renal cell carcinoma treatments, which showed high accuracy ($>90\%$) for data collection from structured data [23]. All used queries for patient inclusion and data collection are available in Supplementary File S7.1. Patients were identified by a combination of the selected chemotherapy treatments in their medication history, mention of the treatment

regimen in the notes and a diagnosis treatment code for breast cancer. All patients and their regimens were manually validated within the software tool. Additionally, for data extracted from unstructured text (tumor receptor characteristics, type of treatment, G-CSF use and incidence of (febrile) neutropenia), we also performed manual validation of the data by reviewing the identified notes within the tool. G-CSF use and incidence of (febrile) neutropenia were further manually validated by EHR review.

2.3 Patient-, disease- and treatment characteristics

All risk factors from the EORTC guideline that could be evaluated in retrospect were included in this study. These were the following patient- and disease characteristics with specified cut-off values per risk factor (RF): age (RF: age >65 years), sex (RF: female gender), length and weight (RF: body surface area <2.0 m²), hemoglobin (RF: hemoglobin <12 g/dl), ALT (RF: abnormal liver transaminases | ALT>35 U/l), AST (RF: abnormal liver transaminases | AST >30 U/l), eGFR (RF: renal disease | eGFR<60 ml/min/1.73m²), absolute neutrophil count (RF: low pretreatment ANC | <2x10⁹ cells/L [24]), serum albumin (RF: albumin <3.5 g/dl), performance status (RF: performance status >0), previous treatments (RF: prior chemotherapy), treatment type (curative or palliative, RF: advanced disease/metastasis) [10].

We did not include the following risk factors: prior episodes of FN, since we estimated the risk prior to the start of cycle one; antibiotic prophylaxis, since patients in the Netherlands do not receive antibiotic prophylaxis; and cardiovascular disease, one or more comorbidities, and prior infections, since the high variability in free-text terminology combined with the uncertainty whether these risk factors are noted structurally in the EHR would result in an incomplete, and potentially incorrect risk estimation.

2.4 Outcome measurements

The primary outcome was the incidence of FN from chemotherapy initiation until 21 days after last treatment cycle. FN is defined as the ANC <0.5x10⁹ cells/L, or ANC <1.0x10⁹ cells/L, predicted to fall below 0.5x10⁹ cells/L within 48 hours, with fever or clinical signs of sepsis. Fever is defined as rise in axillary temperature >38.5°C for 1 hour [10]. Also, the incidence of grade 3 or higher neutropenia was collected, which is defined as ANC <1.0x10⁹ cells/L (Common Terminology Criteria for Adverse Events v.5.0). We included cases of FN and neutropenia if they met the definition based on structured data, or when noted in unstructured text by a treating physician.

2.5 Statistical analysis

Data management and analysis was performed using R 4.1.0 (R CoreTeam, 2021). Descriptive statistics were used to describe patient-, treatment and disease characteristics. The number of patients on HR- or IR chemotherapy regimen receiving PP G-CSF that developed neutropenia or FN were summarized in percentages and visualized in a Sankey plot. With the Student's t-test the number of risk factors between the subgroups was compared. Also, chi-square tests, or Fisher's exact test if expected frequency was lower than five, were performed to compare risk factors between subgroups.

3. Results

In total, 190 breast cancer patients, which received an IR or HR regimen between January 2015 and February 2021, were included. All patients were female and had a median age of 52.6 years (± 11.4 years). Baseline patient-, tumor- and treatment characteristics are shown in Table 7.1.

Table 7.2 presents an overview of the included chemotherapy regimens. In total, patients received 77 HR and 113 IR chemotherapy regimens. Most applied HR regimens were dose dense doxorubicin and cyclophosphamide (ddAC) $\rightarrow$ paclitaxel and carboplatin (32.5%), ddAC $\rightarrow$ paclitaxel (31.2%), and the combination of paclitaxel, doxorubicin, and cyclophosphamide (TAC, 24.7%). Most applied IR regimens were a combination of doxorubicin and cyclophosphamide (AC, 52.2%) and AC $\rightarrow$ docetaxel (25.7%).

Table 7.3 summarizes the proportion of patients that received PP G-CSF, and developed neutropenia, and FN, stratified per risk group. These results are visualized in Figure 7.1. Overall, in 37.9% of chemotherapy regimens PP G-CSF was administered at the start of the treatment regimen, 88.3% during HR treatments and 3.5% during IR regimens. The incidence of severe ($\geq$ grade 3) neutropenia in the overall cohort was 21.1%; 11.1% of the patients developed FN at least once. The incidence of neutropenia and FN combined was higher in patients treated with IR regimens (41.6%) than in HR treated patients (18.2%). However, FN incidence in both groups was comparable and around 10%. In the HR treatment group, 33.3% of the patients who did not receive PP G-CSF developed FN, in contrast to 7.3% of the patients who received PP G-CSF. Further, in the IR treatment group none of the four patients that started with PP G-CSF developed FN and 11.9% in the group who did not receive PP G-CSF.

Table 7.1. Patient-, tumor-, and treatment characteristics

	HR regimen patients, n=77 median (1 st –3 rd quartile) or n (%)	IR regimen patients, n=113 median (1 st –3 rd quartile) or n (%)
Patient characteristics		
Female gender	77 (100)	113 (100)
Age (years)	51 (38–57)	55 (48–64)
Body surface area (m ²)	1.81 (1.72–1.92)	1.85 (1.72–1.92)
Hemoglobin (U/l)	8.5 (8–8.8)	8.4 (8.0–8.8)
Absolute neutrophil count (U/l)	4.1 (3.5–5.5)	4.5 (3.5–5.8)
Albumin (g/dl)	46 (44–48)	45 (44–48)
ALT (U/l)	20 (15–26.3)	20 (15–26.5)
AST (U/l)	21 (19–25)	22 (18–26.25)
eGFR (ml/min/1.73m ²)	89 (81–90)	90 (78–90)
WHO performance status		
0	41 (53.2)	49 (43.4)
1	3 (3.9)	19 (16.8)
2	1 (1.3)	2 (1.8)
Missing	32 (41.6)	43 (38.1)
Tumor receptor characteristics		
Progesterone receptor positive	34 (44.2)	69 (61.1)
Estrogen receptor positive	42 (55.2)	89 (78.8)
Her2 receptor positive	1 (1.3)	12 (10.6)
Missing	1 (1.3)	4 (3.5)
Treatment characteristics		
Previous chemotherapy treatment	4 (5.2)	16 (14.2)
Type of treatment		
Neo-adjuvant	49 (63.6)	32 (28.3)
Adjuvant	25 (32.5)	67 (59.3)
Palliative	1 (1.3)	6 (5.3)
Missing	2 (2.6)	8 (7.0)

IR: intermediate-risk, HR: high-risk, ALT: Alanine transaminase, AST: Aspartate transaminase, eGFR: estimated glomerular filtration rate, HER2: Human epidermal growth factor receptor 2.

We analyzed the presence of additional patient-related risk factors in the IR treated group. A mean of 3.4 risk factors per patient were found, ranging from one to eight risk factors per patient (Figure 7.2A). Patients who received PP G-CSF (n=4) had a mean of 5 risk factors, and without PP G-CSF (n=109) 3.3 (p=0.13). No significant difference was found between the distribution of individual risk factors between both groups (Table 7.4).

Table 7.2. Included high- and intermediate-risk chemotherapy treatment regimens

	Number of patients per treatment, n=190, n (%)	
High-risk treatments	77 (40.5)	
ddAC	3 (3.9)	
ddAC → docetaxel	4 (5.2)	
ddAC → paclitaxel	24 (31.2)	
ddAC → paclitaxel and trastuzumab	1 (1.3)	
ddAC → paclitaxel and carboplatin	25 (32.5)	
Paclitaxel → ddAC	1 (1.3)	
TAC	19 (24.7)	
Intermediate-risk treatments	113 (59.5)	
AC	59 (52.2)	
AC → docetaxel	29 (25.7)	
Cyclophosphamide	2 (1.8)	
Docetaxel	4 (3.5)	
Docetaxel and cyclophosphamide	7 (6.2)	
FEC	1 (0.9)	
FEC, trastuzumab and pertuzumab	5 (4.4)	
FEC → docetaxel	7 (6.2)	

AC: doxorubicin and cyclophosphamide, ddAC: dose dense AC, FEC: 5-fluorouracil, epirubicin and cyclophosphamide, TAC: docetaxel, doxorubicin and cyclophosphamide.

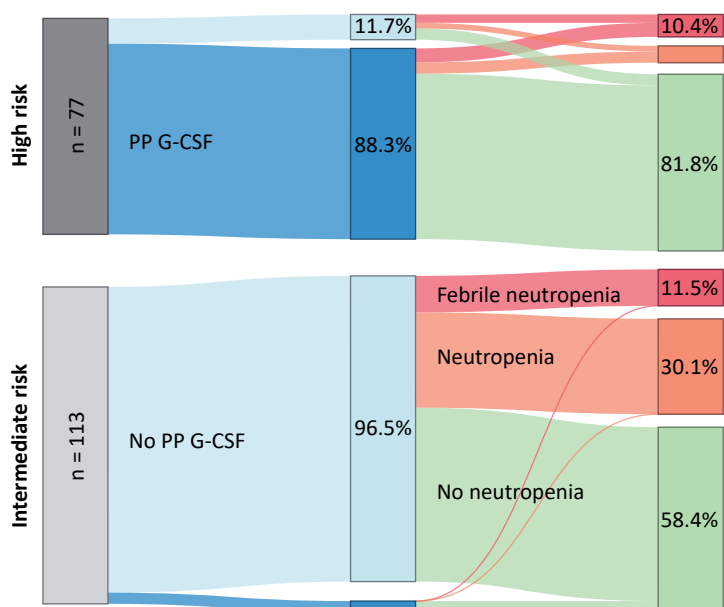


Figure 7.1. Proportion of patients that received primary granulocyte-colony stimulating factor (PP G-CSF) treatment and developed neutropenia (≥ grade 3) or febrile neutropenia stratified for intermediate- and high-risk chemotherapy treatments.

