30-day mortality of patients with advanced cancer

monitoring and machine learning models using extensive health data

Vesteghem, Charles Philippe Edouard

DOI (link to publication from Publisher):
10.54337/aau468602374

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

30-DAY MORTALITY OF PATIENTS WITH ADVANCED CANCER: MONITORING AND MACHINE LEARNING MODELS USING EXTENSIVE HEALTH DATA

BY
CHARLES VESTEGHEM

DISSERTATION SUBMITTED 2022

AALBORG UNIVERSITY
DENMARK

30-day mortality of patients with advanced cancer: monitoring and machine learning models using extensive health data

PhD Dissertation

Charles Vesteghem

Dissertation submitted January 2022

Dissertation submitted: January 3, 2022

Main PhD supervisor: Professor Martin Bøgsted
Department of Clinical Medicine
Aalborg University, Denmark
Department of Haematology
Aalborg University Hospital, Denmark

PhD co-supervisors: Clinical Professor Ursula Falkmer
Department of Clinical Medicine
Aalborg University, Denmark
Department of Oncology
Aalborg University Hospital, Denmark

Associate Professor Rasmus Froberg Brøndum
Department of Clinical Medicine
Aalborg University, Denmark
Department of Haematology
Aalborg University Hospital, Denmark

Professor Karen Dybkær
Department of Clinical Medicine
Aalborg University, Denmark
Department of Haematology
Aalborg University Hospital, Denmark

PhD committee: Associate Professor Henrik Bøggild (chair)
Aalborg University, Denmark

Associate Professor Deirdre Cronin Fenton
Aarhus University, Denmark

Professor Anita Burgun
Paris University, France

PhD Series: Faculty of Medicine, Aalborg University
Department: Department of Clinical Medicine

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-955-4

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2022

Preface

This thesis is intended for biostatisticians, health data scientists, and clinicians with a special interest in biostatistics and machine learning. This work was conducted at the Department of Clinical Medicine, Aalborg University, and the Department of Haematology, Aalborg University Hospital, between January 2019 and December 2021. It was affected, as nearly all activities across the world, by the SARS-CoV-2 pandemic that led to multiple lockdowns in 2020. This hindered the possibility of close collaboration and networking. However, this work was based on registry data that were readily available, which allowed me to conduct this work with limited disruption during this period, even allowing me to focus more on my work. This was also a good occasion to reflect on our busy lives.

On a professional level, after 10 years in start-ups, I was close to innovation but far from science. Joining the Department of Haematology and starting this thesis was a great opportunity for me to dive into the academic world. A PhD is an often-solitary endeavour, but it is a great learning experience in terms of methods and mindset and brings a large panel of transferable skills.

First and foremost, this work would not have been possible without Martin Bøgsted, who together with the late Hans E. Johnsen initially trusted me to be part of the research unit and later worked with Ursula G. Falkmer to define the scope of this PhD. His support and feedback have been invaluable throughout this journey.

This work allowed me to dive into complex topics within health data science, notably machine techniques applied to health registry data. Alongside Martin Bøgsted, Rasmus F. Brøndum was a great support for the biostatistics and machine learning aspects, as well as a friendly companion.

This work also gave me the occasion to better understand cancer and its intrinsic complexity. This understanding was made possible thanks to the precious input from Karen Dybkær based on her expertise in molecular biology and from Ursula G. Falkmer based on unique clinical experience.

I would also like to thank all of my colleagues from the Department of Haematology and Department of Oncology for their support and help, as well as Chloé-Agathe Azencott and the members of her discussion group for our online collaboration.

Finally, such a project also requires the support of one's family; I would like to thank them from the bottom of my heart, especially my wife, Trine, for her love, support, and patience and my father for showing the way. I dedicate this work to my children, Sofia, Oscar, and Victor, as an ode to curiosity.

Charles Vesteghem

Thesis outline

Thesis title:	30-day mortality of patients with advanced cancer: monitoring and machine learning models using extensive health data
PhD student:	Charles Vesteghem
Main supervisor:	Professor Martin Bøgsted, Aalborg University

This thesis is composed of an introduction and three scientific papers:

1. Paper I: Thirty-day mortality following systemic anticancer therapy: Evaluating risk factors without selection bias in a real-world, population-based cohort from 2009 to 2019. Vesteghem C, Brøndum RF, Mouritzen MT, Christensen HS, Bøgsted M, Falkmer UG, Poulsen LØ. Clinical Oncology, submitted 2021 Sep.
2. Paper II: High validity of the Danish National Patient Registry for systemic anticancer treatment registration from 2009 to 2019. Vesteghem C, Brøndum RF, Falkmer UG, Pottegård A, Poulsen LØ, Bøgsted M. Clinical Epidemiology, Accepted 2021 Nov.
3. Paper III: Dynamic risk prediction of 30-day mortality of patients with advanced lung cancer: Comparing five machine learning approaches. Vesteghem C, Weronika MS, Brøndum RF, Falkmer UG, Azencott CA, Bøgsted M. Submitted 2021 Nov.

Five additional papers have been published during this PhD study but are not part of this thesis:

1. Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives. Vesteghem C, Brøndum RF, Sønderkær M, Sommer M, Schmitz A, Bødker JS, Dybkær K, El-Galaly TC, Bøgsted M. Brief Bioinform. 2020 May 21;21(3):936-945. doi: 10.1093/bib/bbz044.
2. Development of a Precision Medicine Workflow in Hematological Cancers, Aalborg University Hospital, Denmark. Bødker JS, Sønderkær M, Vesteghem C, Schmitz A, Brøndum RF, Sommer M, Rytter AS, Nielsen MM, Madsen J, Jensen P, Pedersen IS, Grubach L, Severinsen MT, Roug AS, El-Galaly TC, Dybkær K, Bøgsted M. Cancers (Basel). 2020 Jan 29;12(2):312. doi: 10.3390/cancers12020312.
3. Nationwide Survival Benefit after Implementation of First-Line Immunotherapy for Patients with Advanced NSCLC-Real World Efficacy. Mouritzen MT, Carus A, Ladekarl M, Meldgaard P, Nielsen AWM, Livbjerg A, Larsen JW, Skuladottir H, Kristiansen C, Wedervang K, Schytte T, Hansen KH, Østby AC, Frank MS, Lauritsen J, Sørensen JB, Langer SW, Persson GF, Andersen JL, Frary JMC, Drivsholm LB, Vesteghem C, Christensen HS, Bjørnhart B, Pøhl M. Cancers (Basel). 2021 Sep 28;13(19):4846. doi: 10.3390/cancers13194846.

4. Direct costs of antineoplastic and supportive treatment for progressive multiple myeloma in a tax-based health system. Brøndum RF, Vestergaard AS, Børty L, Vesteghem C, Rytter AS, Nielsen MM, Severinsen MT, Jensen P, Gregersen H, El-Galaly TC, Dybkær K, Ehlers LH, Bøgsted M, Roug AS. Future Oncol. 2021 Sep;17(25):3331-3341. doi: 10.2217/fon-2021-0189. Epub 2021 Jun 22.
5. Patient-reported outcomes in patients with hematological relapse or progressive disease: a longitudinal observational study. Sommer M, Nielsen LK, Nielsen LB, Brøndum RF, Nielsen MM, Rytter AS, Vesteghem C, Severinsen MT, El-Galaly TC, Bøgsted M, Grønkjær M, Jørgensen L. Health Qual Life Outcomes. 2021 Nov 4;19(1):251. doi: 10.1186/s12955-021-01887-6.

English abstract

Systemic anticancer therapies (SACTs) often have severe short- and long-term side effects. Some patients with limited remaining survival time will only experience the short-term side effects with no clinical benefit. Therefore, SACTs should be avoided in these cases to limit the negative impact on their health-related quality of life near death. In Denmark, there are a variety of digital and national health registries that can be coupled and leveraged to monitor the last SACT administrations, including the Danish National Patient Registry (DNPR), which contains administrative data from all hospitals in Denmark (e.g., diagnosis and procedure codes), and clinical databases, such as the histopathological and laboratory registries.

To quantify the frequency of late SACT administration, Earle et al. proposed two widely used indicators that consider the patients who died from cancer. However, these indicators are only applicable in hindsight, as they require information about the cause of death, which limits their clinical applicability. Another approach proposed by Wallington et al. looks at the 30-day mortality following SACT, but it is also conditioned on future events, as it only considers the last SACT within a predefined observation period. In Paper I, we proposed an adapted version based of Wallington et al. that avoids conditioning on future events. This approach allowed us to calculate risk factors more reliably and avoided the potential sampling bias that arises from conditioning on future events. We analysed the data from more than 10,000 cancer patients treated at the Department of Oncology, Aalborg University Hospital, during 2009-2019. We reported differences between malignancies and treatment intent, as well as a downward trend in the frequency of the 30-day mortality following SACT during the study period.

To facilitate implementation of such a monitoring tool on a national level, the indicator must be easily and reliably calculated using available health data. One of the main data sources used in Paper I was the prescription database MedOnc used at the Department of Oncology, Aalborg University Hospital. This database contains information on drug administration to cancer patients, but it is not available nationally. However, the DNPR is available nationally and contains similar information, though, as it contains administrative data, its clinical validity could be questioned. To confirm the validity of this registry for SACTs, in Paper II, we conducted a validation study comparing the DNPR to MedOnc, confirming DNPR's high validity and paving the way for the implementation of a national indicator.

A strategy to limit the frequency of late SACT administration is to help clinicians better assess the risk of early mortality among cancer patients. In Paper III, we compared various machine learning techniques to dynamically predict the 30-day mortality of patients with advanced cancer. In line with other studies, tree-based models outperformed simpler or neural network-based models, and our results showed that most of the information for accurate prediction lies in the biochemical results.

In this thesis, our aim was to tackle late SACT administration in cancer patients through better monitoring and machine learning approaches using extensive health data. We have proposed beneficial improvements that could be implemented nationally in tools for

monitoring and the dynamic prediction of 30-day mortality after these results are confirmed in prospective studies.

Dansk résumé

Systemisk anticancerbehandling (SACT) har ofte alvorlige kort- og langtidsbivirkninger. Nogle patienter med begrænset resterende overlevelsestid vil derfor kun opleve bivirkninger uden at opnå en klinisk fordel. SACT bør derfor undgås i disse tilfælde for at begrænse negativ påvirkning af den sundhedsrelaterede livskvalitet tæt på livets afslutning. Der findes i Danmark en lang række nationale sundhedsregistre, som kan kobles sammen og udnyttes til at danne sig et overblik over SACT-administrationer, som er givet tæt på livets afslutning. Disse omfatter Landspatientregisteret (LPR), som indeholder administrative data fra alle sygehuse i Danmark såsom diagnose- og procedurekoder og kliniske databaser såsom patologi- og laboratoriedatabasen.

For at kvantificere hyppigheden af sene SACT-administrationer har Earle et al. foreslået to populære kvalitetsindikatorer, som betragter patienter, der er døde af kræft. Disse indikatorer kan dog kun bruges retrospektivt, da de kræver information om dødsårsagen, hvilket begrænser deres kliniske anvendelighed. En anden tilgang, foreslået af Wallington et al., kigger på 30-dages dødeligheden efter SACT. Men denne indikator betinger også på fremtidige begivenheder, da den kun betragter den sidste SACT inden for et foruddefineret observationsvindue. I Artikel I foreslår vi en tilpasset version af Wallington et al., der undgår at betinge på fremtidige begivenheder. Denne tilgang giver os mulighed for at beregne risikofaktorer mere pålideligt og undgå den potentielle sampling bias, der opstår som følge af at betinge på fremtidige begivenheder. Vi har analyseret data fra mere end 10.000 kræftpatienter behandlet på Aalborg Universitetshospital i perioden 2009-2019. Vi fandt forskelle i administrationen af SACT tæt på livets afslutning mellem kræfttyper og behandlingsintentioner samt en nedadgående tendens hen over perioden.

For at muliggøre implementeringen af en indikator på nationalt plan skal indikatoren kunne beregnes let og pålideligt ved hjælp af tilgængelige sundhedsdata. En af de vigtigste datakilder, der blev brugt i Artikel I, var ordinationsdatabasen MedOnc, der bliver brugt på Onkologisk Afdeling, Aalborg Universitetshospital. Denne database indeholder information om SACT-administration til kræftpatienter, men denne database er ikke tilgængelig nationalt. LPR er dog tilgængeligt nationalt og indeholder lignende oplysninger, men da LPR indeholder administrative data, kan man overveje LPR's validitet. For at validere dette register mht. SACT har vi i Artikel II gennemført et valideringsstudie, der sammenligner LPR med MedOnc. Studiet bekræfter LPR's høje validitet og baner derved vejen for implementeringen af en national indikator.

En strategi til at begrænse hyppigheden af sen SACT-administration er at hjælpe klinikerne til bedre at vurdere risikoen for tidlig død efter SACT. I Artikel III sammenligner vi forskellige maskinlæringsteknikker til dynamisk forudsigelse af 30-dages dødeligheden for patienter med fremskreden cancer. Vores resultater viser, på linje med andre undersøgelser, at træbaserede modeller udkonkurrerer mere simple eller neurale netværk-baserede modeller og det meste af den prædiktive værdi ligger i de biokemiske resultater.

I denne afhandling var det målet at kunne foreslå en forbedring af håndteringen af sen SACT-administration til kræftpatienter gennem bedre overvågnings- og

maskinlæringstilgange baseret på omfattende sundhedsdata. Vi mener at have foreslået gavnlige redskaber, der potentielt kan implementeres nationalt i værktøjer til overvågning og dynamisk forudsigelse af 30-dages dødeligheden efter SACT, efter disse resultater er blevet bekræftet i prospektive studier.

Résumé en français

Les thérapies anticancéreuses systémiques (SACTs) sont connues pour avoir des effets secondaires souvent graves à court et long terme. Certains patients en fin de vie sont victimes de ces effets secondaires sans bénéfice clinique. Les SACTs doivent donc être évitées dans ces cas pour limiter leur impact négatif sur la qualité de vie liée à la santé de ces patients. Il existe au Danemark un grand nombre de registres nationaux contenant des données de santé numérisées qui peuvent être couplés, notamment pour analyser l'usage de SACT en fin de vie. Cela inclut le Registre National Danois des Patients (DNPR) qui contient des données administratives de tous les hôpitaux danois, telles que les codes pour les diagnostics et procédures et des bases de données cliniques contenant par exemple des résultats d'analyses anatomopathologiques ou biochimiques.

Pour quantifier la fréquence d'administration SACT en fin de vie, Earle et al. ont proposé deux indicateurs fréquemment utilisés qui prennent en compte uniquement les patients ayant décédés par cancer. Ces indicateurs ne sont cependant applicables que de façon rétrospective car ils nécessitent des informations sur la cause du décès, ce qui limite leur applicabilité clinique. Une autre approche proposée par Wallington et al. examine la mortalité à 30 jours après une SACT. Mais cette approche nécessite aussi de conditionner sur des événements futurs, car il ne considère que la dernière SACT dans une fenêtre d'observation prédéfinie. Dans l'article I, nous proposons une version adaptée de Wallington et al. qui évite ce type de conditionnement. Cette approche nous permet de calculer les facteurs de risque de manière plus fiable et évite le biais d'échantillonnage potentiel du fait du conditionnement sur des événements futurs. Nous avons analysé les données de plus de 10 000 patients atteints de cancer traités au Département d'Oncologie de l'Hôpital Universitaire d'Aalborg au cours de la période 2009-2019. Nous avons identifié des différences entre types de cancer et selon l'intention du traitement ainsi qu'une tendance à la baisse au cours de la période d'étude.

Pour faciliter la mise en place d'un outil de suivi au niveau national utilisant cette approche, l'indicateur correspondant doit pouvoir être calculé de manière simple et fiable à partir des données de santé facilement accessibles. L'une des principales sources de données utilisées dans l'article I était la base de données de prescription MedOnc utilisée au Département d'Oncologie de l'Hôpital Universitaire d'Aalborg. Cette base de données contient des informations sur l'usage de traitements anticancéreux, mais elle n'est pas disponible au niveau national. En revanche, le DNPR est disponible au niveau national et contient des informations similaires. Cependant, s'agissant d'un outil administratif, sa validité clinique pourrait être remise en question. Pour confirmer la validité de ce registre pour les SACTs, nous avons mené dans l'article II une étude de validation comparant le DNPR à MedOnc, confirmant la haute fiabilité du DNPR, ouvrant ainsi la voie à la mise en œuvre d'un outil de suivi national.