Table 7.3. Proportion of patients with high- and intermediate risk regimens who received primary prophylaxis using granulocyte colony-stimulating factor (PP G-CSF) from the start of treatment and developed febrile neutropenia, severe ($\geq$ grade 3) neutropenia without fever or no neutropenia

	PP G-CSF	Febrile neutropenia	Neutropenia	No neutropenia	Total
High-risk regimen	Yes, n (%)	5 (7.3)	4 (5.9)	59 (86.8)	68 (100)
	No, n (%)	3 (33.3)	2 (22.2)	4 (44.4)	9 (100)
	Subtotal, n (%)	8 (10.4)	6 (7.8)	63 (81.8)	77 (100)
Intermediate-risk regimen	Yes, n (%)	0 (0)	0 (0)	4 (100)	4 (100)
	No, n (%)	13 (11.9)	34 (31.2)	62 (56.9)	109 (100)
	Subtotal, n	13 (11.5)	34 (30.1)	66 (58.4)	113 (100)
	Total, n (%)	21 (11.1)	40 (21.1)	129 (67.9)	190 (100)

Table 7.4. Risk factors present in the intermediate risk group stratified by PP G-CSF use (A), and patients not receiving PP G-CSF by FN status (B)

Risk factor	PP G-CSF use			Febrile neutropenia		
	Yes n=4 n (%)	No n=109 n (%)	p-value	Yes n=13 n (%)	No n=96 n (%)	p-value
Female gender	4 (100)	109 (100)	-	13 (100)	96 (100)	-
Age >65 years	2 (50)	20 (18.3)	0.17	0 (0)	20 (20.8)	0.060
Body surface area <2 m ²	4 (100)	94 (86.2)	0.58	9 (69.2)	85 (89.5)	0.041
Missing	0	1		0	1	
Hemoglobin <12 g/dl	2 (50.0)	11 (10.1)	0.065	1 (9.1)	10 (9.1)	0.61
Absolute neutrophil count $\leq 5.2 \times 10^9/l$	2 (50.0)	65 (65.6)	0.44	7 (53.8)	58 (67.4)	0.34
Missing	0	10		0	10	
Albumin <3.5 g/dl	0 (0)	2 (2.1)	0.92	0 (0)	2 (2.3)	0.61
Missing	0	12		2	10	
Liver function: ALT ≥ 35 U/l or AST ≥ 30 U/l	2 (50)	20 (18.5)	0.17	2 (15.4)	18 (18.9)	0.55
Missing	0	1		0	1	
Kidney function: eGFR <60 ml/min/1.73m ²	0 (0)	3 (2.8)	1	0 (0)	3 (3.2)	0.67
Missing	0	3		0	3	
WHO performance status >0	1 (33.3)	20 (29.9)	0.66	1 (11.1)	19 (32.8)	0.18
Missing	1	42		4	38	
Previous treatment	2 (50)	14 (12.8)	0.095	0 (0)	14 (14.6)	0.14
Palliative treatment	1 (25.0)	5 (5.0)	0.21	0 (0)	5 (5.5)	0.55
Missing	0	8		2	6	

ALT: alanine transferase, AST: aspartate transferase, eGFR: estimated glomerular filtration rate. P-value <0.05 was defined as significant.

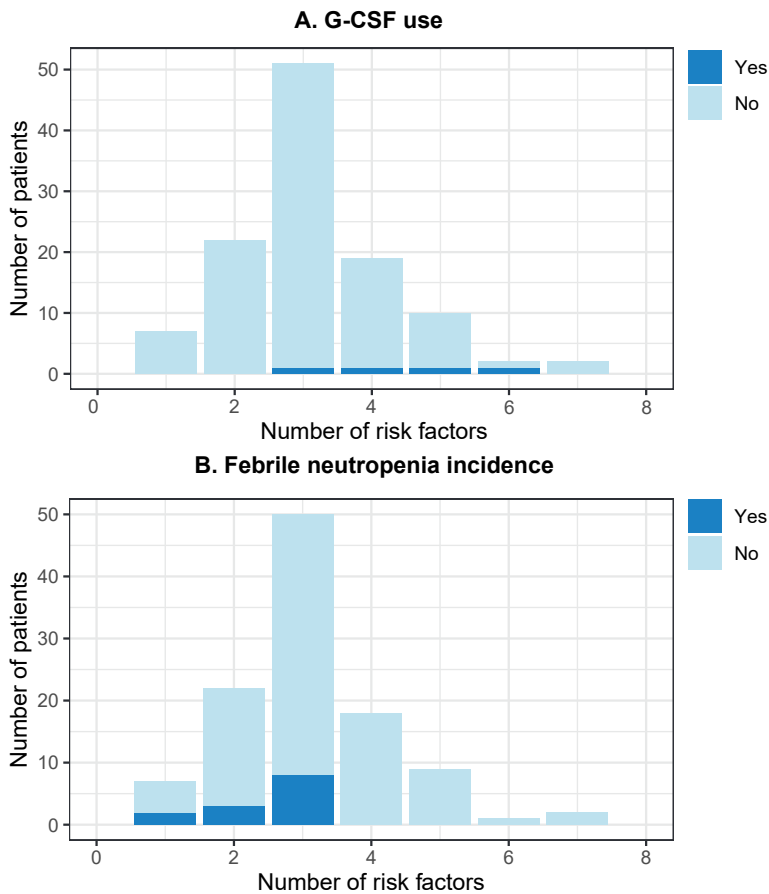


Figure 7.2. Number of confirmed risk factors per patient in the intermediate risk group stratified by G-CSF use (A), and patients not receiving PP G-CSF by FN status (B).

Of the patients on an IR regimen that did not receive PP G-CSF, thirteen (11.9%) developed FN (Table 7.3). Figure 7.2B shows that these patients had a lower mean of risk factors versus the patients who did not develop FN (2.5 versus 3.4; $p=0.0044$). Furthermore, the risk factor BSA below 2 m² was more prevalent in the group who did not develop FN (Table 7.4).

4. Discussion

In this study we investigated whether EHR text mining could be applied to evaluate G-CSF use among breast cancer patients in clinical practice. The high recall rate and the results of the within-program manual validation both indicate that adherence

to guidelines can be performed using EHR text mining. We found that, in general, G-CSFs were not as often administered as primary prophylaxis as indicated in the EORTC guideline. Not all patients (88%) who received HR regimens started with G-CSF prophylaxis. Also, even though almost all patients with IR regimens had two or more of the investigated risk factors, only 4% received PP G-CSF. This resulted in an overall neutropenia incidence of 32%, including 11% FN. Our results indicate that prophylactic treatment with G-CSF should be optimized to further prevent the occurrence of (febrile) neutropenia.

4.1 High-risk chemotherapy treatment

Despite guidelines indicating 100% of the patients on a HR regimen should receive PP G-CSF, we report a PP G-CSF use of 88%. An undertreatment of the HR population was also found by Gawade et al., who reported 76.4% PP G-CSF use [21]. Also, PP G-CSF use over time seems to be improved over the recent years [25]. In the HR regimen patients that received PP G-CSF according to the guidelines, only 7% still developed FN, which is comparable to the incidence of FN (9.5%) found in breast cancer patients receiving PP G-CSF (intention-to-treat) for five days by Clemons et al [26]. However, approximately a third of the patients without prophylaxis developed FN. This difference may be directly related to the lack of primary FN prophylaxis, however numbers in this group are small. In our hospital, most of the omissions of PP G-CSF in the HR group were unintentional. This is potentially related to the fact that G-CSFs are not prescribed and ordered as regular medication in the EHR for hospitalized patients. In the future, these errors could be prevented by incorporating G-CSF prescription into the predefined HR chemotherapy treatment protocol in the EHR.

4.2 Intermediate-risk chemotherapy treatment

Almost none (4%) of the patients treated with an IR chemotherapy received PP G-CSF. As PP G-CSF should only be administered if the cumulative FN risk per individual patient exceeds 20%, the risk estimation of patients receiving IR treatment should be based on additional risk factors [10]. The low use of G-CSF in this population is remarkable as all patients have at least one of the investigated risk factors, female gender, and most have more than two, which may indicate standard G-CSF prescription for these patients. In comparison to the 4% of PP G-CSF use in our study, both Gawade et al. and Bacrie et al. reported around 18% of PP G-CSF use in IR risk patients [21, 27]. Even though the low use of G-CSFs in this population, the FN incidence was moderate

(12%) and a substantial amount of the patients developed grade 3–4 neutropenia (31%). Higher G-CSF use could further have prevented cases of severe neutropenia, which can also result in chemotherapy dose delay or reductions [28, 29].

None of the investigated risk factors were significantly related to PP G-CSF prescriptions in our patients. It is remarkable that 20 patients (18%) aged >65 years did not receive PP G-CSF, since this is the most prominent mentioned risk factor in the guidelines [10]. However, none of the risk factors were significantly associated with the incidence of FN in the group of patients who did not receive PP G-CSF. On the contrary, more risk factors were found in the group without FN, and low BSA specifically was more present in this group. Therefore, even though the small sample sizes, it is questionable to what extent the investigated risk factors substantially contribute to FN development and G-CSF prescription in this IR group. To assess whether dose reductions were applied as alternative preventive measure, we assessed dose reductions on the initial dose and found that approximately 5% of patients started at a dose of 70–80% of the standard starting dose. Thus, in our population dose reductions play only a minor role or not at all in the risk for FN. Also, antibiotic prophylaxis for FN is not used in this population. Therefore, overall, it may have been other, more difficult to assess factors that could have contributed more to both decision-making for PP G-CSF use and real-world FN risk, e.g., risk factors as radiation therapy and not investigated risk factors [12].

Gawade et al., who compared mainly comorbidities to PP G-CSF initiation by retrospectively reviewing a large medical claims database, also showed that risk factors seemed not to influence PP G-CSF initiation in the IR group, although this was suggested in the NCCN guidelines [21]. Furthermore, Lyman et al. compared the model-predicted and physician-predicted FN risk and showed a weak correlation [30]. This underlines that there may be a difference in how physicians weigh patient-, disease- and treatment-related risk factors; therefore, they suggest the need for continuous education on FN risk factors, G-CSF toxicity, guidelines, and appropriate PP G-CSF use. Zooming in, we note that the underlying problem might be the broad definition of some risk factors and their contribution to clinical outcomes. Therefore, besides continuous education, we recommend, firstly, clarification of the risk factors that play a major role in chemotherapy-induced FN, and, secondly, clearer and simple guidelines which state how these risk factors should be weighed. Therefore, the development of a scoring system, comparable to the system developed to estimate the FN risk in patients

who receive low-risk treatments, could also be beneficial for the intermediate- and high-risk group [31].

Moreover, G-CSFs were known to be costly and as a consequence, the choice for treatment above >20% FN risk only was highly related with the cost-effectiveness [15]. However, new biosimilars are proven to be significantly cost-saving and thereby lower the threshold for the application of G-CSF and simplification of the guidelines [32, 33]. This could not only further lower the FN risk, but also the risk on severe neutropenia. Nonetheless, G-CSF use can result in adverse events, potential benefits of G-CSF use should always outweigh risks.

In this study we applied EHR text-mining software. This enabled fast, structured and pseudonymized patient inclusion and data extraction. Using this method, missing data of the included risk factors, in general, was limited. However, not all potential risk factors could (fully) be assessed retrospectively. Partially as the EHR is a secondary source and not all patient data are equally well documented in the EHR, e.g., in this study we could only report the performance status for approximately 60% of the patients. But also, to some extent, because definitions were unclear or too broad, e.g., one or more comorbidities. We still performed manual validation of the critical end-points G-CSF use and FN incidence after text mining, since these were mainly documented in unstructured text. Since only a selection of the data had to be validated, this process was faster compared to complete manual data extraction. Our study shows that a text-mining tool can be an effective method to review adherence to guidelines and that results can be used as a concrete starting point to optimize patient care.

5. Conclusion

By application of text mining to the EHR we were able to review G-CSF use in daily practice in breast cancer patients. PP G-CSF use among HR regimen patients was high, however not maximal, and undertreatment resulted in a higher incidence of FN. Most IR regimen patients had more than two risk factors, and were therefore entitled to the use of PP G-CSF. However, few received PP G-CSF which could have prevented the occurrence of FN and neutropenia. Therefore, current practice is not completely in accordance with the guidelines, in particular for patients treated with IR regimens, and may result in unnecessary toxicity for patients. We conclude that awareness of risk factors related with neutropenia should be enlarged and these risk factors could

be more clearly defined in the guidelines. Finally, our study shows that text-mining methods can be effectively implemented to review daily practice for the evaluation and improvement of patient care.