Une stratégie pour limiter la fréquence d'administration de SACT en fin de vie est d'aider les médecins à mieux évaluer le risque de mortalité précoce des patients cancéreux. Dans l'article III, nous comparons diverses techniques d'apprentissage automatique pour prédire de manière dynamique la mortalité à 30 jours des patients atteints d'un cancer avancé. Nos résultats montrent, conformément à d'autres études, que les modèles basés

sur des arbres de décision surpassent les modèles plus simples ou basés sur des réseaux neuronaux et que la plupart des informations nécessaire à une bonne prédiction résident dans les résultats d'analyses biochimiques.
Dans cette thèse, notre objectif était d'améliorer la gestion des SACTs chez les patients cancéreux en fin de vie grâce à un meilleur suivi et à des méthodes d'apprentissage automatique, en utilisant une large variété de données de santé. Nous avons proposé des solutions qui pourraient être mises en œuvre au niveau national dans des outils de suivi de la mortalité à 30 jours et d'aide à la décision, une fois ces résultats confirmés dans des études prospectives.

Table of contents

Table of contents

Introduction

1. Background

1.1. Cancer

Cancer is currently one of the leading causes of death worldwide[1], and the most common cause of death in people below 65 years of age in Europe[2]. Cancer includes a wide variety of disease types with uncontrolled multiplication and dissemination of neoplastic cells, creating tumours. Death is typically due to tumours developing in other organs, called metastases, which can impair the functioning of these organs. The organs affected by metastases are generally the liver, lungs, brain, and skeleton. Cancer is a genetic disease in the sense that it is primarily caused by alterations in the genetic material of the cells.
The DNA contained in the nucleus of the cell is read by an enzyme, RNA polymerase, which generates messenger RNA from regions of the DNA marked by specific sequences of nucleotides. These regions are called genes (see Figure 1). The generated messenger RNA is used by the ribosome as instructions to produce a protein. This process was described in the Central Dogma of Molecular Biology by Francis Crick[3]. Once generated, the protein folds into a specific form based on the amino acids composing this protein. Once folded, the protein can play a role in transport/storage or as a structural component, messenger, enzyme, or antibody.

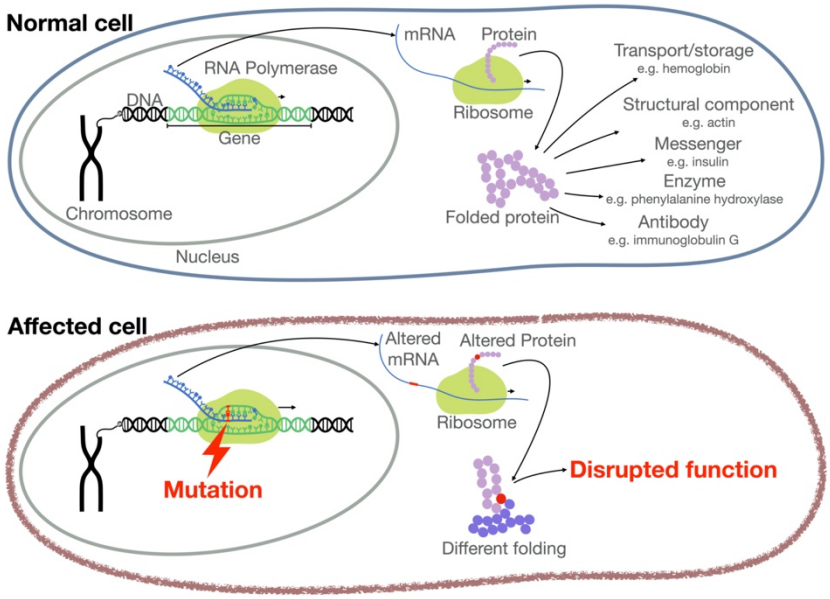


Figure 1. *Protein generation from DNA and impact of a mutation on the folding and, therefore, function of the protein.*

A mutation in a gene can be passed along to the messenger RNA. This altered RNA can be used to produce a protein in which the composition and shape, and therefore function, is disrupted. The original protein could have been part of some control mechanism that is then affected by the mutation. Fortunately, many control mechanisms are present in the cell life cycle to avoid uncontrolled growth.
To circumvent these mechanisms and become malignant, a cell must progressively incorporate changes in its DNA, typically induced by carcinogens. There are three types of carcinogens: physical, chemical, and biological. Physical carcinogens are ionising radiation, such as high-energy UV, which can lead to malignant melanoma. Benzene or arsenic are common examples of chemical carcinogens. Benzene, for example, increases the risk of developing leukaemia. Biological carcinogens are viruses, bacteria, and parasites, such as the human papillomavirus that leads to cervical cancer or the bacteria *Helicobacter pylori* that leads to stomach cancer. Most of these induced genetic alterations are corrected in normal cells, but, in rare occasions, slip through, especially in individuals with genetic predispositions to deficient DNA damage repair mechanisms. A natural selection process is initiated in which some mutations provide a competitive advantage for affected cells, such as by allowing easier and faster division through the disruption of control mechanisms. The required phenotypes of changes for a cell to become malignant were summarised in 10 hallmarks by Hanahan and Weinberg[4] (see Figure 2).

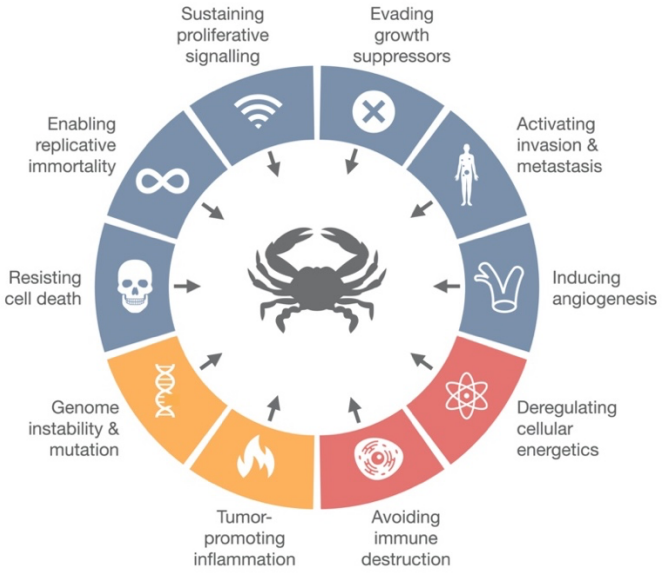


Figure 2. *Updated hallmarks of cancer. The blue hallmarks are the major hallmarks as introduced in the first version of their categorisation. The red hallmarks are two emerging hallmarks, and the orange hallmarks are enabling traits. Based on Hanahan and Weinberg*[5].

Major progress has been made in the last few decades in the treatment of malignancies, with drastic improvement in prognosis for certain diseases, including breast cancer.

However, even in Western countries, most patients will relapse and die from their cancer. Notably, patients with lung or pancreatic cancer have 5-year survival of less than 30%[6].

1.2. Systemic anticancer therapies

Systemic anticancer therapies (SACTs) are treatments for malignancies that have a systemic effect on the patients (i.e., they affect the whole body) as opposed to local anticancer therapies, such as radiotherapy and surgery. SACTs have been used to treat patients for more than 50 years, with the first positive results in the 1940s for haematological malignancies[7]. For solid tumours, these therapies were not considered by most clinicians treating cancer until the 1960s, when they improved the remission rate in combination with local anticancer therapies. This was notably thanks to fluorouracil, which is still currently one of the most widely used drugs (see Table 1).

SACTs rely on drugs that can be administered orally or injected into the patients. Oral drugs must be absorbed first via the digestive system before they can reach the bloodstream. They are also often metabolised in the liver. Once in the bloodstream, the drugs reach all cells in the body, affecting both normal and cancer cells. Notably, due to the brain-blood barrier, some drugs based on larger molecules cannot reach the intracerebral space, complicating the treatment of brain cancers.

As normal cells are also exposed to the drugs used in SACTs, these drugs, depending on type and dose, can have severe short-term and long-term side effects, affecting the health-related quality of life (HRQoL) of patients.

1.2.1. ATC classification

Many different types of drugs for SACT have been developed and manufactured by various pharmaceutical companies under different commercial names. To facilitate the unique identification of drugs, an international classification was put in place by the WHO: the Anatomical Therapeutic Chemical Classification System[8] (ATC). This system assigns a code to the active compound of a drug based primarily on its indication. SACT drugs can primarily be found in the antineoplastic agent category, coded L01. However, endocrine therapy (L02) is also used to treat specific cancer types, notably hormone-sensitive breast cancer. Supportive drugs used to limit the side effects of the SACT drugs are often included in the SACT regimen, such as immunostimulants (L03), antiemetics, and antinauseants (A04). Most frequently used drugs in the "Antineoplastic and immunomodulating agents" category (ATC code L) at the Department of Oncology, Aalborg University Hospital, between 2008 and 2019 are presented in Table 1.

Table 1. *Top 30 "L" drugs used at the Department of Oncology, Aalborg University Hospital, between 2008 and 2019.*

Name	ATC	Records*	Main indication(s)	Type
Fluorouracil	L01BC02	24 583	Colorectal, Pancreatic, Breast	Cytotoxic
Paclitaxel	L01CD01	19 499	Breast, Ovarian	Cytotoxic
Vinorelbine	L01CA04	19 348	Lung, Breast	Cytotoxic
Carboplatin	L01XA02	17 228	Lung, Ovarian	Cytotoxic
Trastuzumab	L01XC03	16 025	Breast, Gastro-oesophageal	Immunotherapy, active
Irinotecan	L01CE02	15 866	Colorectal, Brain	Cytotoxic
Bevacizumab	L01XC07	14 504	Colorectal, Ovarian	Cytotoxic
Oxaliplatin	L01XA03	13 899	Colorectal, Gastro-oesophageal	Cytotoxic
Capecitabine	L01BC06	12 163	Colorectal, Breast	Cytotoxic
Docetaxel	L01CD02	10 868	Breast, Prostatic	Cytotoxic
Gemcitabine	L01BC05	10 211	Pancreatic, Urinary	Cytotoxic
Epirubicin	L01DB03	8 968	Breast, Gastro-oesophageal	Cytotoxic
Cyclophosphamide	L01AA01	6 924	Breast	Cytotoxic
Pegfilgrastim	L03AA13	5 278	All	Supportive (neutropenia)
Cetuximab	L01XC06	5 013	Colorectal	Immunotherapy, active
Etoposide	L01CB01	4 848	Lung	Cytotoxic
Cisplatin	L01XA01	4 233	Lung	Cytotoxic
Fulvestrant	L02BA03	3 173	Breast	Cytotoxic, targeted
Letrozole	L02BG04	2 962	Breast	Cytotoxic, targeted
Lipegfilgrastim	L03AA14	2 876	All	Supportive (neutropenia)
Temozolomide	L01AX03	2 622	Brain	Cytotoxic
Pemetrexed	L01BA04	2 284	Lung	Cytotoxic
Pembrolizumab	L01XC18	1 681	Lung, Urinary	Immunotherapy, passive
Nivolumab	L01XC17	1 625	Lung	Immunotherapy, passive
Palbociclib	L01EF01	1 591	Breast	Cytotoxic, targeted
Pertuzumab	L01XC13	1 560	Breast	Cytotoxic, targeted
Eribulin	L01XX41	1 475	Breast	Cytotoxic
Panitumumab	L01XC08	1 364	Colorectal	Cytotoxic, targeted
Erlotinib	L01EB02	1 354	Lung	Cytotoxic, targeted
Doxorubicin	L01DB01	1 324	Ovarian, Endometrial	Cytotoxic

** The number of prescriptions of the corresponding drug as recorded in the prescription software MedOnc (see Datasets section).*

1.2.2. Cytotoxic drugs

1.2.2.1. Non-targeted drugs

Different classes of drugs can be found among the antineoplastic agents (L01). Historically, the first SACTs to be introduced relied on drugs that were toxic to cells and, thus, are referred to as cytotoxic[7]. These drugs were initially non-targeted, affecting all growing cells in the body. They typically have a cytotoxic effect by blocking the cell cycle at different phases (see Figure 3).

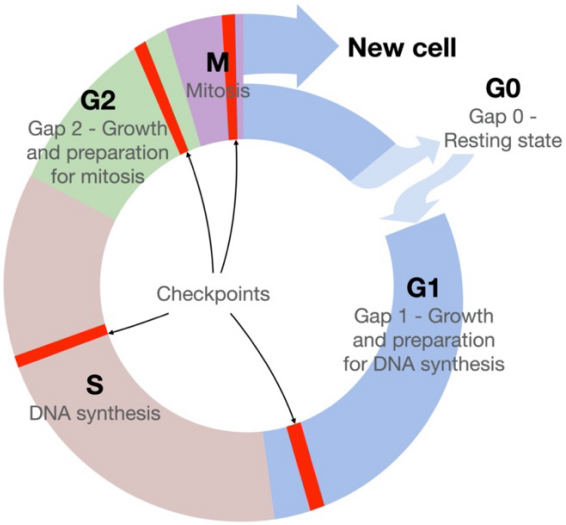


Figure 3. *The cell cycle. During the multiplication process, the cell goes through four stages: G1, S, G2, and M. The cycle has two main outcomes: copying of the DNA (S) and cell division (M). Specific control mechanisms are activated at each checkpoint represented by a red line on the figure.*

The cell cycle is composed of four active phases and a resting phase (see Figure 2)[9]. The resting phase is called Gap 0 (G0). In this quiescent state, the cell cycle does not progress towards cell division. Instead, the cell waits for a stimulus to be reactivated, such as in the case of tissue-specific stem cells when an injury occurs or differentiates into a structural component (e.g., myoblasts). The Gap 1 (G1) phase corresponds to the growth phase when the cell prepares for DNA synthesis. Synthesis (S) is the phase in which DNA is replicated. The Gap 2 (G2) phase is a phase of cell growth after DNA synthesis when the cell prepares for mitosis. Mitosis is the phase in which the cell divides into two cells, which both then enter the G1 phase, continuing the cycle. Checkpoints are milestones in the cell cycle when some control mechanisms are activated to ensure the cell is ready to continue the cycle. For example, the G2/M checkpoint checks for errors in the DNA to prevent potential errors from being passed to daughter cells. It stops the cell cycle as long as the detected errors are not repaired[10].

Blockade of the cell cycle is typically achieved by preventing copying of the DNA (e.g., fluorouracil, gemcitabine, carboplatin), preventing cell division (e.g., paclitaxel,

vinorelbine, eribulin), or damaging the DNA with double-stranded DNA breaks, leading to cell death (e.g., irinotecan, topotecan)[11–13]. For example, fluorouracil inhibits the synthesis of a nucleotide, the pyrimidine thymidylate, which is needed for DNA replication, whereas paclitaxel prevents the destruction of microtubules, structures that are needed for mitosis but must be destroyed for mitosis to finish.

Blocking the cell cycle has a major impact on the fast-dividing tumour cells. Unfortunately, many types of normal cells, such as epithelial cells in the gastrointestinal tract or blood stem cells, are also fast dividing. Therefore, they are also impacted, leading to potentially severe side effects, such as vomiting, diarrhoea, rashes, or acute peripheral neuropathy, as well as life-threatening conditions. As such, a major concern during treatment is infections due to neutropenia. In addition, cancer treatment drugs often induce long-term complications, including cardiac toxicity and an increased risk of secondary cancers.

1.2.2.2. Targeted drugs

To circumvent the drawbacks from non-targeted drugs, great effort has been made to develop drugs that more specifically affect cancerous cells. These drugs affect a specific hallmark of the cancer, stopping its growth or reducing the tumour size. These types of treatments have less of an impact on normal cells but are only usable in specific types of cancer depending on the targeted hallmark. They are typically protein kinase inhibitors (L01EF); protein kinases are enzymes that alter proteins and, thus, their function. Preventing this alteration can suppress the targeted hallmark, forcing cancer cells into a less malignant state. Examples of such drugs are palbociclib and erlotinib, which are primarily used to treat breast and lung cancer, respectively[14,15]. Nevertheless, they still have an effect on different normal cells that rely on these kinases for growth and, thus, also generate side effects. Alongside classical mild chemotherapy-related side effects (nausea, diarrhoea, headache…), they are often associated with more severe side effects, such as neutropenia and stroke.

Another more recent strategy for targeted treatment is to use antibody-drug conjugates, such as trastuzumab emtansine (L01XC14). This type of treatment usually combines an antibody, trastuzumab in the mentioned example, with a cytotoxic compound. The role of the antibody is to preferentially bind a receptor on the cancer cells, triggering the internalisation of the cytotoxic compound and leading to cell death. Typical mild and severe chemotherapy-induced side effects are still present[16]. Notably, the antibody is also often used alone in active immunotherapy (see below).

Endocrine therapies could also be considered as targeted cytotoxic drugs because they impact the growth and multiplication of cells in a targeted manner, but they typically have fewer as well as less severe side effects.

1.2.3. Immunotherapy

Immunotherapy relies on an indirect mechanism to treat cancer. As opposed to cytotoxic therapy, which directly hinders the growth and multiplication of cells, immunotherapy aims at mobilising the immune system against malignant cells. In normal conditions, the immune system recognises cancerous cells and destroys them. To evade the immune system, cancerous cells must either overload the immune system or stay invisible to it.

Thus, immunotherapy has the potential to cure even malignant metastatic disease by leveraging the patient's own immune defences.

1.2.3.1. Active immunotherapy

There are currently two categories of immunotherapy: active and passive. Active immunotherapy forces the patient's own immune system to attack the cancer, such as by using antibodies that bind to cancer cells and then flag these cells to be destroyed by the immune system. Examples of drugs following such a strategy are trastuzumab and cetuximab, which are monoclonal antibodies, highlighted by the "mab" ending in their name. This approach has been extended to molecules that can bind to two binding sites, for example on a T cell and on a cancer cell to force the T cell to attack the cancer cell. These drugs are called bispecific antibodies[17].

Other approaches for active immunotherapy are cancer vaccines, which help the immune system recognise cancer cells, and chimeric antigen receptor T cell (CAR-T cell) therapies. T cells are major components of the immune system. To implement the CAR-T cell approach, T cells need to be extracted from the patient and modified to recognise and attack cancer cells before being reinjected into the patient.

1.2.3.2. Passive immunotherapy

Passive immunotherapy works by either stimulating the immune system (i.e., cytokines), or by disabling defence mechanisms in cancer cells. One such defence mechanism is the presence of programmed death-ligand 1 (PD-L1) at the surface of certain cancer cells. This ligand prevents lymphocytes from attacking the cell. Drugs such as pembrolizumab or nivolumab, as well as monoclonal antibodies, can block this mechanism by rendering lymphocytes insensitive to this ligand.

1.2.3.3. Side effects

Immunotherapies, passive immunotherapies in particular, come with side effects. By altering the general behaviour of the immune response, they can provoke potentially life-threatening auto-immune reactions in patients. They also provoke more common side effects, such as nausea, diarrhoea, or fatigue.

1.2.4. Dynamics of the side effects

In conclusion, all SACTs come with a trade-off between antitumoral benefits and side effects. These side effects affect the HRQoL of patients and can even be linked in some cases to early mortality. Patients with a generalised tumour burden who respond poorly to the treatment and with degraded performance status are at high risk of short-term mortality with or without SACT administration. They will also still experience the short-term side effects of the SACT if given, potentially further reducing their HRQoL (see Figure 4).

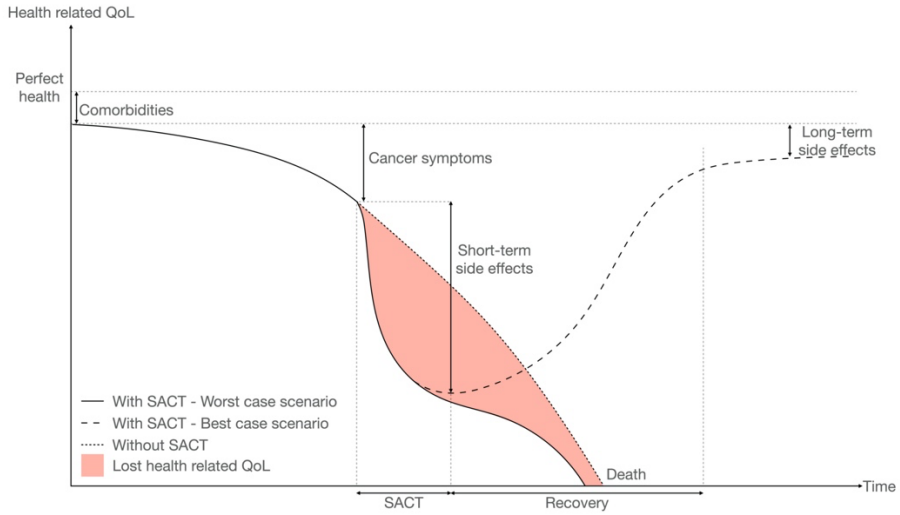


Figure 4. *A model of health-related quality of life for a patient with limited survival.*

Therefore, late SACT administration should be avoided in some cases to limit the risk of negatively impacting the HRQoL of patients close to death.

1.3. Aims of this work

This work had two aims. The first aim was to characterise late SACT usage through a monitoring indicator, as described in Paper I, and investigate the validity of SACT reporting in a national registry to assess the feasibility of implementing this indicator nationally, as described in Paper II. The second aim was to build and compare dynamic predictive models of 30-day mortality that could potentially be implemented in a decision support tool. This decision tool, once validated in a prospective study, would be intended to help clinicians prevent late SACT to limit the risk of harming patients near the end of life due to unnecessary SACT, as well as limit drug spending, as described in Paper III.

2. Materials and methods

2.1. Monitoring indicators for short-term mortality of cancer patients following SACT

The first part of this work was to define late SACT administration and investigate its prevalence and associated risk factors, as described in Paper I. Different monitoring indicators using different definitions for late SACT administration have been proposed, leading to heterogenous results throughout the literature.

2.1.1. Literature review

To draw the landscape of the various monitoring indicators used in similar contexts, we conducted a literature review. We searched in PubMed for articles published between Jan 01, 2010, and Dec 31, 2020, investigating short-term mortality after SACTs. We used the keywords ("30-day mortality" or "30 days mortality" or "early mortality" or "end of life")

and ("anticancer" or "anti-cancer" or "chemotherapy"). From this search, we identified 98 publications. After further filtering to remove duplicates (n=3), articles not focussed on solid tumours (n=18), that were not accessible (n=11), or reported qualitative results (n=26), we found 40 publications.

The corresponding indicators differed mostly in three ways: the inclusion criteria, the time point of interest in the treatment before death, and the threshold of delay between the time point and death. Concerning the inclusion criteria, there were two main strategies: one considered only the dead patients, which was promoted by Earle et al.[18], and the other included all patients treated during a period, typically a calendar year, regardless of whether they were dead[19–26]. The first strategy can be further subdivided into three subcategories based on the circumstances of death: patients who died from any cause, considering if they received a treatment close to death[27,28,37,29–36], for example in the last 6 months, or not[38,39,48,49,40–47]; or patients who died from cancer, often in a palliative context[50–54]. For the second strategy, as defined by Wallington et al.[23], patients can theoretically be included multiple times in different periods if they are treated over more than one period.

Regarding the time point of interest in the treatment, most of the publications considered the last administration or start of the last cycle (36/40, see Table 2). A cycle was commonly a few days long, and these two time points led to similar results in terms of short-term mortality. Concerning the threshold on the delay between the time point and death, 30 days (or 1 month) was used most frequently (39/40).

Table 2. Distribution of references per time point and delay used.

Time point	All	Last SACT administration	Start of SACT
Delay			
All	40	36	8
14 days	20	20	3
30 days	39	35	7
Other	23	21	2

The definition of SACT drugs can influence the time point, as it is based on drug administration or treatment start, and potentially also the cohort if only patients treated with these drugs are included. In some cases, a wide variety of drugs used in cancer treatment were considered, including supportive medications[55], whereas other studies only considered cytotoxic drugs[20].

2.1.2. Conditioned on the future

When the time point and delay have been defined, the cases of late SACT are easily identified, but problems arise when trying to define the cohort for comparison. In other words, the goal is to define a ratio at which the number of late SACTs is the numerator but the denominator can be defined in different ways, as illustrated by the diversity of inclusion criteria found in the literature review. All reviewed articles included one record per patient, which is a common practice in medicine but leads to bias if the inclusion criteria rely on future events.

As introduced in Paper I, the main limitation of the approaches proposed by Earle et al. and Wallington et al. is that they are conditioned on future events. This is a kind of selection bias that can lead to improper assessment of risk factors. Defining a risk factor profile for a patient implies that this patient is evaluated when first seen and not when the outcome is known. In the case of Earle et al., because only patients who died from cancer are included, the risk factors are only applicable for patients who will die from cancer. However, when a patient starts a SACT, dramatically different outcomes are possible, from long-term remission to short-term death due to unsuccessful treatment. The clinician may be able to predict the outcome correctly from known risk factors and experience but cannot be sure of it. To define appropriate risk factors that are usable in a clinical context, all patients should be included. Not considering the patients for whom it went well will create bias toward non-responders.
In addition, an increasing number of patients receive multiple treatments throughout the course of their disease and, using the Earle et al. approach, only the last one is considered. This means that the risk factors that are calculated will only be appropriate for the last treatment. The same limitation is present for Wallington et al.'s approach, as only the last SACT in a period is considered. Figure 5 illustrates different patient trajectories and which SACTs are considered in both cases. A bias toward later treatments is present in both.

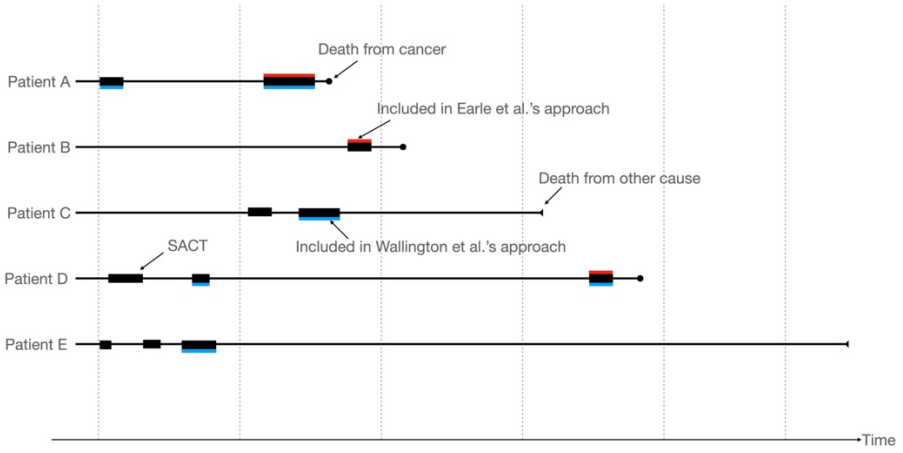


Figure 5. *Examples of SACTs included in Earle et al. and Wallington et al.'s approaches. The black rectangle represents the treatments that each patient actually received. The red overline and blue underline indicate which SACTs would be taken into account in the calculation of Earle et al.'s and Wallington et al.'s indicators, respectively.*

A clinician wants to assess the risk profile of the individual patient before starting SACT to avoid late treatment, which means that each SACT should be considered, but this is not the case, as illustrated in Figure 5.

2.2. Datasets

To characterise the prevalence of late SACT administration, evaluate corresponding risk factors, and build predictive models for the short-term mortality of cancer patients, retrospective data were obtained from a number of registries. In the last few decades, Denmark has invested in solutions to digitise health care data across the country. While some hurdles are still present, notably for clinical work with the lack of a national electronic patient journal, this investment has enabled unique research projects and put Denmark at the forefront of epidemiological studies[56]. The present work was based on registry data from five different sources.

2.2.1. CPR dataset

In Denmark, each citizen is given a number, their Danish civil registration number, which is systematically used across all administrative and health care systems. This number facilitates the coupling of datasets. The corresponding dataset from the Danish Civil Registration System (CPR dataset) also contains information on the individual's sex, date of birth, and date of death, which were used in this work. To improve the security of this data, the CPR numbers were encrypted to enforce pseudo anonymisation in all datasets.

2.2.2. PAS

The five Danish regions manage the hospital system in Denmark. In the North Denmark Region, all interactions of patients at the hospitals are recorded in the Patients Administrative System (PAS). Such systems are in place in all regions and are primarily used to obtain funding from the government to finance the healthcare system. The structure of the PAS data can be decomposed into either in-patient or out-patient contacts and health care procedures. Each procedure is connected to a contact, which typically contains multiple procedures. A contact has associated data, notably the start and end dates and the associated diagnosis. A contact can potentially last for years, especially for chronic diseases. A procedure is defined by the start and end dates and a procedure code. Procedures can be diverse, such as treatment-related (e.g., surgery and chemotherapy) or clinical consultations. All codes used in PAS follow the Danish Healthcare Classification System[57] (SKS). Diagnosis codes in the SKS are similar to codes from the ICD-10 classification[58] developed by the WHO. These diagnosis codes can be used to infer comorbidities, side effects, and the patient's cancer trajectory.

2.2.3. MedOnc

To prescribe, administer, and register the SACTs, clinicians at the Department of Oncology, Aalborg University Hospital, use the prescription software ARIA OIS for Medical Oncology v13.7 (Varian Medical Systems Inc., Palo Alto, CA, USA) (MedOnc). The data contained within this solution includes detailed information on the type of drug used, the frequency of administration, the dose administered, the PRN (Pro Re Nata = as needed) status, the corresponding regimen name, and cycle number. It also contains information on the height and weight used to calculate doses. Most of the information concerning drug administration was only available as text, including the commercial name of the product used with occasional misspellings and the frequency of administration. Reporting issues could also be detected in this dataset. Therefore, extensive data management was needed to extract usable data.

2.2.4. Patobank

The Danish National Pathology Registry (Patobank) dataset contains data on histopathological results. These data inform on the cancer subtype based on topography, morphology, and/or biomarkers. The morphology coding is similar to the International Classification of Diseases for Oncology, 3rd Edition[59] (ICD-O-3). Histopathological information, such as the subtype and biomarkers, can be used to select the appropriate treatment. Examples of biomarkers are PD-L1 expression in lung cancer and BRAF mutation in colorectal cancer.

2.2.5. LABKA

The results of biochemical analyses are collected in the Clinical Laboratory Information System (LABKA). This dataset primarily contains blood test results, such as blood cell counts or ion concentrations. To ensure comparability across Denmark, the Nomenclature, Properties, and Units[60] (NPU) classification is used. Biochemical results are used to monitor the health status of the patients, notably to detect neutropenia or to monitor inflammation.

2.2.6. The DNPR and its validation

The Danish National Patient Registry (DNPR) is a national registry containing information from the patient administrative systems from all five regions in Denmark and, therefore, includes most of the data stored in PAS. All of the datasets that are used are available at the national level or at least in more than just the North Denmark Region apart from MedOnc. The PAS and DNPR contain some information on treatment as procedures, but MedOnc was considered a more reliable data source because it is used by medical personnel, where the DNPR is fed by administrative personnel. However, using MedOnc prevents implementation of the proposed indicator on the national scale. To investigate the validity of the DNPR data for SACTs, we conducted a validation study of this dataset in Paper II.

2.3. Predictive modelling

The main challenge in limiting late SACTs is the ability to accurately predict when it is "late" (i.e., how close to death the patient is). Clinicians tend to be overoptimistic concerning patient survival[61]. In Paper III, we built and compared dynamic predictive models for 30-day mortality in patients with advanced lung cancer with the intent to use these models in a decision support tool. Dynamic refers to the fact that these models should be able to predict the 30-day mortality at any time point in the patient's trajectory. The corresponding outcome is binary (i.e., will the patient die within 30 days or not). The aim was to help clinicians better assess the short-term mortality of their patients at the individual level and make treatment decisions accordingly.