References

1. Teisberg, E., S. Wallace, and S. O'Hara, *Defining and Implementing Value-Based Health Care: A Strategic Framework*. Academic Medicine, 2020. **95**(5): p. 682-685.
2. Casey, J.A., et al., *Using Electronic Health Records for Population Health Research: A Review of Methods and Applications*. Annual Review of Public Health, 2016. **37**(1): p. 61-81.
3. Cowie, M.R., et al., *Electronic health records to facilitate clinical research*. Clin Res Cardiol, 2017. **106**(1): p. 1-9.
4. Savova, G.K., et al., *Use of Natural Language Processing to Extract Clinical Cancer Phenotypes from Electronic Medical Records*. Cancer Res, 2019. **79**(21): p. 5463-5470.
5. Liu, F., A. Jagannatha, and H. Yu, *Towards Drug Safety Surveillance and Pharmacovigilance: Current Progress in Detecting Medication and Adverse Drug Events from Electronic Health Records*. Drug Saf, 2019. **42**(1): p. 95-97.
6. 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.
7. Dreisbach, C., et al., *A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data*. Int J Med Inform, 2019. **125**: p. 37-46.
8. Delespierre, T., et al., *Empirical advances with text mining of electronic health records*. BMC Med Inform Decis Mak, 2017. **17**(1): p. 127.
9. Zardawi, S.J., I. Nordman, and N. Zdenkowski, *A retrospective analysis of nadir-neutropenia directed pegylated granulocyte-colony stimulating factor on febrile neutropenia rates in (neo) adjuvant breast cancer chemotherapy regimens*. Cancer Rep (Hoboken), 2020. **3**(5): p. e1266.
10. Aapro, M.S., et al., *2010 update of EORTC guidelines for the use of granulocyte-colony stimulating factor to reduce the incidence of chemotherapy-induced febrile neutropenia in adult patients with lymphoproliferative disorders and solid tumours*. European Journal of Cancer, 2011. **47**(1): p. 8-32.
11. Averin, A., et al., *Risk of chemotherapy-induced febrile neutropenia in patients with metastatic cancer not receiving granulocyte colony-stimulating factor prophylaxis in US clinical practice*. Support Care Cancer, 2021. **29**(4): p. 2179-2186.
12. Lee, M., et al., *Risk factors for neutropenia and febrile neutropenia following prophylactic pegfilgrastim*. Asia-Pacific Journal of Clinical Oncology, 2019. **15**(4): p. 231-237.
13. Lyman, G.H., E. Abella, and R. Pettengell, *Risk factors for febrile neutropenia among patients with cancer receiving chemotherapy: A systematic review*. Critical Reviews in Oncology/Hematology, 2014. **90**(3): p. 190-199.
14. Salmon, J.P., et al., *Febrile neutropenia (FN) and pegfilgrastim prophylaxis in breast cancer and non-Hodgkin's lymphoma patients receiving high (>20%) FN-risk chemotherapy: results from a prospective observational study*. Support Care Cancer, 2019. **27**(4): p. 1449-1457.
15. Krol, J., et al., *G-CSF in the prevention of febrile neutropenia in chemotherapy in breast cancer patients*. Onkologie, 2006. **29**(4): p. 171-8.
16. Lapidari, P., I. Vaz-Luis, and A. Di Meglio, *Side effects of using granulocyte-colony stimulating factors as prophylaxis of febrile neutropenia in cancer patients: A systematic review*. Critical Reviews in Oncology/Hematology, 2021. **157**: p. 103193.
17. Younis, T., et al., *Cost-effectiveness of febrile neutropenia prevention with primary versus secondary G-CSF prophylaxis for adjuvant chemotherapy in breast cancer: a systematic review*. Breast Cancer Res Treat, 2016. **159**(3): p. 425-32.

18. Aarts, M.J., et al., *Cost Effectiveness of Primary Pegfilgrastim Prophylaxis in Patients With Breast Cancer at Risk of Febrile Neutropenia*. Journal of Clinical Oncology, 2013. **31**(34): p. 4283-4289.
19. Becker, P.S., et al., *NCCN Guidelines Insights: Hematopoietic Growth Factors, Version 1.2020*. J Natl Compr Canc Netw, 2020. **18**(1): p. 12-22.
20. Smith, T.J., et al., *Recommendations for the Use of WBC Growth Factors: American Society of Clinical Oncology Clinical Practice Guideline Update*. Journal of Clinical Oncology, 2015. **33**(28): p. 3199-3212.
21. Gawade, P.L., et al., *Patterns of granulocyte colony-stimulating factor prophylaxis in patients with cancer receiving myelosuppressive chemotherapy*. Supportive Care in Cancer, 2020. **28**(9): p. 4413-4424.
22. de Groot, S., et al., *Fasting mimicking diet as an adjunct to neoadjuvant chemotherapy for breast cancer in the multicentre randomized phase 2 DIRECT trial*. Nat Commun, 2020. **11**(1): p. 3083.
23. van Laar, S.A., et al., *An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma*. Clinical Pharmacology & Therapeutics, 2020. **108**(3): p. 644-652.
24. Jenkins, P. and S. Freeman, *Pretreatment haematological laboratory values predict for excessive myelosuppression in patients receiving adjuvant FEC chemotherapy for breast cancer*. Annals of Oncology, 2009. **20**(1): p. 34-40.
25. Goyal, R.K., et al., *Time trends in utilization of G-CSF prophylaxis and risk of febrile neutropenia in a Medicare population receiving adjuvant chemotherapy for early-stage breast cancer*. Support Care Cancer, 2018. **26**(2): p. 539-548.
26. 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.
27. Bacrie, J., et al., *Febrile neutropenia in adjuvant and neoadjuvant chemotherapy for breast cancer: a retrospective study in routine clinical practice from a single institution*. Support Care Cancer, 2018. **26**(12): p. 4097-4103.
28. 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.
29. Liutkauskienė, 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.
30. Lyman, G.H., et al., *Assessing patients' risk of febrile neutropenia: is there a correlation between physician-assessed risk and model-predicted risk?* Cancer Med, 2015. **4**(8): p. 1153-60.
31. Klastersky, J., et al., *The Multinational Association for Supportive Care in Cancer Risk Index: A Multinational Scoring System for Identifying Low-Risk Febrile Neutropenic Cancer Patients*. Journal of Clinical Oncology, 2000. **18**(16): p. 3038-3051.
32. McBride, A., et al., *Cost-efficiency analyses for the US of biosimilar filgrastim-sndz, reference filgrastim, pegfilgrastim, and pegfilgrastim with on-body injector in the prophylaxis of chemotherapy-induced (febrile) neutropenia*. J Med Econ, 2017. **20**(10): p. 1083-1093.
33. McBride, A., et al., *Economic modeling for the US of the cost-efficiency and associated expanded treatment access of conversion to biosimilar pegfilgrastim-bmez from reference pegfilgrastim*. J Med Econ, 2020. **23**(8): p. 856-863.

Supplementary data

Supplementary File S7.1A. Inclusion criteria

Explanation:

A patient was included for a treatment if it had both a positive DBC-code, and all inclusion criteria for that particular treatment, without exclusion criteria (e.g., for AC/ddAC, a patient needed both a prescription of cyclophosphamide +/- a day from a prescription of doxorubicine, patient had NO (exclusion) prescription of docetaxel in that period, and there was a free text validation of an AC or ddAC course after 1 January 2015, e.g. though mention of "AC kuur" in the free-text notes)

Inclusion criterium	Name data point	Collected answer	Type of data extraction	Searched terms	Used for inclusion or exclusion	Time window	Comments
DBC-code*	DBC	DBC code	DBC	Specialism: 0313 (internal medicine) with Diagnosis: 811 (breast cancer)	inclusion		Diagnose behandel combinatie = diagnosis treatment combination, a code used for reimbursements in the Netherlands.
AC/ddAC treatment use	doxorubicin + cyclophosphamide	Start date cyclophosphamide	medication request	cyclophosphamide	inclusion	between 1 day before and 1 day after start date of first date of use of doxorubicine	
	cyclophosphamide + docetaxel	Start date cyclophosphamide	medication request	cyclophosphamide	exclusion	between 1 day before and 1 day after start date of first date of use of docetaxel	

Supplementary File S7.1A continues on next page.

Supplementary File S7.1A. Continued

Inclusion criterion	Name data point	Collected answer	Type of data extraction	Searched terms	Used for inclusion or exclusion	Time window	Comments
	Verification AC	ddAC AC	text search	ddAC OR dose dense AC OR ((dose dense OR dd) AND (doxorubicin OR adriamycin OR doxorubicine OR doxo) AND (cyclofosfamide OR ciclofosfamide OR endoxan OR cyclofos OR cyclo)) OR ((dose dense OR dd) AND AC) AC OR AC kuur OR ((doxorubicin OR adriamycin OR doxorubicine OR doxo) AND (cyclofosfamide OR ciclofosfamide OR endoxan OR cyclofos OR cyclo))	inclusion	> 1 January 2015	
CD treatment use	Capecitabin + doxetaxel	Start date capecitabine	medication request	capecitabine	inclusion	Between start date first date of use docetaxel and 7 days after	
	Verification CD	DC	text search	(capecitabine OR xeloda) AND (docetaxel OR docetaxol OR taxotere OR doce)	inclusion	> 1 January 2015	
FEC treatment use	Fluorouracil + epirubicin	start date fluorouracil	medication request	fluorouracil	inclusion	between 1 day before and 1 day after start date of first date of use of epirubicine	

Fluorouracil + cyclophosphamide	start date fluorouracil	medication request	fluorouracil	inclusion	between 1 day before and 1 day after start date of first date of use of cyclofosfamide
Verification FEC	FEC	text search	FEC kuur OR FEC OR ((fluorouracil OR fluoruracil OR 5-fluorouracil OR 5-FU) AND (epirubicin OR epirubicine) AND (cyclofosfamide OR cyclofos OR cyclo OR endoxan OR ciclofosfamide))	inclusion	
TAC treatment use	doxorubicine + docetaxel	start date doxorubicine	doxorubicine	inclusion	
cyclophosphamide + docetaxel	Start date cyclophosphamide	medication request	cyclofosfamide	inclusion	
cyclofosfamide + doxorubicine	Start date cyclophosphamide	medication request	cyclofosfamide	exclusion	
Verification TAC	TAC	text search	TAC OR TAC kuur OR TAC kuur OR ((docetaxel OR docetaxol OR taxotere OR doce)AND (doxorubicine OR adriamycin OR doxorubicin OR doxo) AND (cyclofosfamide OR ciclofosfamide OR endoxan OR cyclofos OR cyclo)	inclusion	> 1 January 2015

Supplementary File S7.1A continues on next page.