Similar studies from the literature did not investigate the use of artificial neural network-based approaches or used a limited set of covariates[62–66]. However, one similar study focussed on predicting 6-month mortality for another intent[67].

2.3.1. Measure of performance

When working with binary outcomes, a prediction can be either positive or negative and either true or false. Therefore, there are four possible cases: true-positives, true-negatives, false-positives, and false-negatives (see Figure 6). Performance is typically evaluated using functions of these numbers. The most commonly used metrics are the sensitivity, also called recall or the true-positive rate, and the specificity. In some contexts, the precision, also called the positive predictive value, is also used. The sensitivity is the ratio of true-positives among positives and informs on how good the predictions are at identifying the positive cases. The specificity is the ratio of true-negatives among negatives and informs on how good the predictions were at avoiding false-positives. The precision is the ratio of true-positives among the cases predicted as positive (i.e., how likely is a positive prediction to be correct). The F_1 score is the harmonic mean of the sensitivity and the precision and, thus, summarises these two metrics.

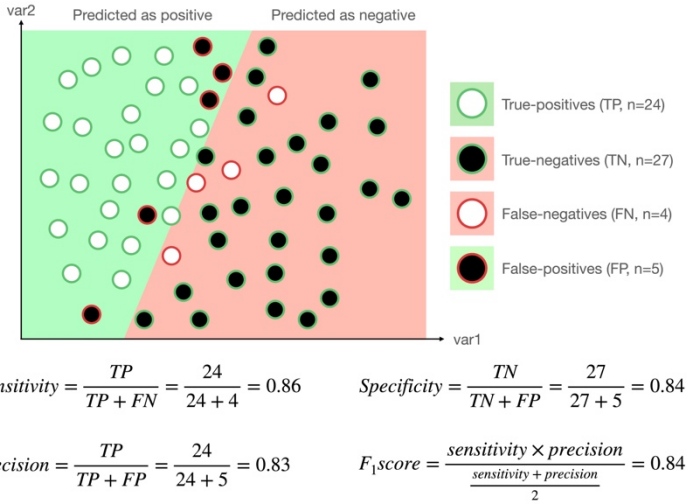


Figure 6. *Illustration of true-positives, false-positives, true-negatives, false-negatives, and associated popular metrics.*

In practice, predictive models for binary outcomes produce probabilities for being positive. The outcome is predicted to be positive if it is above an arbitrary threshold of this probability, and the predictions are used to compute the various performance metrics. To evaluate the overall performance of a model, a classical approach is to calculate the sensitivity and specificity for different values of the threshold and to plot them against each other. In practice, the sensitivity is plotted against 1-specificity, also called the false-positive rate (FPR). This results in a plot called the receiver operating characteristic (ROC) curve. The area under the ROC curve (ROC AUC) is a popular approach for evaluating the predictive power of such models (see Figure 7). One limitation of this approach is that it is sensitive to unbalanced datasets (i.e., if one outcome is much more frequent that the other). In the case of a frequent negative outcome, the model will tend to be very good at predicting true negatives.

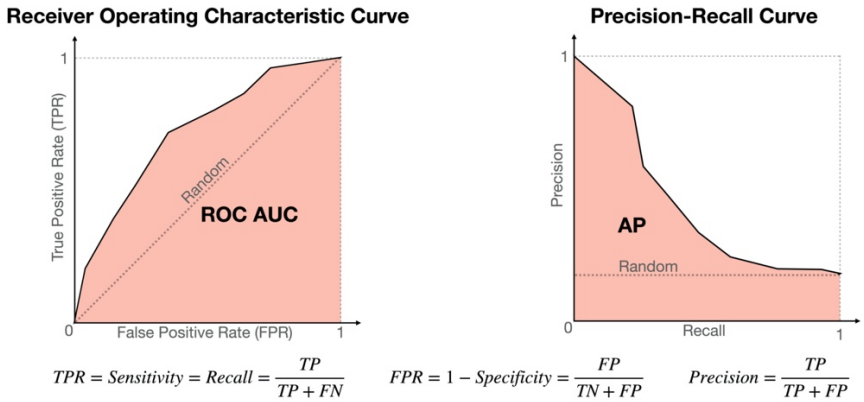


Figure 7. *Example of receiver operating characteristic and precision-recall curves. ROC AUC, area under the ROC curve; AP, average precision*

Therefore, adding more negatives will decrease the FPR for a given threshold while mostly maintaining the sensitivity, leading to an artificial increase in the ROC AUC and making results difficult to compare. To circumvent this limitation with unbalanced datasets, another metric has been proposed, the average precision (AP), which represents the area under the curve of the precision-recall curve[68]. This precision-recall plot displays the precision (or the positive predictive value) against the recall (or sensitivity) for various values of the threshold and calculates the area under the curve. In practice, the values of the threshold used to plot both the ROC and precision-recall curve are chosen based on predictions for each data point, to incrementally vary the sensitivity (or TPR or recall) and calculate the corresponding FPR or precision to plot a new point of the curve.

2.3.2. General learning theory

There are two approaches in statistical learning: supervised and unsupervised learning[69]. In supervised learning, the aim is to build a model capable of predicting an outcome given a number of covariates based on training data for which the covariates and outcome are known. The performance of the model is evaluated based on its ability to predict the outcome. In unsupervised learning, the model is expected to learn from the data without any outcome or label. This is typically used for clustering analyses, in which the model groups the data points into clusters based on a similarity metric. In this PhD study, we exclusively worked with supervised learning with a binary outcome, denoted as 0 or 1, where 0 is considered negative and 1 positive.

The goal of a predictive model is to predict as accurately as possible the outcome y_k based on a set of covariates, x_k, for the k^{th}, $k = 1, \dots, m$ data point (e.g., a patient). The outcomes are collectively referred to as y and the sets of covariates as x. This is done by formulating a predictive function f, taking x_k as input and returning the outcome y_k. The function is supposed to depend on a set of parameters, β, that need to be optimised to maximise the performance of the model. A loss function, L, is defined that measures a distance between the predictions and the known outcomes. A classical function for the

loss function is the residual sum of squares (RSS) used to measure the distance between outcome y_k and prediction $\hat{y}_k$. The prediction $\hat{y}_k$ is a value of function f of the covariates (x_k) and the model parameters (β). Letting m denote the number of data points, this can be stated as:

$$L(x, y, \beta) = \text{RSS}(x, y, \beta) = \sum_{k=1}^{m} (y_k - \hat{y}_k)^2 = \sum_{k=1}^{m} \big(y_k - f(x_k, \beta)\big)^2.$$

Optimisation is achieved by finding the $\hat{\beta}$ that minimises L:

$$\hat{\beta} = \text{argmin}_{\beta} L(x, y, \beta).$$

A more general approach to calculating a loss function is based on maximising the likelihood function, where the likelihood function $\mathcal{L}$ is the density of getting the outcome y knowing the sets of covariates x and the set of parameters β, and ℓ is the likelihood for each datapoint. In the case of a binary outcome, it can be stated as:

$$\mathcal{L}(x, y, \beta) = \prod_{k=1}^{m} \ell(x_k, y_k, \beta),$$

where

$$\ell(x_k, y_k, \beta) = I(y_k = 1)\Pr(y_k = 1 \mid \beta, x_k) + I(y_k = 0)\Pr(y_k = 0 \mid \beta, x_k).$$

In practice, the log-likelihood is used, allowing us to work with a sum instead of a product. The log function is increasing, therefore maximising the log-likelihood is equivalent to maximising the likelihood. As the goal is to minimise the loss function, a negative sign is added:

$$L(x, y, \beta) = -\log\big(\mathcal{L}(x, y, \beta)\big) = \sum_{k=1}^{m} -\log\big(\ell(x_k, y_k, \beta)\big).$$

2.3.3. Overfitting

A key aspect in machine learning is the ability of the models to predict the outcome based on a new set of covariates. The models can become so specific to the training data used for learning, that it will perform poorly when presented with new data. This is called overfitting.

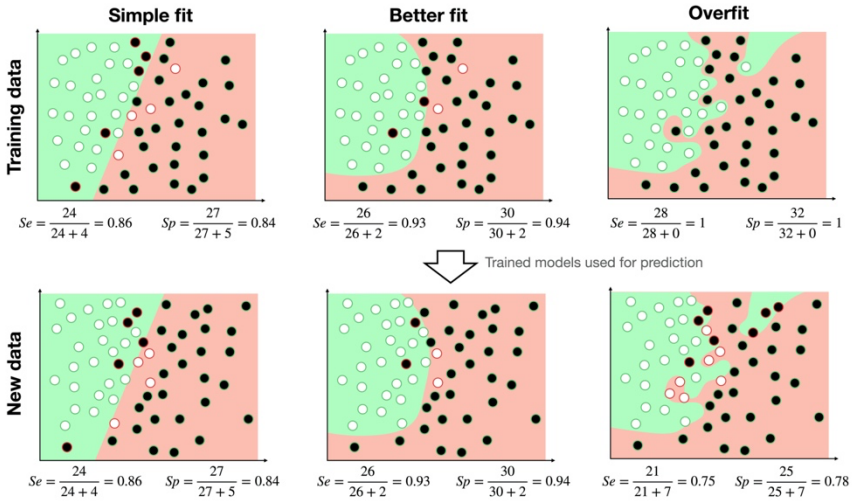


Figure 8. *Illustration of overfitting with three different models. Se, sensitivity; Sp, specificity.*

In Figure 8, three different models were trained on the same data set, with different performances. The first and simplest model performed worst on the training data and the more complex model the best. However, when presented with new data, the first two models maintained their performance while the last model underperformed because it was too specific to the original training data (i.e., this model was not generalisable).
Therefore, to properly evaluate the performance of a predictive model, the performance must be evaluated on different data. Typically, this is achieved by splitting the data into the training data and the test data, where only the training data is used for training and the performances are evaluated on the test data.

2.3.4. Hyperparameter optimisation

Most predictive models can be configured to limit overfitting, but this gives a risk of underfitting, with the model not learning sufficiently from the data. This configuration is done through hyperparameters, such as by using a coefficient to penalise the size of parameters used in the model through a loss function. The difference between hyperparameters and parameters is that the hyperparameters are set before the learning phase starts, whereas the parameters are optimised during the learning phase. To find the optimum values for the hyperparameters, the performance of the models for the tested hyperparameter combinations must be evaluated. The test set could be used, but this will also lead to overfitting because the hyperparameters are then optimised based on the same data set that will be used to evaluate the performance of the model. The final evaluation of the performance of the selected model should be made on a data set that was not used in either the learning phase or for the selection of the hyperparameters to achieve a better assessment of the generalisability of the model. Ideally, an external data set should be used, but this is not often available. To solve this problem, the test set can be split into a validation set and a final test set. The validation set is used to optimise the

hyperparameters, and the final performance of the model is assessed using the final test set.

Different methods are available for finding the best hyperparameters. A naïve approach would be to systematically explore the hyperparameter space for the best combinations. This approach is called a grid search. A more sophisticated and effective approach is Bayesian optimisation, which is based on Bayes' rule[70]. Bayes' rule states how to update the current probability of proposition A, $\Pr(A)$, called the prior probability based on observed evidence B, leading to a new probability called the posterior probability, $\Pr(A|B)$.

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}$$

This approach can be extended to a version of Bayes' rule involving a prior distribution π on an unknown variable, z, leading to a posterior distribution on the unknown value z, given the set of observations o.

$$\pi(z|o) = \frac{\pi(z)\mathcal{L}(o|z)}{\int_z \pi(z)\mathcal{L}(o|z)dz},$$

where $\mathcal{L}$ the likelihood of observing o given z.

In Bayesian optimisation of hyperparameters, the first step is to define a prior distribution on the loss function by a Gaussian process (GP) on the hyperparameter space. In practice, a selection of points, h_k, $k = 1, \ldots, N_I$, is used as indices for the Gaussian process, where N_I is the number of indexed points. For each indexed point, a Gaussian distribution of mean μ_0 and covariance function Σ_0 is initially defined. The mean μ_0 is typically a constant over the hyperparameters, and covariance function Σ_0 is assumed to be a smooth function of the distance between a pair of points in the hyperparameter space and zero for points where the loss function was observed. Using Bayes' rule, a new GP is calculated on the indexed points as the posterior distribution on the loss function given all the observed losses. The approximate loss function, i.e., the mean function of the posterior GP, is called the surrogate function[71]. Next, we use the maximum of an acquisition function defined from the posterior GP to decide the set of hyperparameters for which the loss function should be evaluated. As an acquisition function, one could, for example, choose the variance function of the GP because this will push the point away from already explored areas. However, more sophisticated approaches exist based on the expected improvement or maximum entropies[70]. Once a new data point for the loss function is observed, an updated posterior is generated, leading to a new acquisition function and, therefore, a new set of hyperparameters for which the loss function should be evaluated. This process is repeated until a specific number of iterations is reached or when a convergence criterion is met (see Figure 9).

This will eventually lead to a concentration of hyperparameters around the minimum value of the loss function. Compared to the grid search, which can realistically only handle a few dimensions, Bayesian optimisation is capable of handling higher

dimensional spaces, in practice up to 20[ref70]. This approach has been implemented in the Python library keras-tuner[72].

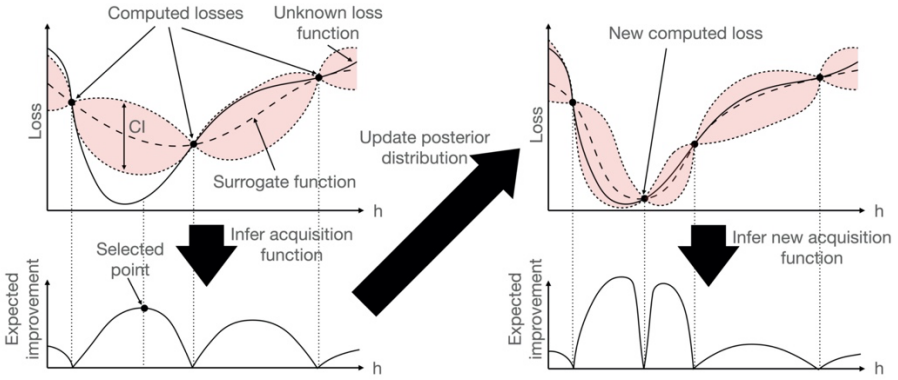


Figure 9. *Example of an iteration in Bayesian optimisation of hyperparameters. Confidence intervals (CIs) describe typical realisations of the prior on the loss function and maximum variation. h is a value in the hyperparameter space.*

2.3.5. Predictive models

In this work, we benchmarked five different machine learning techniques to build predictive models for the 30-day mortality.

2.3.5.1. Logistic regression with elastic net regularisation

Logistic regression is a regression technique in which the logit-function of the probability of getting the outcome 1 for the k^{th} data point is calculated as a linear combination of n covariates $x_{ik}, i = 1, \ldots, n$, with corresponding coefficients β_i and β_0, where β_0 is an offset independent of covariates.

$$\text{logit}\big(\Pr(y_k = 1 \mid \beta, x_k)\big) = \beta_0 + \sum_{i=1}^{n} \beta_i x_{ik},$$

where

$$\text{logit}(x) = \log\left(\frac{x}{1-x}\right).$$

The coefficient $\hat{\beta}$ used in the linear combination is selected to minimise the negative log-likelihood function L, which aims at maximising the probability of outcome y given the coefficient β and sets of covariates x.

$$\hat{\beta} = \text{argmin}_\beta(L(x, y, \beta))) = \text{argmin}_\beta \left(\sum_{k=1}^{m} -\log\left(\ell(x_k, y_k, \beta)\right) \right)$$

The main issues with this type of model are potential overfitting when many covariates are used and interpretability when correlated or multicollinear covariates are present. To

solve these problems, regularisation mechanisms have been proposed, such as ridge regression, addressing most instabilities due to correlation or multicollinearity, and the least absolute shrinkage and selection operator (LASSO), tackling the overfitting by enforcing covariate selection. The regularisation mechanism in the Ridge regression comes from a penalty added in the loss function L based on the sum of the squares of the values of the coefficients, with λ being the regularisation coefficient.

$$l(x, y, \beta, \lambda) = -\log(L(x, y, \beta)) + \lambda \sum_{i=1}^{n} {\beta_i}^2$$

The regularisation in LASSO is achieved by applying a penalty proportional to the sum of the absolute values of the coefficient.

$$l(x, y, \beta, \lambda) = -\log(L(x, y, \beta)) + \lambda \sum_{i=1}^{n} |\beta_i|$$

These two methods address two different issues and have been combined into one regularisation, the elastic net regularisation[73]. The corresponding hyperparameters are the regularisation strength, λ, and the elastic net mixing parameter between the Ridge regression and LASSO, α.

$$l(x, y, \beta, \lambda, \alpha) = -\log(L(x, y, \beta)) + \lambda \left(\alpha \sum_{i=1}^{n} {\beta_i}^2 + (1 - \alpha) \sum_{i=1}^{n} |\beta_i| \right)$$

2.3.5.2. Random forest classifier

A classical approach in predictive modelling of binary outcomes is to use decision trees. The idea is to initially find the covariate and the corresponding threshold for continuous variables, or the value for categorical variables, that best separates the positives and negatives, and then repeat the process in each branch. This separation criteria can be stated as a question with a "Yes" or "No" answer; for example, "Is the value for the covariate age above 65?" or "Is the patient a male?" A tree is characterised by its depth, corresponding to how many times the splitting procedure is performed. The splitting procedure can be stopped at any time point based on the performance criteria. An example of a decision tree based on two covariates giving a probability of being positive is presented in Figure 8. Each covariate can be reused multiple times for splitting in the same tree. A decision tree for a binary outcome does not have to provide a binary prediction, but can also generate a probability of being positive or negative.

As seen in Figure 10, the predictive function based on a decision tree, f, splits the covariate space into N_D domains: $R_j,\ j = 1, \ldots, N_D$ with a corresponding probability b_j of belonging to R_j.

$$f(x_k) = \sum_{j=1}^{N_D} b_j \, I(x_k \in R_j)$$

Importantly, N_D, and therefore R_j and b_j, depends on the hyperparameters used to build the tree, which are principally stopping criteria, such as the maximum tree depth.

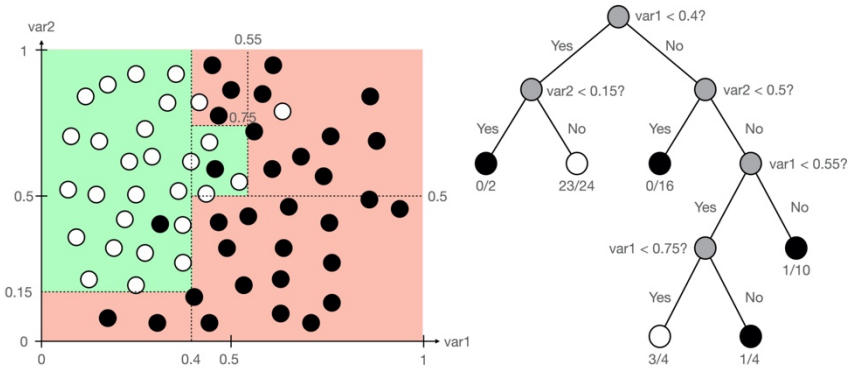


Figure 10. *Example of a decision tree with two covariates, var1 and var2. In the covariate space on the left, the positive cases are represented by white circles and the negative cases by black circles. In the decision tree on the right, grey circles are decision nodes. The end nodes, also called leaves, are represented by white or black circles based on the probability of being positive using a 50% threshold.*

A problem with decision trees is that they tend to overfit if their depth is too large, but they learn very little with a low depth. To circumvent this limitation, ensemble approaches that create multiple trees and aggregate the results have been proposed (see Figure 11).

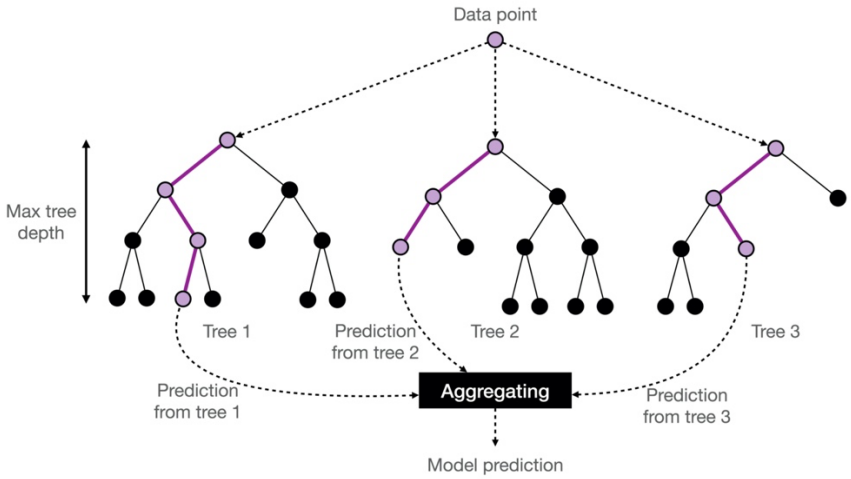


Figure 11. *Example of the prediction process for a tree-based ensemble model with three trees and a maximum depth of 3.*

The prediction function for the whole model, f, is a function g, often the mean, of the outcomes of each tree, $f_j,\ j = 1, \ldots, N_T$, where N_T is the number of trees.

$$f(x_k) = g\big(f_1(x_k), f_2(x_k), \ldots, f_l(x_k)\big) = \frac{1}{N_T}\sum_{j=1}^{N_T} f_j(x_k)$$

One famous tree-based ensemble technique is called random forest (RF)[74]. A RF model generates an ensemble of trees, also known as a forest, for which covariates and records are randomly selected from the dataset to train each tree. This type of model is primarily optimised on the number of trees and the maximum depth of these trees.

2.3.5.3. Gradient boosting classifier

Another tree-based ensemble approach that has gained traction in recent years is tree-based gradient boosting[75] (GB). In contrast to the RF, covariates and data are not removed randomly; instead, this approach follows an iterative process in which the difference between the outcome and the predictions, called residuals, is used to build a new tree that will be included in the ensemble model (see Figure 12).

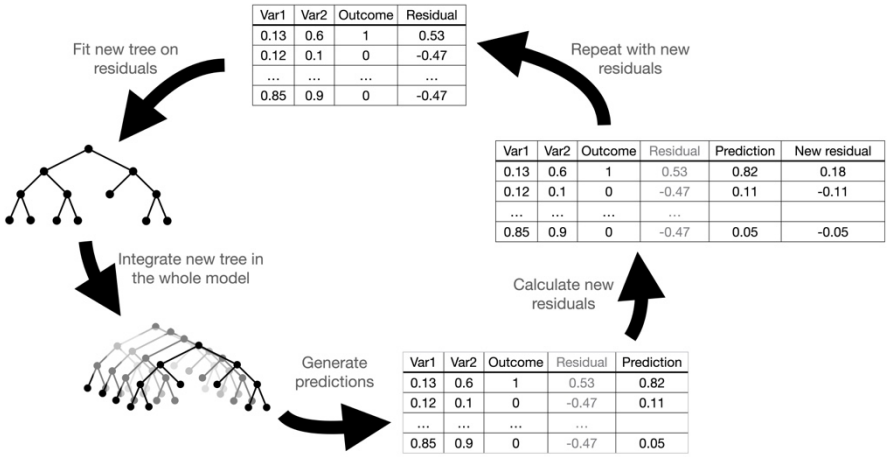


Figure 12. *Iterative learning process for a tree-based gradient boosting classifier.*

The process is started by calculating the mean of the outcomes as the function for the first fitted model F_0. From the previous model function, F_{j-1}, residuals are calculated as the difference between the outcomes and the predictions. These residuals are used to fit a new tree with predictive function f_j. After this new tree is fitted, a weighting coefficient, γ_j, is calculated for f_j to minimise the loss function, L, on the function for the whole model, F_j, combining F_{j-1} and f_j.

$$F_j(x_k) = F_{j-1}(x_k) + \gamma_j f_j(x_k)$$
$$\text{with } \gamma_j = \text{argmin}_\gamma \left(\sum_{k=1}^{m} L\left(y_i, F_{j-1}(x_k) + \gamma f_j(x_k)\right) \right)$$

Typically, the weighted residual predictions for each tree function, $\gamma_j f_j$, are further weighted by the learning rate, λ, before the new residuals are calculated to facilitate convergence.

$$F_j(x_k) = F_{j-1}(x_k) + \lambda\gamma_j f_j(x_k)$$

where the predictive function, f, is the output of the last iteration, and N_T is the number of trees.

$$f(x_k) = F_{N_T}(x_k)$$

Like the RF, this type of model can primarily be optimised on the number of trees and the maximum depth of these trees, but also on the learning rate λ. The number of trees represents the number of times the iterative process is performed.

2.3.5.4. Multilayer perceptron

Artificial neural network methods have gained considerable traction in the last decade due to exceptionally good results in the fields of image and voice recognition and natural language processing, which were made possible by improvements in processing speed. The corresponding models rely on programmatic neurons that apply a linear or non-linear function to a linear combination of inputs to generate an output. This function, called an activation function, is often of a sigmoid type, such as the one used in the logistic regression:

$$N_{\mathrm{jk}}(y_{j-1}) = \sigma\left(\beta_{jk0} + \sum_{l=1}^{n_{j-1}} \beta_{jkl}\, y_{(j-1)l}\right) = y_{jk},$$

where j is the layer number, k the neuron number in the layer, N_{jk} the output function for the corresponding neuron, σ the activation function, n_j the number of neurons in layer j, y_j the output vector of dimension n_j from layer j composed of the values y_{jk}, β_{jk0} the bias, and β_{jkl} the weight used in the corresponding neuron. These neurons are organised in layers, with each layer feeding into the next (see Figure 13).

In predictive modelling with binary outcomes, the final layer contains only one neuron that outputs a probability (i.e., the predictive function f). The model learns by back-propagation; the prediction is compared to the actual outcome and the weights and bias for each neuron updated going backwards in the network. This back-propagation is handled by an algorithm called the optimiser. Different versions of this algorithm are available[76]. The size of this update is regulated by the learning rate.
A simple architecture for an artificial neural network model is multilayer perception (MLP) as shown in Figure 13.

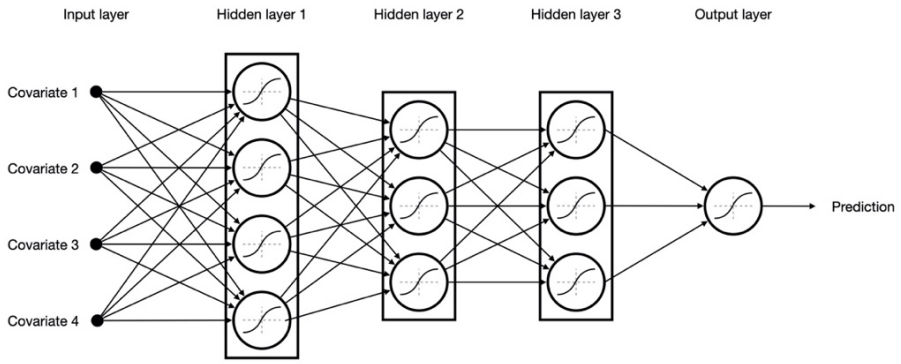


Figure 13. *Example of a multilayer perception with three hidden layers with 4, 3, and, 3 neurons, respectively.*

These types of models contain many hyperparameters, including the number of layers, the number of neurons and activation method at each layer, the learning rate, and the

optimiser. In addition, a dropout step can be placed between each layer to randomly remove some covariates to avoid too much overfitting, adding an additional hyperparameter to regulate the number of covariates being removed. These hyperparameters can typically be determined by Bayesian optimisation.

2.3.5.5. LSTM

One of the main limitations of the MLP for natural language processing or voice recognition is their lack of memory. Understanding a text and recognising words in speech are typically dependent on the context and, thus, keeping track of the previous data (i.e., words in the case of natural language processing) can help dramatically improve performance. To tackle this issue, recurring neural networks have been developed in which the prediction is calculated from a sequence of data points, with the output at each data point in the sequence being dependant on the data for the new data point but also the output data at the previous data point. However, naïve implementation of such networks leads to instability in the fitting process. The long-short term memory (LSTM) model[77] was proposed to limit this problem. It is based on the idea of an internal state that helps select the information to be kept while going through the sequence (see Figure 14).

The layers comprising this model are similar to the layers used in the MLP and, thus, lead to similar hyperparameters (i.e., the number of layers, the number of neurons, the activation method at each layer, the learning rate, and the optimiser). Dropout mechanisms can also be included. Furthermore, the output of the LSTM can be fed into an MLP, further complexifying the model.

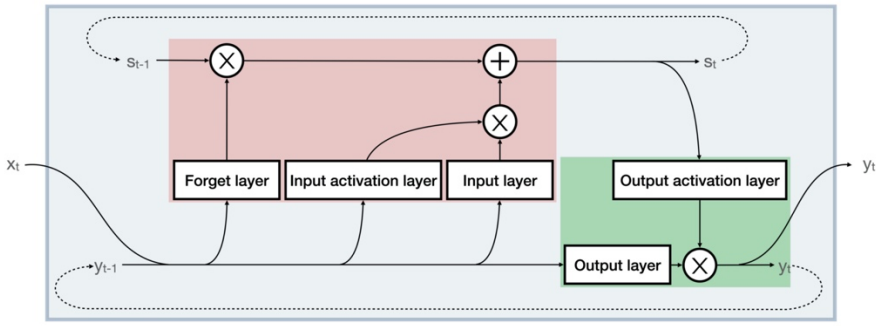


Figure 14. *Example of an LSTM model. The red block is for the management of the internal state s_t and the green block is for the generation of the output y_t, x_t is the data for the event at time t in the sequence.*

2.3.6. Explainability

In medical research, understanding how covariates affect predictions is of particular interest. Clinicians usually require understanding the reasons behind a clinical decision, and decision support tools should be able to provide that in some form[78]. Therefore, tree-based ensemble and artificial neural network models can be challenging to implement in a clinical context due to their lack of interpretability. An approach based on game theory aims to make the predictions from any model more interpretable. In game theory, Shapley values provide an estimate of the marginal contribution of each

player to a result so it can be allocated fairly to all players. The marginal contribution is the difference between the result with or without the considered player. This approach can be applied to a predictive model by considering covariates as players contributing to the outcome.
This approach has been implemented in a Python library that calculates SHapley Additive exPlanation (SHAP) values[79] for each covariate. SHAP values are local estimations of the marginal effect of each covariate on the prediction of each data point.

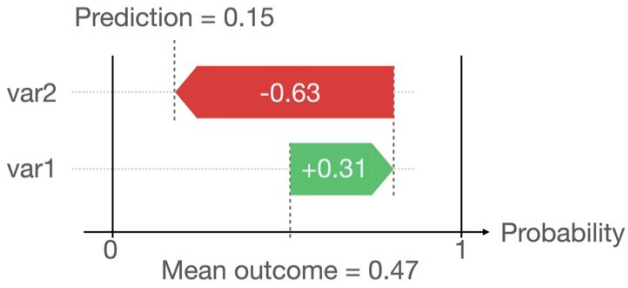


Figure 15. *Decomposition of the difference for one data point between the mean outcome and the prediction in contributions from each covariate in a predictive model with two covariates.*

One of the main advantages of this approach is that it is model agnostic (i.e., it can be applied to any predictive model). For a specific data point, it decomposes the difference between the mean outcome and the prediction in contributions from each covariate (see Figure 15). By calculating these SHAP values for all data points or a random sample of them, it is possible to estimate the overall contribution of each covariate. Though the SHAP approach is considered to be the standard, the explainability of machine learning models is a subject of ongoing research, with alternative strategies being developped[78].