Supplementary File S7.1A. *Continued*

Inclusion criterion	Name data point	Collected answer	Type of data extraction	Searched terms	Used for inclusion or exclusion	Time window	Comments
TC treatment use	docetaxel + cyclofosfamide	start date docetaxel	medication request	docetaxel	inclusion		
	docetaxel + doxorubicine	start date docetaxel	medication request	docetaxel	exclusion		
	cyclofosfamide + doxorubicine	Start date cyclophosphamide	medication request	cyclofosfamide	exclusion		
	verification TC	TC	text search	((docetaxel OR docetaxol OR taxotere OR doce) AND (cyclofosfamide OR ciclofosfamide OR endoxan OR cyclofos OR cyclo)) OR TC kuur	inclusion	>1 January 2015	
monotherapy cyclophosphamide treatment use	cyclofosfamide	Start date cyclophosphamide	medication request	cyclofosfamide	inclusion		
	cyclofosfamide + doxorubicine	Start date cyclophosphamide	medication request	cyclofosfamide	exclusion		
	cyclofosfamide + docetaxel	Start date cyclophosphamide	medication request	cyclofosfamide	exclusion		
	verification cyclophosphamide	cyclophosphamide	text search	cyclofosfamide OR ciclofosfamide OR endoxan OR cyclofos OR cyclo	inclusion	>1 January 2015	

Monotherapy docetaxel treatment use	docetaxel	start date docetaxel	medication request	docetaxel	inclusion
	docetaxel + doxorubicine	start date docetaxel	medication request	docetaxel	exclusion
	docetaxel + cyclofosfamide	start date docetaxel	medication request	docetaxel	exclusion
	docetaxel + pertuzumab	start date docetaxel	medication request	docetaxel	exclusion
	docetaxel + capecitabine	start date docetaxel	medication request	docetaxel	exclusion
	Verification docetaxel	docetaxel	text search	docetaxel OR docetaxol OR taxotere OR doce	inclusion
					>1 January 2015

Supplementary File S7.1B. Data collection per treatment

Name data point	Collected answer	Type of data extraction	Searched terms/keywords	Keyword search restricted to	Time window	Comments
G-CSF use						
primair GCSF	Yes	text search	Filgrastim OR filgrastim OR accofil OR neupogen OR pegfilgrastim OR fulphila OR pelgraz OR pelmeg OR ziextenzo OR neulasta OR lipegfilgrastim OR lipegfilgrastim OR lonquex OR G-CSF OR granulocyte-colony-forming stimulating factor OR koloniestimulerende factoren OR primair G-CSF OR GCSF			Between 1 month before verification treatment AND 1 month after verification treatment
	No					
All cases found are manually checked within the Ctcue user window and by manual EHR review						
(Febrile) Neutropenia						
Neutropenia	0.5-1.0	measurements	Neutrofiele Granulocyten OR neutrofielen between 0.5 AND 1			Newest result between 1 week before verification treatment AND 1 month after stop date treatment
	<0.5		Neutrofiele Granulocyten OR neutrofielen below 0.5			
All cases found are manually checked within the Ctcue user window and by manual EHR review						

Fever	Yes	Measurements Vital signs	temperatuur patient OR patient temperatuur OR temperatuur ≥38 Temp ≥38	
Verification FN	FN No FN	text search	Febrile neutropenie OR Neutropenic Fever OR neutropene Koorts OR neutropenie OR neutropene	Result between verification treatment AND 1 month after stop date treatment
Patient characteristics				
Length	value	vital signs	length	Most recent result before start treatment
Weight	value	vital signs	gewicht	Most recent result before start treatment
eGFR	value	laboratory measurement	eGFR MDRD OR eGFR CKD-EPI	Most recent result before start treatment
creatinin	value	laboratory measurement	creatinine	Most recent result before start treatment
ALAT	value	laboratory measurement	ALAT OR ALAT GPT	Most recent result before start treatment
ASAT	value	laboratory measurement	ASAT OR ASAT GOT	Most recent result before start treatment
Hemoglobin	value	laboratory measurement	Hemoglobine OR Hb	Most recent result before start treatment

Supplementary File S7.1B. *Continued*

Name data point	Collected answer	Type of data extraction	Searched terms/keywords	Keyword search restricted to	Time window	Comments
albumin	value	laboratory measurement	albumine		Most recent	result before start treatment
neutrophil count	value	laboratory measurement	neutrofielen OR neutrofiële granulocyten		Most recent	result before start treatment
Disease characteristics						
Progesterone receptor	Positive	text search	progesteronreceptor positief OR PR positief OR progesteron receptor positief OR oestrogeen en progesteronreceptor positief			Selection order based on answer order, only if unknown answer will be manually validated
	negative					
	Unknown		progesteron receptor negatief OR progesteronreceptor negatief OR PR negatief			
			progesteron OR progesterone OR progesteronreceptor			

HER2 receptor	Positive negative Unknown	text search	her2 positief OR her2neu positief OR her2neu receptor positief her2neu receptor negatief OR her2neu negatief OR her2 negatief OR er2-neu negatief her2 OR her2neu OR her2neu receptor OR her2receptor	Selection order based on answer order, only if unknown answer will be manually validated
Estrogen receptor	Positive negative Unknown	text search	oestrogeenreceptor positief OR oestrogeen receptor positief OR oestrogeen en progesteronreceptor positief OR oestrogeenreceptor en progesteronreceptor positief ER positief oestrogeenreceptor negatief OR oestrogeen receptor negatief OR er negatief oestrogeenreceptor OR oestrogeen OR estrogen	Selection order based on answer order, only if unknown answer will be manually validated
Triple negative	triple negatief	text search	triple negative breast cancer OR TNBC OR triple negatief OR triple negatieve OR triple negatieve OR triple negative	Selection order based on answer order, only if unknown answer will be manually validated

Chapter 8

Liver and kidney function in patients with Covid-19 treated with remdesivir

Sylvia A. van Laar, Mark G.J. de Boer, Kim B. Gombert-Handoko,
Henk-Jan Guchelaar, Juliëtte Zwaveling, on behalf of the
LUMC-Covid-19 research group

Br J Clin Pharmacol. 2021; 87(11): 4450-4454

Abstract

For the treatment of Covid-19 patients with remdesivir, poor renal- and liver function were both exclusion criteria in randomized clinical trials (RCTs) and contra-indication for treatment. Also, nephrotoxicity and hepatotoxicity are reported as adverse events. We retrospectively reviewed renal- and liver functions of covid-19 103 patients who received remdesivir in the 15 days after treatment initiation. Approximately 20% of the patient population met RCT exclusion criteria. In total, 11% of the patients had a decrease in estimated glomerular filtration rate larger than 10 ml/min/1.73m². Also, 25% and 35% had increased alanine transaminase and aspartate transaminase levels, respectively. However, serious adverse events were limited. Therefore, based on these preliminary results, contra-indications based on kidney- and liver function should not be absolute for remdesivir treatment in patients with Covid-19 if these functions are monitored regularly. A larger patient cohort is warranted to confirm our results.

1. Introduction

Nephrotoxicity and hepatotoxicity were reported as adverse drug events in patients with coronavirus disease 2019 (Covid-19) treated with remdesivir a viral RNA-dependent, RNA polymerase inhibitor [1, 2]. Renal toxicity due to remdesivir is attributed to two different mechanisms. First, remdesivir triphosphate itself has low potential for mitochondrial toxicity as it weakly inhibits the mammalian DNA and RNA polymerases, and therefore may cause mitochondrial injury in renal tubular epithelial cells [2, 3]. However, after treatment with similar type of antiviral drugs (e.g. tenofovir), this adverse effect only occurred after long-term treatment [4]. Second, the intravenous preparation of remdesivir contains solfobutylether- β -cyclodextrin (SBECD) as vehicle, since remdesivir itself has poor water solubility. Accumulation of SBECD was found in patients with a creatinine clearance <50 ml/min, and associated with liver necrosis and obstruction of renal tubules in animal studies with high doses (2, 3, 5). However, a clinical trial investigating remdesivir as treatment for Ebola did not report any severe hepatotoxicity or nephrotoxicity [5].

Nonetheless, Covid-19 patients with impaired kidney- (estimated glomerular filtration range [eGFR] <30 or <50 ml/min/1.73m²) or liver function (alanine transaminase [ALT] or aspartate transaminase [AST] five times upper limit of normal [ULN]; ALT male: 45 U/l, female: 35 U/l; AST male: 35 U/l, female: 30 U/l) were excluded from randomised clinical trials (RCTs) with remdesivir [1, 6, 7]. These impaired functions were included as contra-indications for treatment in the summary of product characteristics [1, 8].

We report changes in renal- and liver functions during treatment with remdesivir in a Covid-19 patient population, regardless of the initial laboratory values. Hereby, additional information was obtained about the incidence and severity of nephro- and hepatotoxicity due to remdesivir treatment in clinical practice.

2. Patients and methods

We included hospitalized (Leiden University Medical Center, The Netherlands) adult patients on a regular ward with oxygen supplementation who started 5-day remdesivir treatment (intention to treat) between August 17 and November 4, 2020. We received a waiver for the need of informed consent from the Medical Ethics Review Committee

of LUMC, Leiden. All patients had a severe acute respiratory syndrome coronavirus-2 infection confirmed by polymerase-chain-reaction. Per patient, age, sex, remdesivir prescription data, and laboratory parameters (eGFR, AST, ALT) were extracted from the electronic health record using CTcue text mining software (CTcue B.V., Amsterdam, The Netherlands). We determined renal- and liver function at start of remdesivir, and change of parameters in a follow-up period of 15 days after remdesivir treatment initiation, both per patient and as mean change per day. Adverse event classification was according to the Common Terminology Criteria for Adverse Events version 5.0. Patients were excluded if no laboratory measurements were available.

3. Results

In total, 103 patients were included in this study, of whom at least one of the laboratory measurements was available, the majority of them being male (68%), and the median (interquartile range, IQR) age was 64 (56–75) years (Table 8.1).

Figure 8.1 shows the change in eGFR, AST and ALT compared to the baseline measurements per individual patient and mean change per day after remdesivir initiation.

Table 8.1. Patient characteristics

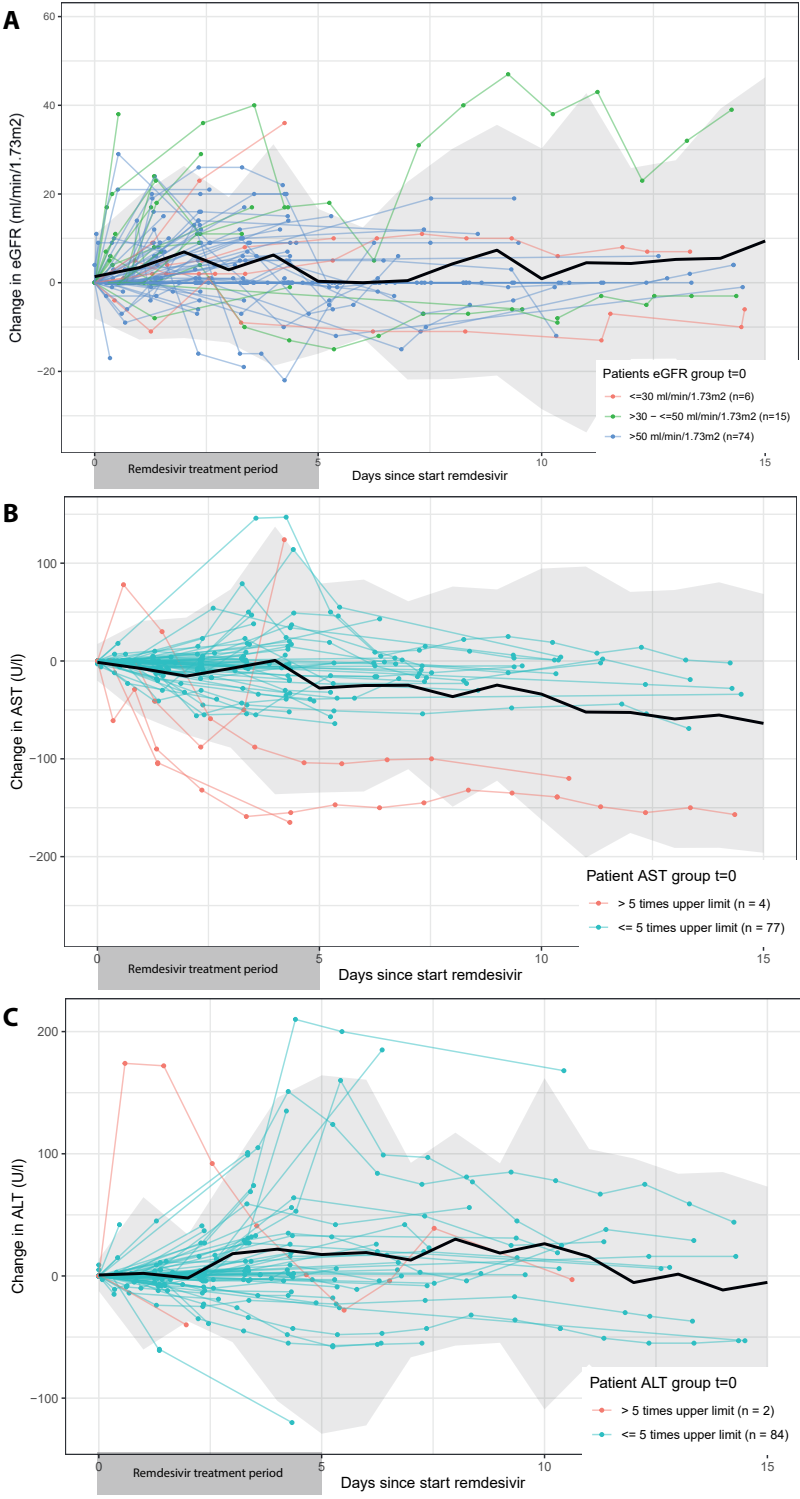
	Total population n=103
Sex, male, n (%)	70 (68)
Age, median (1 st & 3 rd qu.)	64 (56–75)
Baseline eGFR	n=95
eGFR, median (1 st & 3 rd qu.)	74.0 (54.5–87.0)
>50 ml/min/1.73m ² , n (%)	74 (78)
30–50 ml/min/1.73m ² , n (%)	15
<30 ml/min/1.73m ² , n (%)	6
Baseline ALT	n=95
ALT, median (1 st & 3 rd qu.)	32.0 (22.0–55.8)
<ULN, n (%)	54
ULN - 5x ULN, n (%)	36
>5x ULN, n (%)	5
Baseline AST, median (1 st & 3 rd qu.)	n=81
<ULN, n (%)	33
ULN - 5x ULN, n (%)	44
>5x ULN, n (%)	5