References

1. IARC/WHO. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer. Published 2021. Accessed October 6, 2021.
2. Eurostat. Causes of death statistics.
3. Crick F. Central Dogma of Molecular Biology. *Nature*. 1970;227(5258):561-563. doi:10.1038/227561a0
4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
6. Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol*. 2019;20(11):1493-1505. doi:10.1016/S1470-2045(19)30456-5
7. DeVita VT, Chu E. A history of cancer chemotherapy. *Cancer Res*. 2008;68(21):8643-8653. doi:10.1158/0008-5472.CAN-07-6611
8. WHO Collaborating Centre for Drug Statistics Methodology. Anatomical Therapeutic Chemical (ATC) classification system. https://www.whocc.no/atc/structure_and_principles/. Accessed March 3, 2021.
9. Cooper GM. *The Cell: A Molecular Approach, 2nd Edition*. Sunderland; 2000.
10. Cuddihy AR, O'Connell MJ. Cell-cycle responses to DNA damage in G2. *Int Rev Cytol*. 2003;222:99-140. doi:10.1016/s0074-7696(02)22013-6
11. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer*. 2003;3(5):330-338. doi:10.1038/nrc1074
12. Drugs.com. Paclitaxel. https://www.drugs.com/monograph/paclitaxel.html. Accessed November 9, 2021.
13. Pommier Y. Drugging Topoisomerases: Lessons and Challenges. *ACS Chem Biol*. 2013;8(1):82-95. doi:10.1021/cb300648v
14. Xu H, Yu S, Liu Q, et al. Recent advances of highly selective CDK4/6 inhibitors in breast cancer. *J Hematol Oncol*. 2017;10(1):1-12. doi:10.1186/s13045-017-0467-2
15. Raymond E, Faivre S, Armand JP. Epidermal Growth Factor Receptor Tyrosine Kinase as a Target for Anticancer Therapy. *Drugs*. 2000;60(1):15-23. doi:10.2165/00003495-200060001-00002
16. Wolska-Washer A, Robak T. Safety and Tolerability of Antibody-Drug Conjugates in Cancer. *Drug Saf*. 2019;42(2):295-314. doi:10.1007/s40264-018-0775-7
17. Fan G, Wang Z, Hao M, Li J. Bispecific antibodies and their applications. *J Hematol Oncol*. 2015;8(1). doi:10.1186/s13045-015-0227-0
18. Earle CC, Neville BA, Landrum MB, et al. Evaluating claims-based indicators of the intensity of end-of-life cancer care. *Int J Qual Heal Care*. 2005;17(6):505-509. doi:10.1093/intqhc/mzi061
19. Jones GS, McKeever TM, Hubbard RB, Khakwani A, Baldwin DR. Factors

influencing treatment selection and 30-day mortality after chemotherapy for people with small-cell lung cancer: An analysis of national audit data. *Eur J Cancer*. 2018;103:176-183. doi:10.1016/j.ejca.2018.07.133

20. Wilson M, Mak W, Firth M, Deva S, Findlay M. Mortality within 30 days of systemic anticancer therapy at a tertiary cancer centre: Assessing the safety and quality of clinical care. *N Z Med J*. 2017;130(1460):63-72.
21. McCracken JA, Dabscheck A, Coperchini M, et al. Prospective analysis of 30-day mortality following palliative chemotherapy at a tertiary cancer centre. *Cancer Rep*. 2018;1(1):e1135. doi:10.1002/cnr2.1135
22. Burgers JA, Damhuis RA. 30-Day Mortality After the Start of Systemic Anticancer Therapy for Lung Cancer: Is It Really a Useful Performance Indicator? *ERJ Open Res*. 2018;4(1):00030-02018. doi:10.1183/23120541.00030-2018
23. Wallington M, Saxon EB, Bomb M, et al. 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study. *Lancet Oncol*. 2016;17(9):1203-1216. doi:10.1016/S1470-2045(16)30383-7
24. Gibson AJW, Li H, D'Silva A, et al. Factors associated with early mortality in non-small cell lung cancer patients following systemic anti-cancer therapy: A 10 year population-based study. *Lung Cancer*. 2019;134(February):141-146. doi:10.1016/j.lungcan.2019.06.003
25. Yoong J, Seah JA, Hamilton K, Teo LN, Chong G. Mortality within 30days of receiving systemic anti-cancer therapy at a regional oncology unit: What have we learned? *Asia Pac J Clin Oncol*. 2012;8(4):325-329. doi:10.1111/j.1743-7563.2011.01498.x
26. Silverman R, Smith L, Sundar S. Benchmarking 30 Day Mortality After Palliative Chemotherapy for Solid Tumours. *Clin Oncol (Royal Coll Radiol*. 2014;26(1):236. doi:10.1016/j.clon.2013.12.005
27. Andreis F, Rizzi A, Rota L, Meriggi F, Mazzocchi M, Zaniboni A. Chemotherapy use at the End of Life. A Retrospective Single Centre Experience Analysis. *Tumori*. 2011;97(1):30-34. doi:10.1177/030089161109700106
28. Colla CH, Morden NE, Skinner JS, Hoverman JR, Meara E. Impact of payment reform on chemotherapy at the end of life. *Am J Manag Care*. 2012;8(3S):e6s-e13s. doi:10.1200/jop.2012.000539
29. Jung D, Hwang S, You HJ, Lee J. The realities and associated factors of palliative chemotherapy near the end of life in the patients enrolled in palliative care unit. *Korean J Fam Med*. 2012;33(1):44-50. doi:10.4082/kjfm.2012.33.1.44
30. Ortiz JS. Chemotherapy at the end of life: Up until when? *Clin Transl Oncol*. 2012;14(9):667-674. doi:10.1007/s12094-012-0847-6
31. Karim SM, Zekri J, Abdelghany E, Dada R, Munsoor H, Ahmad I. Time from last chemotherapy to death and its correlation with the end of life care in a referral hospital. *Indian J Med Paediatr Oncol*. 2015;36(1):55-59. doi:10.4103/0971-5851.151792
32. Taberner Bonastre P, Taberner Bonastre MT, Soler Company E, Pérez-Serrano

Lainosa MD. Chemotherapy near the end of life; assessment of the clinical practise in onco-hematological in adult patients. *Farm Hosp*. 2016;40(1):14-24. doi:10.7399/fh.2016.40.1.8918

33. Rautakorpi LK, Seyednasrollah F, Mäkelä JM, et al. End-of-life chemotherapy use at a Finnish university hospital: a retrospective cohort study. *Acta Oncol (Madr)*. 2017;56(10):1272-1276. doi:10.1080/0284186X.2017.1332424
34. Massa I, Nanni O, Foca F, et al. Chemotherapy and palliative care near end-of life: examining the appropriateness at a cancer institute for colorectal cancer patients. *BMC Palliat Care*. 2018;17(1):86. doi:10.1186/s12904-018-0339-8
35. Zhang Z, Chen M-L, Gu X-L, Cheng W-W. Use of palliative chemotherapy near the end of life: a retrospective cohort study. *Ann Palliat Med*. 2020;9(5):2809-2816. doi:10.21037/apm-20-273
36. Hiramoto S, Tamaki T, Nagashima K, et al. Prognostic factors in patients who received end-of-life chemotherapy for advanced cancer. *Int J Clin Oncol*. 2019;24(4):454-459. doi:10.1007/s10147-018-1363-7
37. Urvay S, Civelek B, Özaslan E, Sürel AA. Chemotherapy at the End of Life. *J Palliat Care*. 2020. doi:10.1177/0825859720946505
38. Greer JA, Pirl WF, Jackson VA, et al. Effect of early palliative care on chemotherapy use and end-of-life care in patients with metastatic non-small-cell lung cancer. *J Clin Oncol*. 2012;30(4):394-400. doi:10.1200/JCO.2011.35.7996
39. Goksu SS, Gunduz S, Unal D, et al. Use of chemotherapy at the end of life in Turkey. *BMC Palliat Care*. 2014;13(1):51-56. doi:10.1186/1472-684X-13-51
40. Adam H, Hug S, Bosshard G. Chemotherapy near the end of life: A retrospective single-centre analysis of patients' charts. *BMC Palliat Care*. 2014;13(1):26-30. doi:10.1186/1472-684X-13-26
41. Lee HS, Chun KH, Moon D, Yeon HK, Lee S, Lee SH. Trends in receiving chemotherapy for advanced cancer patients at the end of life. *BMC Palliat Care*. 2015;14(1):4-9. doi:10.1186/s12904-015-0001-7
42. Huo J, Du XL, Lairson DR, et al. Utilization of Surgery, Chemotherapy, Radiation Therapy, and Hospice at the End of Life for Patients Diagnosed With Metastatic Melanoma. *Am J Clin Oncol*. 2015;38(3):235-241. doi:10.1097/COC.0b013e31829378f9
43. Zerillo JA, Stuver SO, Fraile B, Dodek AD, Jacobson JO. Understanding oral chemotherapy prescribing patterns at the end of life at a comprehensive cancer center: Analysis of a Massachusetts payer claims database. *J Oncol Pract*. 2015;11(5):372-377. doi:10.1200/JOP.2015.003921
44. Kempf E, Tournigand C, Rochigneux P, Aubry R, Morin L. Discrepancies in the use of chemotherapy and artificial nutrition near the end of life for hospitalised patients with metastatic gastric or oesophageal cancer. A countrywide, register-based study. *Eur J Cancer*. 2017;79:31-40. doi:10.1016/j.ejca.2017.03.029
45. Mathew A, Achkar T, Abberbock S, et al. Prevalence and determinants of end-of-life chemotherapy use in patients with metastatic breast cancer. *Breast J*. 2017;23(6):718-722. doi:10.1111/tbj.12905

46. Low D, Merkel EC, Menon M, et al. Chemotherapy use at the end of life in Uganda. *J Glob Oncol*. 2017;3(6):711-719. doi:10.1200/JGO.2016.007385
47. Rochigneux P, Raoul JL, Beaussant Y, et al. Use of chemotherapy near the end of life: What factors matter? *Ann Oncol*. 2017;28(4):809-817. doi:10.1093/annonc/mdw654
48. Edman Kessler L, Sigfridsson J, Hatzidaki D, et al. Chemotherapy use near the end-of-life in patients with metastatic breast cancer. *Breast Cancer Res Treat*. 2020;181(3):645-651. doi:10.1007/s10549-020-05663-w
49. Sánchez-Cuervo M, García-Basas L, Gómez de Salazar-López de Silanes E, Pueyo-López C, Bermejo-Vicedo T. Chemotherapy Near the End of Life in Onco–Hematological Adult Patients. *Am J Hosp Palliat Med*. 2020;37(8):641-647. doi:10.1177/1049909119901133
50. Zhu Y, Tang K, Zhao F, et al. End-of-life chemotherapy is associated with poor survival and aggressive care in patients with small cell lung cancer. *J Cancer Res Clin Oncol*. 2018;144(8):1591-1599. doi:10.1007/s00432-018-2673-x
51. Massard V, Salleron J, Krakowski I, Conroy T, Weber B. Chemotherapy at the end of life: Factors of prescription. *J Palliat Med*. 2015;18(8):658-659. doi:10.1089/jpm.2015.0134
52. Fang P, Jagsi R, He W, et al. Rising and falling trends in the use of chemotherapy and targeted therapy near the end of life in older patients with cancer. *J Clin Oncol*. 2019;37(20):1721-1731. doi:10.1200/JCO.18.02067
53. Sheng J, Zhang Y, He X, et al. Chemotherapy Near the End of Life for Chinese Patients with Solid Malignancies. *Oncologist*. 2017;22(1):53-60. doi:10.1634/theoncologist.2016-0013
54. Liu TW, Chang WC, Wang HM, et al. Use of chemotherapy at the end of life among Taiwanese cancer decedents, 2001-2006. *Acta Oncol (Madr)*. 2012;51(4):505-511. doi:10.3109/0284186X.2011.653440
55. Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end-of-life cancer care from administrative data. *J Clin Oncol*. 2003;21(6):1133-1138. doi:10.1200/JCO.2003.03.059
56. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National patient registry: A review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-490. doi:10.2147/CLEP.S91125
57. Sundhedsdatastyrelsen. Disease Classification System - SKS (in Danish). https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer/klassifikation-sygdomme. Accessed March 3, 2021.
58. WHO. ICD-10 Version:2016. https://icd.who.int/browse10/2016/en. Published 2016. Accessed March 26, 2020.
59. IARC/WHO. International Classification of Diseases for Oncology. http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=100&Itemid=577. Accessed March 26, 2020.
60. Joint Committee on Nomenclature P and U (C-S-N of the I and I, Pontet F, Magdal Petersen U, et al. Clinical laboratory sciences data transmission: the NPU

coding system. *Stud Health Technol Inform*. 2009;150(1):265-269.
61. Amano K, Maeda I, Shimoyama S, et al. The accuracy of physicians' clinical predictions of survival in patients with advanced cancer. *J Pain Symptom Manage*. 2015;50(2):139-146.e1. doi:10.1016/j.jpainsymman.2015.03.004
62. Simmons CPLL, McMillan DC, McWilliams K, et al. Prognostic Tools in Patients With Advanced Cancer: A Systematic Review. *J Pain Symptom Manage*. 2017;53(5):962-970.e10. doi:10.1016/j.jpainsymman.2016.12.330
63. Hamano J, Takeuchi A, Yamaguchi T, et al. A combination of routine laboratory findings and vital signs can predict survival of advanced cancer patients without physician evaluation: a fractional polynomial model. *Eur J Cancer*. 2018;105:50-60. doi:10.1016/j.ejca.2018.09.037
64. Adelson K, Lee DKK, Velji S, et al. Development of imminent mortality predictor for advanced cancer (IMPAC), a tool to predict short-term mortality in hospitalized patients with advanced cancer. *J Oncol Pract*. 2018;14(3):e168-e175. doi:10.1200/JOP.2017.023200
65. Renfro LA, Goldberg RM, Grothey A, et al. Clinical Calculator for Early Mortality in Metastatic Colorectal Cancer: An Analysis of Patients From 28 Clinical Trials in the Aide et Recherche en Cancérologie Digestive Database. *J Clin Oncol*. 2017;35(17):1929-1937. doi:10.1200/JCO.2016.71.5771
66. Uneno Y, Taneishi K, Kanai M, et al. Development and validation of a set of six adaptable prognosis prediction (SAP) models based on time-series real-world big data analysis for patients with cancer receiving chemotherapy: A multicenter case crossover study. *PLoS One*. 2017;12(8):1-13. doi:10.1371/journal.pone.0183291
67. Parikh RB, Manz C, Chivers C, et al. Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer. *JAMA Netw open*. 2019;2(10):e1915997. doi:10.1001/jamanetworkopen.2019.15997
68. Tomašev N, Harris N, Baur S, et al. Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nat Protoc*. May 2021. doi:10.1038/s41596-021-00513-5
69. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Vol 26. New York, NY: Springer New York; 2009. doi:10.1007/978-0-387-84858-7
70. Mockus J. On Bayesian Methods for Seeking the Extremum. *Optim Tech*. 1974:400-404.
71. Frazier PI. A Tutorial on Bayesian Optimization. 2018;(Section 5):1-22.
72. O'Malley T, Bursztein E, Long J, et al. Keras-tuner. https://github.com/keras-team/keras-tuner. Published 2019.
73. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x
74. Ho TK. Random decision forests. *Proc Int Conf Doc Anal Recognition, ICDAR*. 1995;1:278-282. doi:10.1109/ICDAR.1995.598994
75. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001;29(5). doi:10.1214/aos/1013203451
76. Chollet F, Others. Keras. https://keras.io/. Published 2015.

77. Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput*. 1997;9(8):1735-1780. doi:10.1162/neco.1997.9.8.1735
78. Belle V, Papantonis I. Principles and Practice of Explainable Machine Learning. *Front Big Data*. 2021;4(July):1-25. doi:10.3389/fdata.2021.688969
79. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. *NeurIPS Proc*. May 2017.