At start of treatment, the median baseline eGFR in 95 patients was 74.0 (IQR 54.5–87.0) mL/min/1.73m². In total, ten (10.5%) patients had a decrease in eGFR of more than 10 mL/min/1.73m² and the maximum decrease in eGFR was 21 mL/min/1.73m². Of all patients, 21 (22.1%) started treatment with an eGFR below the trial exclusion limit of 50 mL/min/1.73m² and two of them (9.5%) had a decrease in eGFR of more than 10 mL/min/1.73m² (Figure 8.1A).

The baseline median AST was 40.0 (IQR 27.0–58.5) U/L in 81 patients, of which 33 patients started with a normal AST and 48 patients with AST above ULN, including four patients out of the trial exclusion limit of 5x ULN. After start with remdesivir, thirteen patients (39%) with baseline AST in normal range had an increase in AST > ULN (grade 1 adverse event), one patient (3%) had an increase in AST above 3x ULN (grade 2 adverse event) and one patient AST 5x > ULN (grade 3 adverse event). Of the 48 patients where baseline was initially above ULN, two of them met the exclusion limit, five (10.4%) had an increase of 1.5x baseline (grade 1 adverse event). The maximum increase in AST was 147 U/L (Figure 8.1B).

The baseline median ALT was 32.0 (IQR 22.0–55.8) U/L in 95 patients, 54 started remdesivir treatment with a normal ALT and 41 with ALT above ULN, of which two patients with ALT > 5x ULN. In total 19 (35%) of patients starting with a normal ALT had an increase of ALT > ULN (grade 1 adverse event), three patients (6%) had ALT > 3x ULN (grade 2) and one patient had ALT > 5x ULN (grade 3). Nine patients (22%) that started remdesivir treatment with ALT above ULN had an increase in ALT > 1.5x baseline (grade 1), however in none of them > 3.0x baseline (grade 2) was reported. The maximum increase in ALT was 210 U/L (Figure 8.1C).

We included eleven patients with at least one contra-indication based on kidney or liver function. Of these, nine patients received the complete full dose five-day treatment. In one patient the remdesivir treatment was ended when dismissed from the hospital and in one patient treatment was ended due to hepatitis, which was suspected to be viral or ciprofloxacin induced. However, in none of these patients, remdesivir-induced toxicity was a reason to end treatment.



4. Discussion and conclusion

Overall, in 103 hospitalized Covid-19 patients who received a maximum of 5 days remdesivir treatment, no severe nephrotoxicity and in two patients grade 3 hepatotoxicity was found in the first fifteen days after treatment initiation.

The incidence of decreased glomerular filtration rate was comparable with data reported in the RCTs last year [6, 7]. A quarter of our population had AST elevation and a third ALT elevation, which is more than the adverse events reported in RCTs, 3.4–5% and 2.3–6%, respectively. However, Beigel et al. only reported adverse events of grade 2 or higher; Wang et al. and Goldman et al. did not specify the grade of adverse events included in the study [1, 6, 7]. Most of the liver transaminase elevation in patients in our study was mild (grade 1). Only 1% showed grade 3 elevation, whereas Goldman et al. reported grade 3 and 4 adverse events in 2–6% of the patients [7]. None of the patients that started with transaminases above ULN, including patients meeting exclusion criteria, had grade 2 or higher transaminase elevation.

It is important to emphasize that remdesivir is not the only independent factor which could cause nephrotoxicity or hepatotoxicity in these patients. Both Covid-related and non-Covid related risk factors are known, of which age of 65 years and older is one [9–10]. However, in our patients age was not identified as an independent risk factor for nephro- or hepatotoxicity. Remdesivir is extensively metabolized; for the prodrug and its metabolites, respectively, no and modest (<2-fold) accumulation was observed after multiple dosing of 150 mg for 14 days [11]. Also, comparable clearance was shown between a patient with an impaired and a patient with normal kidney function [12]. Therefore, although remdesivir metabolites are mostly renally excreted, we do not expect that poor renal function will cause clinically relevant blood level elevation within five days of treatment resulting in increased toxicity or efficacy.

Figure 8.1. Changes in renal and liver function parameters after start of remdesivir treatment in patients with Covid-19. The data originates from the electronic health record of one hospital (Leiden University Medical Center, The Netherlands). In this figure changes in eGFR (Figure 8.1A), AST (Figure 8.1B) and ALT (Figure 8.1C) laboratory measurements compared to baseline are shown. Baseline measurement was the most recent before start (max. 30 days), or if not available, the first measurement within 24 hours after the remdesivir prescription in the electronic health record. All colored data represent individual patient measurements. Patients in blue would have been excluded from clinical trials, based on their baseline measurements (eGFR: <50 ml/min/1.73m², AST: >5 times upper limit of normal, ALT: >5 times upper limit of normal). In black the mean change per day is shown including a 95% confidence interval in grey.

In a real-world population, we found that the incidence of decreased eGFR during treatment with remdesivir was comparable to the RCTs and was at most moderate. Furthermore, we reported a higher incidence of patients with non-severe transaminase elevation. however, with only two patients with a reversible severe adverse event, this was not more than in RCTs. Patients meeting contra-indications seem not to develop more renal- or hepatotoxicity on remdesivir treatment. Therefore, there may be no reason for these contra-indications to be absolute. However, monitoring remains necessary. In case of a further decline of organ functions during treatment with remdesivir, we believe it is the role of the clinician to assess the origin of the impairment.

Our observations indicate that kidney- and liver dysfunction should not be an absolute contra-indication for the use of remdesivir in Covid-19 patients, and that by regularly monitoring kidney- and liver function, treatment with remdesivir can be justified in these patients. Verification of these findings in a larger cohort is needed.

References

1. Beigel, J.H., et al., *Remdesivir for the Treatment of Covid-19 — Final Report*. New England Journal of Medicine, 2020. **383**(19): p. 1813-1826.
2. Adamsick, M.L., et al., *Remdesivir in Patients with Acute or Chronic Kidney Disease and COVID-19*. J Am Soc Nephrol, 2020. **31**(7): p. 1384-1386.
3. Ackley, T.W., et al., *A Valid Warning or Clinical Lore: An Evaluation of Safety Outcomes of Remdesivir in Patients with Impaired Renal Function from a Multicenter Matched Cohort*. Antimicrobial Agents and Chemotherapy, 2020: p. AAC.02290-20.
4. Herlitz, L.C., et al., *Tenofovir nephrotoxicity: acute tubular necrosis with distinctive clinical, pathological, and mitochondrial abnormalities*. Kidney Int, 2010. **78**(11): p. 1171-7.
5. Mulangu, S., et al., *A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics*. N Engl J Med, 2019. **381**(24): p. 2293-2303.
6. 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.
7. 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.
8. Sciences, G., *Summary of Product Characteristics: Veklury (remdesivir)*. 2020.
9. Nadim, M.K., et al., *COVID-19-associated acute kidney injury: consensus report of the 25th Acute Disease Quality Initiative (ADQI) Workgroup*. Nature Reviews Nephrology, 2020. **16**(12): p. 747-764.
10. Mercado, M.G., D.K. Smith, and E.L. Guard, *Acute Kidney Injury: Diagnosis and Management*. Am Fam Physician, 2019. **100**(11): p. 687-694.
11. Humeniuk, R., et al., *Safety, Tolerability, and Pharmacokinetics of Remdesivir, An Antiviral for Treatment of COVID-19, in Healthy Subjects*. Clinical and Translational Science, 2020. **13**(5): p. 896-906.
12. Yan, V.C. and F.L. Muller, *Captisol and GS-704277, but Not GS-441524, Are Credible Mediators of Remdesivir's Nephrotoxicity*. Antimicrobial Agents and Chemotherapy, 2020. **64**(12): p. e01920-20.

Part IV

General discussion and summaries

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 (DBC's), 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 automatized 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 it 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 or 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 algorithm 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. Liutkauskienė, 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.

Chapter 10

English summary

Nederlandse samenvatting

English summary

In the Netherlands, approximately 120,000 people are diagnosed with cancer, and 45,000 die from cancer on a yearly basis. A significant amount of these patients receive a treatment with anti-cancer drugs. Over the last decennia, the number of drug treatment options has greatly expanded. Next to chemotherapy, targeted therapies such as tyrosine kinase inhibitors and monoclonal antibodies, including immune checkpoint inhibitors, have become available.

The randomized controlled trial (RCT) has been the golden standard to investigate treatment efficacy and safety before marketing authorization, due to their strict in- and exclusion criteria, highly protocolized treatment regimens, and rigid statistical analysis plan. However, this comes with the disadvantage of limited external validity, and limited insights in, for example, rare or long-term adverse events (AEs). Also, due to the high unmet medical need, treatments can receive approval based on surrogate endpoints as progression-free survival (PFS) and recurrence-free survival (RFS), since it can take years before median overall survival is reached. And, the number of initial registration trials with a single-arm or non-randomized design is rising. Therefore, a knowledge gap between the trial and clinical practice remains.

The interest in real-world data (RWD), healthcare data generated in daily clinical practice, has risen, since RWD can be complementary to trial results. The electronic health record (EHR) is an important source of longitudinal patient data for the evaluation of oncologic treatments. However, since most RWD in EHR is stored in unstructured free-text notes, manual chart review has been the standard method of data extraction, which is highly time-consuming and error-prone. Text mining is a range of techniques that can automatize information extraction from unstructured text. This might improve the efficiency of the collection of data from EHR.