Paper I: Thirty-day mortality following systemic anticancer therapy: Evaluating risk factors without selection bias in a real-world, population-based cohort from 2009 to 2019

Submitted to Clinical Oncology, September 2021

Authors

Charles Vesteghem[1,2,3], Rasmus Froberg Brøndum[1,2,3], Mette Thune Mouritzen[1,3,4], Heidi Søgaard Christensen[1,2,3], Martin Bøgsted[1,2,3], Ursula G Falkmer[1,3,4], Laurids Østergaard Poulsen[1,3,4]

Affiliations

[1] Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[2] Department of Haematology, Aalborg University Hospital, Aalborg, Denmark
[3] Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark
[4] Department of Oncology, Aalborg University Hospital, Aalborg, Denmark

Description

In this paper, we analysed the 30-day mortality following SACT using an adapted indicator, avoiding selection bias from arising by conditioning on future events present in similar studies. Conditioning on future events means that the inclusion criteria rely on data that are not available at the actual time of inclusion. Similar studies included patients conditioned on not getting subsequent SACT in a specific timeframe or before death. Our indicator considering each individual SACT circumvents this issue. We compared a large variety of malignancies over an 11-year period using both our adapted indicator and a reference indicator subject to selection bias.

1. Introduction

Systemic anticancer therapies (SACTs) often require lengthy drug administration procedures at hospitals and frequently induce severe side effects[1–4]. Patients with limited residual life expectancy may not benefit from the treatment and only experience the short-term side effects, thus, reducing the patients' quality of life[5]. SACT should be avoided in these cases[6].

To monitor the usage of SACT near the end of life, primarily two approaches have been used. One, proposed by Earle et al.[7], considers exclusively patients who die from cancer . While the criterion on the cause of death is not an issue for monitoring, it becomes a problem when calculating risk factors. Indeed, including only patients who died from cancer leads to a selection bias in the cohort definition by conditioning the inclusion on future events[8]. Conditioning on death from cancer will for example exclude long term survivors who died from other causes.

Another approach was proposed by Wallington et al.[9]. It suggests examining 30-day mortality from the start of the last SACT cycle in a calendar year. Their indicator, referred to as Wallington's indicator in the following, does not condition on death or its cause and, thus, allows for more prospective studies. As Wallington's indicator only considers the last SACT given within a chosen observation interval for each patient, there is a selection bias towards inclusion of later lines. This selection bias may thus lead to unreliable calculation of risk factors for use in a clinical context.

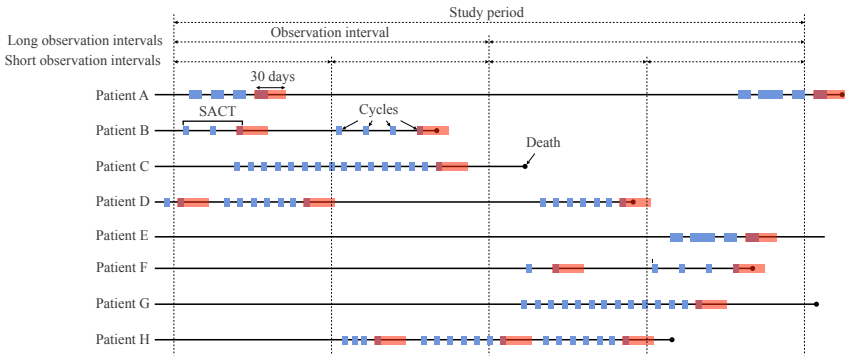
This study aimed to adapt the endpoint of Wallington's indicator to improve the clinical applicability. A second aim was to compare risk factors found with both indicators in the same dataset. The final aim was to obtain standard values for 30-day mortality following SACT for the improved indicator, over the period 2009–2019 for the most common solid cancers in the North Denmark Region.

2. Materials and methods

2.1. The improved indicator

SACT is defined as treatment including antineoplastic agents (i.e., Anatomical Therapeutic Chemical [ATC] classification[10] code L01). A cycle is defined as a set of drug prescriptions given on consecutive days. A SACT regimen is defined as a treatment based on the drugs used and the administration protocol. Consecutive cycles with the same regimen were grouped as one SACT, if the interval between two consecutive cycles was less than 60 days. SACTs were characterized using the regimen names, e.g., FOLFOX, to obtain their intent, palliative, or curative. The line number represents the number of palliative SACTs administered to the patient. Some regimens can be chosen with either a curative or palliative intent and were referred to as multi-intent regimens.

For each SACT, a dichotomous outcome is considered, describing whether the patient died within 30 days of the start of the last cycle of this SACT. Thus, the value for the improved indicator in a given observation interval is the average of the 30-day mortality outcomes for all SACTs that ended in this interval (see Figure 1).



Long observation intervals

Death within 30 days	1	2
Last SACTs	5	5
SACTs	7	7
Wallington's indicator	1/5 = 20%	2/5 = 40%
Improved indicator	1/7 = 14%	2 / 7 = 28%

Short observation intervals

Death within 30 days	0	1	1	1
Last SACTs	3	3	3	3
SACTs	4	3	4	3
Wallington's indicator	0/3 = 0%	1/3 = 33%	1/3 = 33%	1/3 = 33%
Improved indicator	0/4 = 0%	1/3 = 33%	1/4 = 25%	1/3 = 33%

Figure 1. *An example of a calculation of the improved indicator compared to Wallington's indicator using two different lengths for the observation interval on the same study period for 8 patients receiving 14 SACTs.*
The impact of the duration of the observation interval is illustrated in both cases using two interval lengths, the long observation interval length being twice the short observation interval one. The limits of the observation intervals are represented with vertical dashed lines. A SACT is considered for an observation interval if it ends in this interval. The last SACTs value represents the number of SACTs ended in an observation interval when considering for each patient only the last SACT. It is equal to the number of patients who ended a SACT in the considered observation interval. The difference between the values for the last SACTs and the SACTs illustrates the exclusion of some SACTs and therefore the selection bias.

2.2. Study design and participants

All patients from the North Denmark Region diagnosed with solid tumors before 31/12/2019 and alive after 01/01/2009 (N=29,937) were screened using the Patients Administrative System (PAS) from the North Denmark Region based on the diagnosis codes. Among these patients, 24,496 had one of the included malignancies (see Supplementary Table 1). In the period 2009–2019, 10,672 patients received SACTs without being referred to other regions. Among these patients, 459 were excluded due to their participation in clinical trials. The final cohort of 10,213 patients received 16,622 SACTs (see Supplementary figure 1).

The clinical data were extracted from the PAS, and the treatment data were obtained from the prescription software ARIA OIS for Medical Oncology v13.7 (Varian Medical Systems Inc., Palo Alto, CA, USA) (MedOnc). The PAS data consisted of all diagnoses and procedures coded according to the Danish Disease Classification System[11]. This classification system is similar to the ICD-10 classification for diagnoses. Dates of death were obtained from the Danish Civil Registration System (CPR). Data for each SACT

consisted of sex, age, comorbidities according to Charlson's Comorbidity Index[12] (CCI), current malignancy, treatment intent (curative or palliative), regimen, year at the start of treatment, line number, and death within 30 days of the start of the last cycle. The comorbidities were extracted from the diagnosis codes found in the PAS (see Supplementary Table 2) and updated at each SACT.

2.3. Statistical methods

The improved and Wallington's indicators were both calculated over the 11-year period per diagnosis and treatment intent as well as for all diagnoses per year and treatment intent. Wallington's indicator was calculated with an observation interval of one year, taking into consideration only the last cycle of the last SACT for each patient who ended a SACT in each interval. To estimate the effect of the observation interval, Wallington's indicator was also calculated with an observation interval of a quarter and compared to the yearly values. Additionally, the improved indicator for palliative treatments was calculated over the 11-year period per line number and per drug combination.

A multivariate logistic regression was performed for both indicators using period, age, sex, comorbidities, number of treatment lines, and type of malignancy as independent variables to identify potential risk factors. Death within 30 days of the start of the last cycle of either each SACT or last SACT in a given observation interval was used as the dependent variable for the improved and Wallington's indicators, respectively. The corresponding effect estimates are presented as odds ratios (ORs). A threshold of 0.05 was used to define the statistical significance of p-values, and 95% confidence intervals (CIs) were used for the ORs and survival estimates.

30-day mortality per diagnosis, line number, and regimen were also calculated, for which only SACTs given in first or second line were considered.

Data management and statistical analyses were performed using SAS Enterprise Guide 8.3 (SAS Institute Inc., Cary, NC, USA) and Python 3.8 in Jupyter notebooks[13]. The Python library statsmodel v0.11[14] was used for the regressions.

3. Results

3.1. Study population

The characteristics of the study population are presented in Table 1. The majority of the 10,213 patients included in this study were women (60%) due to the size of the female cancer cohorts (breast, ovarian, and uterine cancers, n = 3331).

Table 1. *Study population characteristics (overall and based on the cancer diagnosis) and the improved and Wallington's indicators per diagnosis and SACT intent.*

	Overall	Brain	Lung	Breast	Gastroesophageal	Pancreatic	Colorectal	Ovarian	Uterine	Prostate	Urinary
N	10 213	403	2 563	2 556	532	565	2 081	507	268	450	288
Males %	40	60	50	1	75	55	57	0	0	100	69
Mean age at diagnosis years (range)	64 (19–94)	60 (20–84)	68 (33–93)	57 (25–89)	65 (32–84)	67 (39–87)	66 (19–94)	67 (19–88)	64 (21–89)	71 (48–87)	68 (37–89)
Tx	16 622	907	4 030	3 979	922	809	3 506	1 108	395	580	386
Palliative Tx n (%)	10 006 (62)	403 (44)	3 505 (87)	1 614 (41)	490 (53)	712 (88)	1 855 (54)	464 (57)	117 (57)	567 (98)	279 (72)
Lines n (range)	1.7 (1–10)	1.7 (1–6)	1.6 (1–7)	2.3 (1–10)	1.5 (1–8)	1.4 (1–4)	2.0 (1–10)	1.9 (1–7)	1.4 (1–4)	1.3 (1–5)	1.4 (1–6)

N, number of patients; Males, percentage of male patients; Age, average age at diagnosis in years; N-y survival. %, the overall survival percentage from diagnosis for patients treated with SACTs in N years, as in 2, 5, and 10 years; Tx, total number of SACTs given; Palliative Tx, number of SACTs given with palliative intent; Lines, the number of palliative SACTs given to patients treated with at least one palliative SACT. Values between parentheses show the range for the 'Age' and 'Lines' columns, the 95% confidence interval for the survival columns, and the proportion in the percent of palliative SACTs given among treatments with known intent for the 'Palliative Tx' column.

Patients treated for advanced or metastatic disease received an average of 1.7 SACT lines. On average, prostate cancer patients received only 1.3 lines, while breast cancer patients were treated on average with 2.3 lines. For lung, pancreatic, and prostate cancer, patients were predominantly given palliative SACTs (87%, 88%, and 98%, respectively). In contrast, breast cancer patients mainly received curative SACTs (59%).

3.2. The improved indicator compared to Wallington's indicator

3.2.1. Per diagnosis and intent

As seen in Table 2, the 30-day mortality following SACT was higher for palliative SACTs than for curative SACTs across malignancies (10.3% vs 1.3% for the improved indicator, 13.1% vs 1.5% for Wallington's indicator).

Table 2. *The improved and Wallington's indicators per diagnosis and SACT intent.*

	Overall	Brain	Lung	Breast	Gastroesophageal	Pancreatic	Colorectal	Ovarian	Uterine	Prostate	Urinary
30-day mortality - Improved indicator											
For curative Tx % (ratio)	1.3 (75/ 5944)	5.5 (27/ 491)	1.2 (6/511)	0.2 (4/ 2306)	1.2 (5/426)	0.0 (0/94)	0.8 (12/ 1562)	5.4 (19/ 350)	0.0 (0/88)	7.7 (1/13)	1.0 (1/103)
For palliative Tx % (ratio)	10.3 (1004/ 9760)	8.6 (34/ 394)	14.3 (486/ 3410)	6.7 (106/ 1577)	15.0 (71/ 473)	14.8 (102/ 689)	6.7 (123/ 1832)	6.2 (28/ 450)	6.9 (8/116)	3.6 (20/ 550)	9.7 (26/ 269)
For multi-intent Tx % (ratio)	2.1% (11/ 532)	0.0% (0/1)	None	0.0% (0/12)	0.0% (0/2)	None	2.0% (1/49)	2.4% (7/288)	1.7% (3/179)	None	0.0% (0/1)
Overall % (ratio)	6.7 (1090/ 16236)	6.9 (61/ 886)	12.5 (492/ 3921)	2.8 (110/ 3895)	8.4 (76/ 901)	13.0 (102/ 783)	4.0 (136/ 3443)	5.0 (54/ 1088)	2.9 (11/ 383)	3.7 (21/ 563)	7.2 (27/ 373)
30-day mortality - Wallington's indicator											
For curative Tx % (ratio)	1.5 (74/ 5021)	8.8 (26/ 296)	1.3 (6/471)	0.2 (4/ 2103)	1.7 (5/296)	0.0 (0/83)	0.9 (12/ 1376)	8.8 (19/ 215)	0.0 (0/72)	9.1 (1/11)	1.0 (1/98)
For palliative Tx % (ratio)	13.1 (994/ 7595)	11.3 (34/ 302)	17.8 (481/ 2706)	9.4 (106/11 30)	18.9 (70/ 371)	17.8 (102/ 574)	9.0 (121/ 1349)	8.1 (26/ 322)	7.8 (8/102)	3.9 (20/ 510)	11.4 (26/ 229)
For multi-intent Tx % (ratio)	2.5% (11/ 437)	0.0% (0/1)	None	0.0% (0/12)	0.0% (0/2)	None	3.6% (1/28)	3.0% (7/236)	1.9% (3/157)	None	0.0% (0/1)
Overall % (ratio)	8.3 (1079/ 13053)	10.0 (60/ 599)	15.3 (487/ 3177)	3.4 (110/ 3245)	11.2 (75/ 669)	15.5 (102/ 657)	4.9 (134/ 2753)	6.7 (52/ 773)	3.3 (11/ 331)	4.0 (21/ 521)	8.2 (27/ 328)

The values shown for the improved and Wallington's indicators are in %. The 'Multi-intent' column contains the values for SACT regimens that can be used for both curative and palliative intents. The values between parentheses show the corresponding ratio. For the improved indicator, the numerator is the number of SACTs followed by the death of the patient within 30 days of the start of the last cycle, and the denominator is the total number of SACTs over the 11-year period. For Wallington's indicator, the denominator is the total number of patients who ended a treatment in a year, and the numerator is the number of these patients who died within 30 days of the start of their last cycle in the same year.

Considering all intents, there were large disparities between malignancies, ranging from below 3.5% for breast and uterine cancer SACTs (2.8% and 2.9% for the improved indicator, respectively and 3.4% and 3.3% for Wallington's indicator, respectively) to above 12% in 30-day mortality for lung and pancreatic cancer SACTs (12.5% and 13% for the improved indicator, respectively, and 15.3% and 15.5% for Wallington's indicator, respectively).

For palliative SACTs, the 30-day mortality was above 14% for lung, gastroesophageal, and pancreatic cancers (14.3%, 15%, and 14.8% for the improved indicator, respectively and 17.8%, 18.9%, and 17.8% for Wallington's indicator, respectively), while it was less than 4% for prostate cancer (3.6% for the improved indicator, 3.9% for Wallington's

indicator). For curative SACTs, the 30-day mortality was less than 2%, except for brain, ovarian, and prostatic cancers (5.5%, 5.4%, and 7.7% for the improved indicator, respectively and 8.8%, 8.8%, and 9.1% for Wallington's indicator, respectively).
Overall, the 30-day mortality with Wallington's indicator was consistently higher than with the improved indicator, especially for palliative SACTs given to gastroesophageal cancer patients (18.9% vs 15%).

3.2.2. Per year and treatment intent

Over time, the improved indicator showed an overall downward trend for the 30-day mortality from 8.0% in 2009 to 3.8% in 2019 (see Figure 2A).