The aim of this thesis was to investigate if a text-mining tool is suitable for the collection of RWD from EHRs to evaluate cancer treatments in clinical practice.

In general, the amount of RWD available of oncologic patients and their disease and treatment trajectory has significantly grown. All RWD sources have their own strengths and limitations and therefore can be used to investigate various aspects of oncologic drug treatments in daily practice. In **Chapter 2** we have provided a non-exhaustive overview of the most used RWD sources, including traditional sources

as the case report form, electronic health records, administrative claims data, and more emerging fields as patient reported outcomes, cancer registries, mobile applications, wearable devices, and social media data. Data from these sources can, already, generate significant insights when used to estimate real-world effectiveness, monitor pharmacovigilance, calculate cost-effectiveness, describe general treatment utilization patterns, and support personalized treatment choice of cancer treatments. However, for RWD to reach its full potential in the evaluation of oncologic treatments, sources have to be more harmonized and complete, the connection between RWD sources has to be improved and artificial intelligence could be used to further enhance insights, while keeping patient privacy in mind.

In **Chapter 3** we validated the text-mining tool CTcue for data collection from EHRs in a cohort of 100 metastatic renal cell carcinoma (mRCC) patients that received a total of 175 systemic anti-cancer treatments. Both by text mining and manual chart review patients were identified and baseline patient characteristics, disease characteristics, and treatment outcomes were extracted. The comparison between the data sets showed that patients could be adequately identified and therefore included with the text-mining tool. Furthermore, treatment outcomes, including the overall survival (OS) and PFS, could be determined with no significant difference between median survival rate. Also, all structured data could be collected with high accuracy. However, information extraction of unstructured data, for example comorbidities and adverse drug events, showed varying results on accuracy (F1-scores ranging from 51.4–100). This showed that some queries still need to be optimized. Finally, this study showed that the data extraction by text mining was seven times faster, and only took 12 minutes per patient spent on data validation versus 86 minutes by manual chart review. These findings show that using a text-mining tool the efficiency of RWD collection from EHR can be improved considerably, especially since the queries have the potential to be reused in other hospitals and other oncologic populations.

After validation of the text-mining tool, it was used to evaluate multiple treatment aspects. Part 2 focussed on the treatment effectiveness and treatment patterns.

First, in **Chapter 4** text mining EHR was used to evaluate mRCC treatments of patients that were treated both in the Leiden University Medical Center and the Haga Teaching Hospital. Between January 2015 and May 2020, 122 patients started with systemic anti-cancer treatment. Most of these patients started on pazopanib (n=70), sunitinib (n=34), and nivolumab plus ipilimumab (n=28). Of the patients that finished their

first-line treatment, 64% received follow-up treatment; subsequent, 62% of the patients that finished second-line treatment continued on a third-line. This study showed a multitude of treatment paths, however most used treatments in the second- and third line were cabozantinib and nivolumab. Furthermore, for first-line pazopanib and sunitinib we were able to determine the median PFS (pazopanib: 16.3 months; sunitinib: 6.9 months) and median OS (pazopanib: 39.3 months; sunitinib: 28.1 months). These overall survival outcomes were probably better than expected from the pivotal trials. The extensive follow-up treatments patients received, may have contributed to the improved outcomes. Furthermore, the used EHR text-mining method was easily applied for evaluation of treatments in two hospitals, which may be useful to expand in the future, especially in the rapidly evolving field of renal cell carcinoma treatments.

Next, a part of the queries used in **Chapter 4** were also applied to review patterns and effectiveness of population of patients with metastatic hepatocellular carcinoma (HCC). However, as hepatocellular carcinoma has many known potential prognostic factors, these were added in this study in **Chapter 5**, to further investigate which patients are most likely to benefit from treatment. The population of 94 patients included between 2015 and 2021 first showed that sorafenib was the most used first-line treatment (88% of the patients). Only a limited number of patients (16%) started a second-line treatment, mostly regorafenib. This finding was not unexpected, as only from 2017 onwards systemic treatments were registered next to sorafenib. Further, patients receiving first-line sorafenib had a median PFS of 4.1 months and median OS of 10.5 months. However, at least 28% of the patients did not meet eligibility criteria. Specifically patients with a Child-Pugh score of B or C were associated with worse outcomes, and these patients in general might not clinically benefit of treatment. This study showed that reuse of queries in another population was possible, and that the collection of RWD for the evaluation of the current, but also upcoming HCC treatment patterns and outcomes can be performed efficiently with our text-mining method.

Part 3 focussed on RWD on treatment safety. Even though new drug treatments are monitored before the marketing approval, post approval monitoring is needed to identify AEs occurring in subgroups, especially patients not meeting trial eligibility criteria, AEs that occur later after the start of a medication, and rare events.

Compared to palliative treatments, the acceptance of AEs is lower for adjuvant treatments in the curative setting. For melanoma, only since 2018 three adjuvant treatments are registered. Therefore, we focussed in **Chapter 6** on the tolerability and safety of

these adjuvant melanoma treatments. For tolerability, the tool was used to identify the most relevant part of the EHR in regards of the end of treatment; for safety, text mining was used to identify common AEs. The study included 122 patients, 54 on nivolumab, 48 on pembrolizumab and 20 on dabrafenib plus trametinib. When first focussing on the reasons to terminate treatment – the tolerability – it showed that patients receiving dabrafenib plus trametinib significantly more often ended treatment due to toxicity compared to patients on nivolumab and pembrolizumab ($p=0.024$). Treatment-limiting toxicity from dabrafenib plus trametinib was primarily a combination of reversible AEs, including pyrexia and fatigue. For the immune checkpoint inhibitors it were mainly immune-related AEs. Regarding safety, this study showed results comparable to the clinical trial data, except for a higher incidence of treatment-limiting AEs for both nivolumab and dabrafenib plus trametinib, thyroid-related irAEs for nivolumab and pembrolizumab, and some mild AEs. Furthermore, to have an indication of effectiveness, the recurrence-free survival at year 1 was estimated and found to be comparable to trial results. Overall, this study showed that an extensive text-mining approach can be a valuable method to collect real-world safety data on cancer treatments.

Furthermore, some AEs are severe but can be prevented with prophylactic treatment. Granulocyte colony-stimulating factors (G-CSFs) to prevent chemotherapy induced febrile neutropenia (FN) is one of those. However, primary prevention is only indicated for high risk patients in guidelines. By application of text mining to the EHR we were able to review G-CSF use in daily practice in breast cancer patients in **Chapter 7** by not only identifying patients and treatments from structured data, but also FN risk factors, G-CSF used and incidence of FN from unstructured data. Among the 77 patients receiving high-risk treatment, 88% started with G-CSF treatment. The undertreatment resulted in a higher incidence of FN. Most of the 113 patients on intermediate risk treatments, had two or more risk factors, they probably should have started with primary G-CSF prevention, however, few did. Therefore, we concluded that G-CSF use in current practice is not completely in accordance with the guidelines, in particular for patients treated with intermediate risk treatment regimens, and may result in unnecessary toxicity for patients. Additionally, the awareness of risk factors related with neutropenia should be enlarged and these risk factors could be more clearly defined in the guidelines. Finally, text-mining methods can be effectively implemented to review daily practice for the evaluation and improvement of patient care.

At last, the data-extraction method was used ad hoc, for a fast estimation of the severity of AEs of a potential treatment of Covid-19. In **Chapter 8** we reacted on a signal that the treatment of Covid-19 with remdesivir could result in renal- and liver damage and investigate the changes in renal- and liver function after treatment initiation. Reviewing the renal- and liver functions of 103 Covid-19 patients in the first 15 days after remdesivir initiation showed that at least 20% would not have met trial eligibility criteria. Also, even though up to one-third of the patients had changes in their renal- and liver function, a limited amount could be labelled as serious. Our observations indicate that kidney- and liver dysfunction should not be an absolute contra-indication for the use of remdesivir in Covid-19 patients, and that by regularly monitoring kidney- and liver function, treatment with remdesivir can be justified in these patients. However, verification of these findings in a larger cohort was needed. In addition, this study nicely illustrates that text mining can be easily applied to address ad hoc questions arising from daily clinical practice.

In **Chapter 9** the methods and findings of the studies in this thesis are discussed and future perspectives are presented. This thesis, in general, showed that text mining is a suitable method to collect RWD and review oncologic drug treatments, including treatment patterns, effectiveness, safety, prognostic factors and guideline adherence. Even though the EHR is an extensive RWD source on systemic anti-cancer treatments, it still comes with challenges in regards to secondary data use, especially combined with text mining. Not only the variability, but also omissions, inconsistencies and errors in the EHR influence the accuracy of text mining. Therefore, further improving the data quality in the EHR is essential to establish meaningful real-world evidence. Also, the next step in improving the text mining might be by introducing machine-learning based natural language processing queries based on large datasets. Plus, guidelines how to use EHR text mining should be developed in order to standardize steps as data validation and cut-off values. Finally, RWD should be actively used to create insights, and be an active part of treatment evaluation, in regards to treatment policies, costs and shared decision making.

In the end, RWD must result in better insights into all aspects of cancer treatments, especially what treatment would benefit an individual patient best in every phase of their disease, and therefore optimize systemic cancer therapy.

Nederlandse samenvatting

In Nederland worden jaarlijks ongeveer 120.000 patiënten gediagnostiseerd met kanker en overlijden 45.000 patiënten aan kanker. Een significant deel van deze patiënten met kanker krijgt een behandeling met geneesmiddelen. In de laatste decennia is het aantal beschikbare geneesmiddelen hiervoor enorm uitgebreid. Naast chemotherapie is er nu ook nu ook *targeted therapy*, doelgerichte therapie, beschikbaar. Dit zijn de tyrosine kinase remmers en de monoklonale antilichamen, inclusief de checkpointremmers.

De *randomized controlled trial* (RCT, gerandomiseerd onderzoek) is de gouden standaard om onderzoek te doen naar de effectiviteit en veiligheid van geneesmiddelen voor het verkrijgen van marketingautorisatie, omdat ze strikte in- en exclusiecriteria, hoog geprotocolleerde behandelregimes en een rigide statistisch analyseplan hebben. Echter, de beperkte externe validiteit en het beperkte inzicht in bijvoorbeeld zeldzame of langtermijn bijwerkingen zijn nadelen. Door de hoge medische urgentie van deze nieuwe behandelingen, worden behandelingen op basis van surrogaat eindpunten als progressievrije overleving (PFS) en recidievrije overleving (RFS) goedgekeurd, omdat het jaren kan duren voordat de mediaan van de algehele overleving is bereikt.. Daarnaast neemt het aantal registratietrials met maar één studiearm of een niet-gerandomiseerd design toe. Door deze snelle ontwikkelingen, blijft de kennis over de effecten van geneesmiddelen in de praktijk vaak nog achter.

De interesse in *real-world data* (RWD), gezondheidsdata gegenereerd in de dagelijkse praktijk, is toegenomen, aangezien deze RWD aanvullend kan zijn aan de trialresultaten. Het elektronische patiëntendossier (EPD) is een belangrijke bron van longitudinale patiëntendata voor de evaluatie van oncologische behandelingen. Echter, sinds de meeste RWD in het EPD is opgeslagen in ongestructureerde, vrije tekstnotities, is handmatig dossieronderzoek lang de standaardmethode voor dataextractie geweest. Dit is zowel tijdrovend als foutgevoelig.

Text mining is een proces dat een reeks aan technieken omvat waarmee geautomatiseerd informatie uit ongestructureerde teksten geëxtraheerd kan worden. Hierdoor kan mogelijk ook de dataverzameling uit EPD's efficiënter plaatsvinden. Het doel van deze thesis was om te onderzoeken of een text-mining programma geschikt is voor data-extractie uit EPD's om op deze wijze de geneesmiddelbehandelingen van kanker in de praktijk te evalueren.

In het algemeen is de hoeveelheid RWD die beschikbaar is over oncologische patiënten en hun ziekte- en behandeltraject significant toegenomen. Alle bronnen van RWD hebben hun voor- en nadelen, en kunnen gebruikt worden om verschillende aspecten van oncologische geneesmiddelbehandelingen in de dagelijkse praktijk te evalueren. In **hoofdstuk 2** hebben we een niet-uitputtend overzicht van de meest gebruikte RWD-bronnen gegeven, inclusief traditionele bronnen als het *case report form*, een vragenlijst per patiënt, elektronische patiëntendossiers, administratieve declaratiegegevens, en de nieuwere bronnen als de patiënt-gerapporteerde uitkomsten, kankerregistraties, mobiele applicaties, draagbare technologie en sociale media. Data uit deze bronnen kan nu al significante inzichten genereren. Bijvoorbeeld om de *real-world* effectiviteit in te schatten, de farmacovigilantie te monitoren, de kosteneffectiviteit te berekenen, algemene geneesmiddelbehandelpatronen te bepalen en gepersonaliseerde geneesmiddelkeuzes te ondersteunen. Echter, om het volledige potentieel van deze RWD te benutten voor het evalueren van oncologische geneesmiddelen, moeten bronnen meer geharmoniseerd, compleet en ook met elkaar te combineren worden. Bovendien kan kunstmatige intelligentie behulpzaam zijn om extra inzichten te creëren, terwijl rekening gehouden blijft worden met de privacy van de individuele patiënt.

In **hoofdstuk 3** hebben we text-mining tool CTcue gevalideerd door data uit het EPD van het Leids Universitair Medisch Centrum te verzamelen van een cohort van 100 patiënten met gemetastaseerd niercelcarcinoom die in totaal 175 systemische anti-kankerbehandelingen kregen. Als eerste werden zowel met text mining als met handmatig dossieronderzoek patiënten geïdentificeerd en vervolgens de patiënt- en ziektekaracteristieken bij de start van de behandeling en de behandeluitkomsten verzameld. De vergelijking tussen de beide datasets liet zien dat met de text-mining tool patiënten adequaat geïdentificeerd en daarmee geïncludeerd konden worden. Ook konden de behandeluitkomsten, PFS en overleving (OS), bepaald worden zonder significant verschil in mediane overleving tussen beide methodes. Verder kon gestructureerde data verzameld worden met hoge accuraatheid. De informatie-extractie van ongestructureerde data, bijvoorbeeld van bijwerkingen en comorbiditeiten, liet resultaten zien die nogal verschilden betreffende de accuraatheid (F1-scores tussen 51.4–100). Hieruit concludeerden we dat sommige zoekopdrachten nog kunnen worden geoptimaliseerd. Als laatste liet deze studie zien dat de dataextractie met text mining zeven keer sneller was dan de handmatige dataverzameling: het kostte maar 12 minuten per patiënt, besteed aan dataverificatie, vergeleken met 86 minuten bij de handmatige dataverzameling. Deze bevindingen laten zien dat het gebruik van een

text-mining tool de verzameling van RWD uit het EPD significant efficiënter kan maken, ook omdat de opgestelde zoekopdrachten de potentie hebben om hergebruikt te worden voor de dataverzameling in andere ziekenhuizen en oncologische populaties.

Na validatie is de text-mining tool gebruikt om verschillende aspecten van behandelingen te evalueren.

In deel 2 wordt gefocust op de effectiviteit van behandelingen en behandelpatronen. Eerst is, in **hoofdstuk 4**, text mining gebruikt om de behandelingen te evalueren van patiënten met gemetastaseerd niercelcarcinoom in zowel het Leids Universitair Medisch Centrum als het Haga ziekenhuis. Tussen januari 2015 en mei 2020 zijn 122 patiënten gestart met een systemische anti-kankerbehandeling. De meeste van deze patiënten startten op pazopanib (n=70), sunitinib (n=34), en nivolumab met ipilimumab (n=28). 64% van de patiënten die een eerstelijnsbehandeling heeft afgerond tijdens de studieperiode ontving een tweedelijnsbehandeling, en daarvan startte weer 62% een derdelijnsbehandeling. Deze studie liet een grote variatie aan behandelpaden zien; cabozantinib en nivolumab waren de meest gebruikte behandelingen in tweede en derde lijn. Verder was het in deze studie voor de eerstelijnsbehandelingen pazopanib en sunitinib mogelijk om de mediane PFS (pazopanib: 16.3 maanden; sunitinib: 6.9 maanden) en mediane OS (pazopanib: 39.3 maanden; sunitinib: 28.1 maanden) te bepalen. Deze overlevingsuitkomsten lijken beter dan verwacht op basis van de registratietrials. De uitgebreide opvolging met tweede- en derdelijnsbehandelingen die deze patiënten ontvingen kan mogelijk aan deze betere uitkomsten hebben bijgedragen. Verder was het praktisch eenvoudig om de EPD text-mining methode toe te passen voor de evaluatie van geneesmiddelen in twee ziekenhuizen. Dit kan relevant zijn om het aantal ziekenhuizen voor dataverzameling uit te breiden in de toekomst, zeker omdat het veld van niercelcarcinoombehandelingen op dit moment snel ontwikkelt.

Vervolgens is een deel van de zoekopdrachten uit **hoofdstuk 4** gebruikt om de patronen en effectiviteit van de behandeling van patiënten met gemetastaseerd hepatocellulair carcinoom (HCC) te evalueren. Aangezien er voor HCC vele mogelijke prognostische factoren zijn, is in **hoofdstuk 5** voor een selectie van deze factoren onderzocht of deze de uitkomsten van de behandelingen beïnvloeden. De populatie van 94 patiënten die hun behandeling startte tussen 2015 en 2021 ontving als eerstelijnsbehandeling voornamelijk sorafenib (88% van de patiënten). Een select deel van de patiënten (16%) kreeg een tweedelijnsbehandeling, voornamelijk met regorafenib. Dit is niet geheel onverwacht, aangezien pas vanaf 2017 andere systemische behandelingen naast

sorafenib werden geregistreerd. Patiënten die een eerstelijnsbehandeling met sorafenib ontvingen hadden een mediane PFS van 4.1 maanden en een mediane OS van 10.5 maand. Echter, 28% van deze patiënten voldeed niet aan de in- en exclusiecriteria van de trials onderzocht in deze studie. De patiënten met een Child-Pugh score van B of C werden in deze studie geassocieerd met een relatief slechtere uitkomst, en deze patiënten hebben mogelijk geen enkel klinisch voordeel bij deze behandeling. Als laatste laat deze studie zien dat het hergebruiken van de zoekopdrachten in een andere populatie mogelijk is, en dat de verzameling van RWD voor de evaluatie van huidige, maar ook aankomende HCC-behandelingen, zowel de patronen als de uitkomsten, efficiënt uitgevoerd kan worden met deze text-mining methode.

In deel 3 is gefocust op RWD over de veiligheid van oncologische behandelingen. Ondanks dat bijwerkingen (AEs) gemonitord worden voor de toelating van een geneesmiddel tot de markt, zijn post-autorisatiestudies nodig om bijwerkingen in de praktijk te monitoren. Dit gaat bijvoorbeeld over bijwerkingen in patiënten die niet aan de in- en exclusiecriteria van trials zouden voldoen, maar ook over bijwerkingen die pas na een langere periode van behandelen optreden of zeldzame bijwerkingen.

In vergelijking met de palliatieve oncologische behandelingen is de acceptatie van bijwerkingen bij adjuvante behandelingen in de curatieve setting lager. Voor de behandeling van melanoom zijn er sinds 2018 drie adjuvante behandelingen geregistreerd. In **hoofdstuk 6** hebben we de tolerantie en veiligheid van deze drie adjuvante melanoombehandelingen bestudeerd. Om uit te zoeken of bijwerkingen de reden van het stoppen van de geneesmiddelbehandeling waren, is de text-mining tool gebruikt. Hiermee zijn de meest relevante stukken EPD aan het einde van de behandeling te selecteren, om zo de reden van stoppen te kunnen bepalen. Daarnaast zijn ook de meest voorkomende AEs uit de trials geïdentificeerd in de tekst om een inschatting te maken van de veiligheid van de geneesmiddelen. In deze studie zijn 122 patiënten geïnccludeerd; 54 kregen nivolumab, 48 kregen pembrolizumab en 20 patiënten werden behandeld met dabrafenib en trametinib. Patiënten die behandeld werden met dabrafenib en trametinib stopten significant vaker vanwege bijwerkingen dan patiënten behandeld met nivolumab en pembrolizumab ($p=0.024$).

De bijwerkingen van dabrafenib en trametinib die reden tot stoppen waren, waren voornamelijk combinaties van reversibele AEs, inclusief pyrexie en vermoeidheid. Bij de immuuncheckpointremmers waren het voornamelijk immuungerelateerde bijwerkingen. Wat betreft veiligheid laat deze studie vergelijkbare resultaten zien met

de registratietrials, met uitzondering van een hogere incidentie van AEs die reden tot stoppen waren voor zowel nivolumab als dabrafenib met trametinib, schildkliergerelateerde AEs voor zowel nivolumab als pembrolizumab, en een aantal milde AEs. Om een indicatie van de effectiviteit te krijgen, was de RFS op 1 jaar geschat, deze was voor alle behandelingen vergelijkbaar met de studieresultaten. In het algemeen laten de resultaten van deze studie zien dat een uitgebreide text-mining aanpak een waardevolle methode kan zijn om in de praktijk veiligheidsdata van kankerbehandelingen te verzamelen.

Sommige AEs zijn ernstig, maar kunnen (deels) voorkomen worden door een profylactische behandeling. Granulocyt kolonie-stimulerende factoren (G-CSFs) ter preventie van chemotherapie-geïnduceerde febrile neutropenie (FN) is een van deze behandelingen. Primaire preventie met G-CSFs is volgens de richtlijnen geïndiceerd voor patiënten met een hoog risico op FN. Door text mining van EPDs toe te passen waren we in staat om het juiste gebruik van G-CSFs in de dagelijkse praktijk te evalueren in een populatie van borstkankerpatiënten, zoals beschreven in **hoofdstuk 7**. In deze studie hebben we patiënten en behandelingen geïdentificeerd uit gestructureerde data en risicofactoren voor FN, G-CSF-gebruik en FN-incidentie uit ongestructureerde data. Van de 77 patiënten die een behandeling ontvingen met een hoog risico op FN startte 88% met een G-CSF behandeling. De onderhandeling resulteerde in een hogere incidentie van FN. De meeste van de 113 patiënten die behandeling ontvingen met een gemiddeld risico op FN hadden twee of meer risicofactoren. Daarom hadden deze patiënten mogelijk primaire preventie met een G-CSF moeten krijgen, echter dit werd bij weinig patiënten gedaan.

Daarom concludeerden we dat de toepassing van G-CSF in de dagelijkse praktijk niet helemaal in lijn is met de richtlijnen. Dit gold vooral voor patiënten die een behandeling met een gemiddeld risico ontvingen en daardoor onnodig werden blootgesteld aan toxiciteit. We doen een aanbeveling om de bewustwording betreffende de mogelijke risicofactoren voor neutropenie te vergroten en de risicofactoren duidelijker te definiëren in de richtlijnen. Tot slot laat deze studie zien dat text mining ook effectief kan worden geïmplementeerd om het volgen van richtlijnen in de dagelijkse praktijk te evalueren en zo patiëntenzorg te verbeteren.

Als laatste is de data-extractiemethode gebruikt voor een onderwerp net buiten de scope van deze thesis, namelijk voor een snelle inschatting van de ernst van de bijwerkingen van een behandeling van Covid-19. Slechte nier- en leverfunctie waren beide contra-indicaties voor de behandeling met remdesivir, echter werd dit in de praktijk alsnog gebruikt. Om een inschatting te maken van de invloed van remdesivir op de

nier- en leverfunctie, hebben we de veranderingen in nier- en leverfunctie na de start van de behandeling onderzocht in **hoofdstuk 8**. In deze studie zijn de nier- en leverfuncties van 103 Covid-19 patiënten in de eerste 15 dagen na start weergegeven. Twintig procent van de patiënten voldeed niet aan de trial in- en exclusiecriteria bij start op basis van de nier- en leverfunctie. Echter, ondanks dat een derde van de patiënten veranderingen in de functies liet zien, had een beperkte hoeveelheid patiënten ernstige verslechtering van hun nier- en leverwaarden. Onze observaties geven aan dat een matige tot slechte nier- en leverfunctie bij start geen absolute contra-indicatie zou moeten zijn voor het gebruik van remdesivir bij Covid-19, mits frequente monitoring van de functies tijdens de behandeling van deze patiënten. Echter was verificatie van deze bevindingen in een groter cohort nodig. Op dit moment is remdesivir niet meer geïndiceerd voor de behandeling van Covid-19.

In **hoofdstuk 9** zijn de methode en de bevindingen van de studies in deze thesis bediscussieerd en zijn de toekomstperspectieven geschetst. Deze thesis laat in het algemeen zien dat text mining een geschikte methode is voor het verzamelen van praktijkdata en het evalueren van de uitkomsten van oncologische geneesmiddelenbehandelingen, inclusief de behandelpatronen, effectiviteit, veiligheid, prognostische factoren en het volgen van richtlijnen. Ondanks dat het EPD een uitgebreide bron van RWD is voor de evaluatie van systemische anti-kankerbehandelingen, zijn er uitdagingen wat betreft het secundair gebruik van data, specifiek gecombineerd met text mining. Niet alleen de variabiliteit, maar ook omissies, inconsistenties en fouten in het EPD beïnvloeden de accuraatheid van de resultaten van het tekst-minen. Daarom is het noodzakelijk om de datakwaliteit in de EPDs verder te verbeteren. Ook kan het tekst-minen zelf worden verbeterd met bijvoorbeeld het introduceren van zoekopdrachten gebaseerd op *machine-learning* en *natural language processing* getraind op grote datasets. Aanvullend moeten er richtlijnen worden ontwikkeld over hoe EPD text mining moet worden toegepast, met gestandaardiseerde stappen als datavalidatie en afkapwaarden. Als laatste kan RWD actief gebruikt worden om inzichten te creëren en kan het deel uitmaken van de evaluatie van behandelingen wat betreft het behandelbeleid, de kosten en in de gedeelde besluitvorming.

Uiteindelijk moet de beschikbaarheid van RWD resulteren in betere inzichten in alle aspecten van kankerbehandelingen, met name welke behandeling het meest geschikt is voor een individuele patiënt in elke fase van de ziekte, en daarmee de systemische behandeling van kanker verbeteren.

Appendices

Curriculum Vitae

List of publications

Dankwoord

Curriculum Vitae

Sylvia van Laar was born in The Hague on November 7th 1993. She completed her secondary school (gymnasium) at Gymnasium Novum in Voorburg in 2012.

Next, she started her bachelor of Pharmacy at the University of Groningen. During a research internship, she investigated the association between drugs introducing oxidative stress and congenital disorders under supervision of prof. dr. B. Wilffert.

In 2015, she started her master's degree in Pharmacy at Utrecht University. During this period, she performed a research internship investigating risk factors for medication errors at admission at the Leiden University Medical Center under supervision of M.M. Ebbens and dr. K.B. Gombert-Handoko.

Afterward, in 2018, she started her PhD at the Leiden University Medical Center under the supervision of prof. dr. H.J. Guchelaar, dr. J. Zwaveling and dr. K.B. Gombert-Handoko. She investigated whether text mining was a suitable method for collecting real-world data from electronic health records to evaluate systemic anti-cancer therapies in clinical practice.

Since December 2022, she is working as policy advisor at the Ministry of Health, Welfare and Sports at the Information Policy Department.

List of publications

In this thesis

van Laar SA, Gombert-Handoko KB, Guchelaar HJ, Zwaveling J. An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma. *Clinical Pharmacology & Therapeutics*. 2020; 108: 644-652.

van Laar SA, de Boer MGJ, Gombert-Handoko KB, Guchelaar HJ, Zwaveling J. Liver and kidney function in patients with Covid-19 treated with remdesivir. *British Journal of Clinical Pharmacology*. 2021; 87(11): 4450-4454.

van Laar SA, Gombert-Handoko KB, Groenwold RHH, van der Hulle T, Visser LE, Houtsma D, Guchelaar HJ, Zwaveling J. Real-World Metastatic Renal Cell Carcinoma Treatment Patterns and Clinical Outcomes in The Netherlands. *Frontiers in Pharmacology*. 2022; 13: 803935.

van Laar SA, Gombert-Handoko KB, Wassenaar S, Kroep JR, Guchelaar HJ, Zwaveling J. Real-world evaluation of supportive care using an electronic health record text-mining tool: G-CSF use in breast cancer patients. *Supportive Care in Cancer*. 2022; 30: 9181-9189.

van Laar SA, Kapiteijn E, Gombert-Handoko KB, Guchelaar HJ, Zwaveling J. Application of electronic health record text mining: real-world tolerability, safety and efficacy of adjuvant melanoma treatments. *Cancers*. 2022; 14(21): 5426.

van Laar SA, Gombert-Handoko KB, Guchelaar HJ, Zwaveling J. Use of real-world data sources to evaluate cancer treatments: current practice and future perspectives. *Manuscript in preparation*.

van Laar SA, van der Weg W, Gombert-Handoko KB, van der Hulle T, Guchelaar HJ, Zwaveling J. Application of an electronic health record text-mining tool: treatment patterns and outcomes for Dutch patients with hepatocellular carcinoma. *Manuscript in preparation*.

Not in this thesis

Ebbens MM, **van Laar SA**, Wesselink EJ, Gombert-Handoko KB, van den Bemt PMLA. Prospective validation of a risk prediction model to identify high-risk patients for

medication errors at hospital admission. *Annals of Pharmacotherapy*. 2018; 52(12): 1211-1217.

Abedian Kalkhoran H, Zwaveling J, Storm BN, **van Laar SA**, Portielje JEA, Codrington H, Luijten D, Brocken P, Smit EF, Visser LE. A text-mining approach to study the real-world effectiveness and potentially fatal immune-related adverse events of PD-1 and PD-L1 inhibitors in older patients with stage III/IV non-small cell lung cancer. *BMC Cancer*. 2023; 23: 247.

Dankwoord

Op de weg naar een proefschrift, word je door velen ondersteund, geïnspireerd, geadviseerd. Hen wil ik hier graag voor bedanken.

Allereerst dank aan mijn (co-)promotoren. Henk-Jan, bedankt voor je vertrouwen in dit project, scherpe blik en het altijd toe kunnen voegen van extra perspectieven. Beste Juliëtte en Kim, dank voor jullie enorme betrokkenheid en alle ideeën, suggesties, maar ook zeker het zijn van de stok achter de deur in dit traject wat we al pionierend door zijn gegaan. Ik zie al onze momenten, van kantoor, maar ook zeker toen we over moesten naar online, als zeer waardevol.

Dank aan Jochem de Boer, Roel Lakmaker, en alle andere medewerkers van CTCue voor het technisch mogelijk maken van dit onderzoek.

Mijn medepromovendi en kamergenoten: Carin, Cathelijne, Maaïke, Michel, Anyue, Lianne, Marieke, Emma, Sylvia K., Rineke, Iris, Thomas, Claire, Qinglian, Henok en Hanieh, dank voor alle goede discussies en de gezelligheid. De variatie in onze onderzoeken heeft mij ook extra verdieping en verrijking van kennis gebracht. Maaïke, dank voor het zijn van mijn steun en toeverlaat bij alles gerelateerd en minder gerelateerd aan het doen van onderzoek bij KFT. Of we nu een bureau, een Teams-chat of een chalet in de Veluwe deelden, onze samenwerking maakte mijn werkdagen altijd beter. Daarom ben ik blij dat je ook vandaag dicht bij me zal staan.

Dank aan de collega's bij de cytostatica voor de gezelligheid op maandagen die ik met jullie heb mogen doorbrengen. Onderzoek doen met data uit het patiëntendossier kan pas goed als je ook weet hoe die praktijk werkt.

Dank aan de bachelor- en masterstagiaires Anouk, Amber, Doaa, Emma en Willem voor jullie bijdragen.

Merian, wat ben ik dankbaar voor de traditie die wij hebben opgebouwd, nadat we elkaar hebben leren kennen bij de cursus time management. Alle gesprekken met jou als m'n Grip buddy hebben mij heel erg geholpen m'n doelen helder te krijgen en te behalen.

Lieve Eva en Vera, wat was het fijn om jullie in de buurt te hebben. Ik ben dankbaar voor al die avondjes met heerlijk eten en goede gesprekken, ook als het weer even druk was. Vera, jij weet als geen ander dat een PhD doen, ongeacht de faculteit, niet vanzelf gaat.

Niet alle inspiratie, motivatie en afleiding komt van dichtbij. Dank aan alle Volters uit Leiden, Nederland en Europa, (kandidaat) Tweede Kamerleden, werkgroep bestuur wetenschappelijk bureau en met name het CO-team, Julia, Mette, Jori, Robin, Mels, Marc, Maurice en Diederik voor alles wat ik samen met jullie heb mogen aangaan, meemaken en vieren afgelopen jaren en mijzelf zo heb kunnen ontwikkelen.

Dank aan mijn lieve jaarclubgenoten, Babette, Carlijn, Emmy, Jikke, Pascalle en Sophie. Een deel van jullie heb ik beloofd hier te bedanken voor het meelesen met m'n sollicitatiebrief vijf jaar geleden, het is gelukt! Ik ben eindeloos dankbaar voor onze vriendschap die door de jaren heen is gegroeid, veranderd en ons al op veel plekken heeft gebracht, en tegelijkertijd nog steeds zo dichtbij voelt als waar het begon, in Groningen. Gelukkig. Babette, met jou op reis is altijd een succes, blij dat ik ook vandaag op je kan rekenen.

Als laatst dank aan mijn familie. Mama, dank voor het echt altijd voor me klaar staan. Papa voor het zijn van de supporter langs de lijn. Lieve Linda en Maurits, wat blij ben ik met jullie als zus en broer. Tijdens mijn promotietraject brachten onze wegen ons geografisch weer dicht bij elkaar. Ik vind het inspirerend hoe jullie creatief, ondernemend, (maatschappelijk) bewust en vooral geheel eigen door het leven gaan, dank daarvoor.