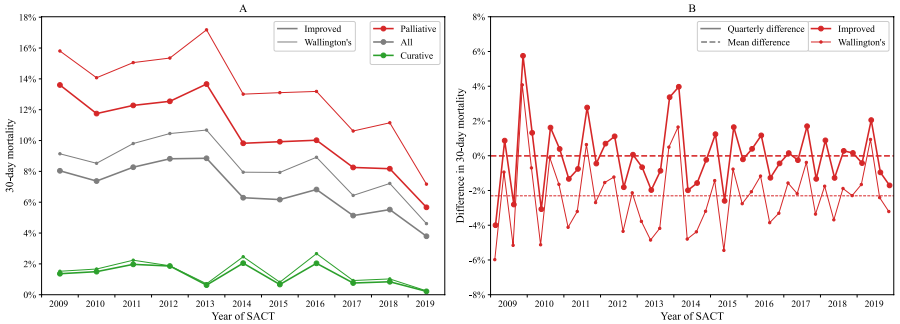


Figure 2. *30-day mortality per SACT intent and year (A) and the difference in 30-day mortality between the quarterly values and corresponding yearly values for palliative SACTs (B). The mean difference shows the mean of all differences between the quarterly and corresponding yearly values.*

This decline is notable for palliative SACTs, decreasing from 13.6% in 2009 to 5.7% in 2019, while the 30-day mortality following curative SACTs remained low over the study period. A similar pattern was seen for Wallington's indicator with a downward trend, especially for palliative SACTs, which ranged from 15.8% in 2009 to 7.2% in 2019.
The mean difference between the quarterly and yearly 30-day mortalities for palliative SACTs was below 0.1% and above 2% for the improved and Wallington's indicators, respectively (see Figure 2B).

3.3. Logistic regressions

Figure 3 shows the results of multivariate regressions for both the improved and Wallington's indicators.
No significant effect on 30-day mortality was found for comorbidities, sex, age group or line number in neither of the considered indicators. The period 2018–2019 is associated with a significant decrease in the 30-day mortality for both indicators compared to the period 2009–2011. A significant decrease was also found for the period 2015-2017 for the improved indicator. Lung, gastroesophageal, and pancreatic cancer diagnoses had a significantly worse 30-day mortality than breast cancer using the improved indicator. Inversely, prostate cancer had a significantly better 30-day mortality compared to breast cancer.
Overall, no major difference could be found in terms of risk factors between the improved and Wallington's indicators. A tentative difference was observed for the 75+

years age group, with a borderline significantly lower 30-day mortality for Wallington's indicator (OR: 0.84, CI: 0.70 to 1.01, p = 0.065), while it was far from significant for the adapted 30-day mortality indicator (OR: 0.93, CI: 0.77 to 1.11, p = 0.4).

Variable	Patient-years	Pal. SACTs	Wallington's	P-value	Improved	P-value
Sex						
Female	4 115	5 324	Used as reference		Used as reference	
Male	3 612	4 482	1.16 (0.99, 1.36)	0.075	1.14 (0.98, 1.34)	0.099
Year						
2009-2011	1 829	2 219	Used as reference		Used as reference	
2012-2014	2 051	2 585	1.18 (0.97, 1.43)	0.090	1.09 (0.90, 1.31)	0.368
2015-2017	2 251	2 908	0.89 (0.73, 1.09)	0.252	0.79 (0.65, 0.97)	**0.021**
2018-2019	1 585	2 083	0.64 (0.51, 0.81)	**<0.001**	0.56 (0.45, 0.71)	**<0.001**
Age						
18-44	251	342	0.81 (0.50, 1.31)	0.385	0.81 (0.51, 1.31)	0.393
45-59	1 383	1 864	1.00 (0.83, 1.21)	0.990	0.95 (0.79, 1.15)	0.612
60-74	4 365	5 556	Used as reference		Used as reference	
75+	1 728	2 044	0.84 (0.70, 1.01)	0.065	0.93 (0.77, 1.11)	0.407
Comorbidities						
Myocardial infarction	278	340	0.98 (0.68, 1.42)	0.917	1.03 (0.72, 1.47)	0.884
Peripheral vascular disease	381	465	1.10 (0.81, 1.49)	0.527	1.15 (0.86, 1.54)	0.358
CVA or TIA	532	669	0.95 (0.72, 1.26)	0.718	0.96 (0.73, 1.27)	0.788
COPD	609	722	1.08 (0.84, 1.38)	0.554	1.16 (0.91, 1.48)	0.242
Connective tissue disease	204	256	1.05 (0.69, 1.59)	0.833	1.02 (0.67, 1.53)	0.941
Peptic ulcer disease	305	370	0.88 (0.61, 1.28)	0.505	0.88 (0.61, 1.27)	0.498
Diabetes	232	286	1.19 (0.81, 1.75)	0.364	1.23 (0.84, 1.79)	0.288
CKD	119	142	1.09 (0.62, 1.91)	0.757	1.14 (0.66, 1.98)	0.641
Line number						
1st line	4 723	5 936	Used as reference		Used as reference	
2nd line	1 771	2 288	1.07 (0.91, 1.28)	0.410	1.06 (0.90, 1.26)	0.486
3+ line	1 233	1 582	0.98 (0.79, 1.21)	0.820	1.03 (0.83, 1.28)	0.773
Diagnosis group						
Breast	1 145	1 583	Used as reference		Used as reference	
Brain	305	394	1.24 (0.79, 1.96)	0.348	1.41 (0.90, 2.21)	0.136
Lung	2 716	3 446	2.10 (1.62, 2.72)	**<0.001**	2.26 (1.75, 2.92)	**<0.001**
Gastroesophageal	382	478	1.66 (1.12, 2.46)	**0.011**	1.89 (1.29, 2.78)	**0.001**
Pancreatic	575	690	2.20 (1.58, 3.06)	**<0.001**	2.59 (1.87, 3.59)	**<0.001**
Colorectal	1 416	1 830	0.85 (0.62, 1.16)	0.312	0.92 (0.68, 1.26)	0.615
Ovarian	349	451	1.00 (0.63, 1.58)	0.995	1.13 (0.72, 1.77)	0.596
Uterine	103	117	1.12 (0.54, 2.29)	0.767	1.40 (0.68, 2.86)	0.362
Prostatic	510	550	0.44 (0.26, 0.75)	**0.002**	0.58 (0.34, 0.99)	**0.045**
Urinary	226	267	1.35 (0.83, 2.20)	0.226	1.61 (0.99, 2.60)	0.055

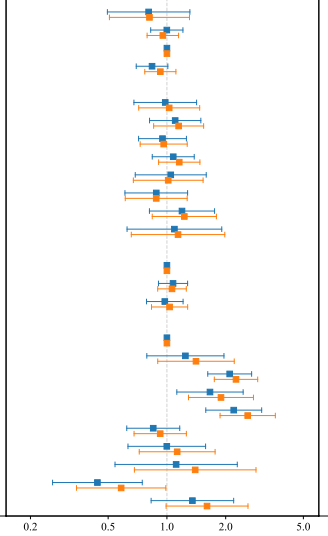


Figure 3. *Logistic regression results for palliative SACTs for the improved and Wallington's indicators. Year, year range at start of the SACT; Age, age range of the patient at start of the SACT. The comorbidities were defined as in Charlson's Comorbidity Index, and the PAS codes used for each comorbidity are available in Supplementary Table 2. Only comorbidities with a prevalence of >1% in the cohort were considered. Note that using the line number with Wallington's indicator is theoretically not appropriate but was included for comparison with the improved indicator.*

3.4. 30-day mortality following palliative SACTs per line number and drug combination

The 30-day mortality using the improved indicator, shown for specific line number in Figure 4A, does not reveal any clear shared pattern across malignancies.

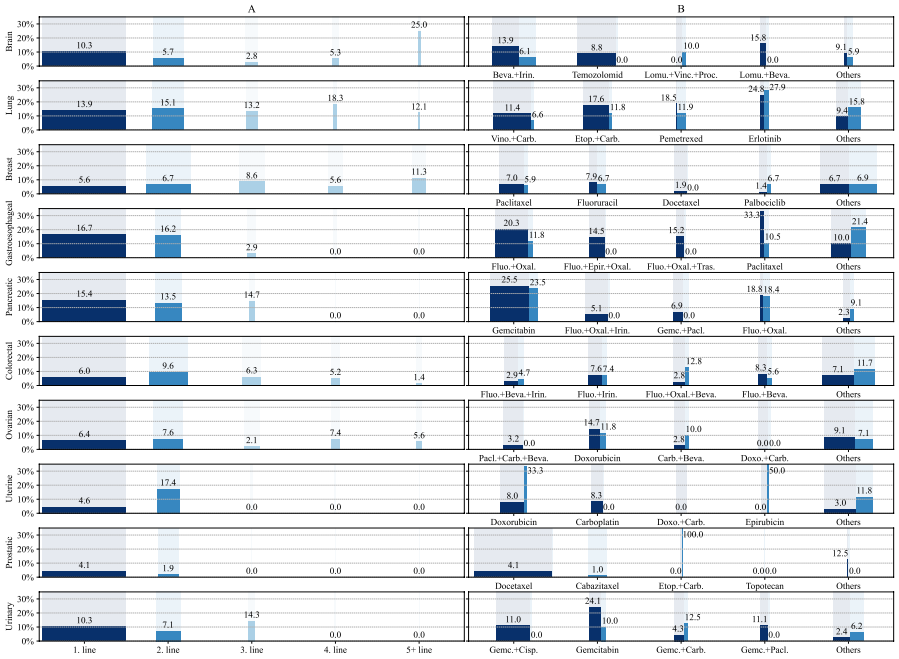


Figure 4. *30-day mortality following palliative SACTs per malignancy stratified by line number (A) and regimen (B). For each malignancy type, the width of the bar was proportional to the number of corresponding SACTs, normalized by the number of patients. For line number (A), the treatments after the 4th line were grouped in a '5+ line' category. For regimen (B), only the two first lines are included, and the corresponding top four regimens are displayed individually alongside other regimens grouped in the 'Others' category. Capecitabine is considered equivalent to fluorouracil and has thus been grouped with it. The same was true for panitumumab with cetuximab. Abbreviations: Beva.: bevacizumab, Carb.: carboplatin, Cisp.: cisplatin, Doxo.: doxorobucin, Epir.: epirubicin, Etop.: etoposide, Fluo.: fluorouracil/capecitabine, Gemc.: gemcitabine, Irin.: irinotecan, Lomu.: lomustin, Oxal.: oxaliplatin, Pacl.: paclitaxel, Proc.: procarbazin, Tras.: trastuzumab, Vinc.: vincristine, Vino.: vinorelbin. See Supplementary Table 3 and Supplementary Table 4 for the raw values.*

For example, for colorectal and uterine cancers, the 30-day mortality is lower in the first line than in the second line. Conversely, for brain, prostate, and urinary cancers, the 30-day mortality was lower in the second line than in the first line. For the remaining malignancies, the 30-day mortality remained mostly stable between the first and second lines.

Large differences in 30-day mortality were observed for the four most frequently administered regimens by diagnosis group and line number (Figure 4B). Patients who received gemcitabine monotherapy tended to have high 30-day mortality (25.5% in the first line and 23.5% in the second line for pancreatic cancers, and 24.1% in the first line for urinary cancers). The highest 30-day mortality (27.1%) was seen for erlotinib, given to lung cancer patients, with similar values for the first and second lines (24.8% and 27.9%, respectively).

4. Discussion

4.1. Main findings

We defined a quality indicator describing the 30-day mortality following the last cycle of SACT based on Wallington et al.'s approach. This indicator is adapted to the clinical context by avoiding selection bias and summarizes how often a SACT was followed by death within 30 days. Our proposed indicator allows for a more valid assessment of risk factors of the patients in a clinical context. However, limited differences were found between risk factors identified using the improved and Wallington's indicators for the present dataset.

Overall, we report a significant downward trend for the 30-day mortality following SACT using both indicators for palliative SACTs over an 11-year period. This decrease was not necessarily expected, despite the increased worldwide attention to close-to-death treatment of cancer patients. Recent advances in cancer treatment could have led to an increase in 30-day mortality. For example, in the case of protein kinase inhibitors, some patients benefit from continued treatment close to death[15,16]. This was illustrated in our study of erlotinib for lung cancer patients, which had the highest 30-day mortality. The fact that the 30-day mortality decreased over the period could be due to an increased attention of the clinicians towards earlier discontinuation of treatment.

Our study also found large differences in 30-day mortality between malignancies and between treatment intent. Unsurprisingly, treatments administered to patients with metastatic or advanced cancer had the highest 30-day mortality compared to treatments given as curative SACTs. The 30-day mortality following curative SACTs was 2% for some years, which we consider unacceptably high, but the values in recent years have been consistently low.

The groups with the highest 30-day mortality were those including patients with metastatic lung, gastroesophageal, and pancreatic cancers, all showing values above 14%. These values might partly be explained by widely spread tumor, several tumor-related symptoms, and poor performance status, notably among lung cancer patients[17].

Overall, the number of treatment lines did not seem to have a clear impact on 30-day mortality, with different patterns observed across malignancies. The 30-day mortality was expected to be higher in the second line than in the first line due to the progression of the disease. However, for brain, prostate, and urinary cancers, this was not the case. One explanation could be differences in toxicity profiles according to the type of malignancy and the line number. Another explanation could be that rapidly progressing disease may hinder the opportunity for second-line treatment, and only patients with less aggressively growing tumors are offered subsequent treatments.

Among the regimens, gemcitabine monotherapy and erlotinib monotherapy had the highest 30-day mortalities. As mentioned above, the high 30-day mortality for erlotinib could be explained by the maintained clinical benefit close to death. Gemcitabine is predominantly administered to frail patients with advanced urinary and pancreatic cancers[18], frailty that might not be taken well enough into consideration by clinicians, notably because this treatment might be the one and only option for these patients.

The considered comorbidities had no significant impact on 30-day mortality. This could be explained by their limited role in the 30-day mortality or by appropriate comorbidity adjustment in the clinical treatment decision making.

4.2. Critical Assessment

4.2.1. Study population

The main strengths of this study are the population-based design, coverage of all the major cancer groups, and extension over a wide timeframe with a high level of detail from a single-center setting. The single-center setting could also be considered as a limitation. However, due to the homogeneity of the healthcare system and treatment guidelines in Denmark, we expect that the conclusions can be extrapolated to the entire Danish cancer patient cohort. Nevertheless, a national study is required to confirm this assumption. Additionally, pooling regimens of different types, for example cytotoxic and targeted, as well as cancer types with significantly different prognoses, such as small cell and non-small cell lung cancers, may lead to results that are not representative of any of the regimens or subtypes. Investigating the 30-day mortality for individual regimen types, cancer subtypes, or rare malignancies (e.g., head-and-neck cancers and sarcomas) would require access to a larger cohort. This could be achieved by extending the study to the entire Danish cancer patient cohort as done with another indicator by Mattsson et al.[19].

4.2.2. Using healthcare data registries

In this study, the main data sources were electronic health records (EHRs) and administrative data, which we refer to as healthcare data registries (HDRs). Since such datasets are susceptible to biases like informed presence bias[20,21], we only considered actively followed patients, whose data are less prone to these biases.

One of the main advantages of using HDRs over quality databases is that they do not require additional reporting from clinicians. This makes it possible to build continuous quality monitoring tools. A disadvantage is the relative inaccessibility of some clinical parameters. For example, performance status, which is a known predictor of survival, is currently only recorded as text in patient journals.

Nevertheless, an increasing amount of healthcare data is currently being digitalized, and the quality of the stored data has been reported to be good, particularly in Denmark[22]. This should facilitate the development of HDR-based and clinically applicable quality indicators.

4.2.3. Using WHO ATC classification

An international consensus on the definition of a SACT is warranted since differences in the definition can significantly affect the results and impede the comparison of studies. We therefore use the WHO ATC classification as reference. We only included antineoplastic agents as defined by this classification, that is, drugs with an ATC code starting with L01. The endocrine therapies (ATC code starting with L02) were primarily excluded due to:

(1) their less severe toxicity profile
(2) oral administration, which implies fewer visits to the hospital, impeding the assessment of treatment compliance and impair the reporting in HDRs.

Furthermore, while these treatments are often included in studies following Earle et al.'s approach, they were also excluded in Wallington et al.'s study.

4.2.4. Characteristics of the indicator

The "30-day mortality" endpoint is a common endpoint in healthcare systems, notably in surgery. This endpoint was used by Wallington et al., and we thus decided to use this approach as a reference to define 30-day mortality. The main strength of our improved 30-day mortality indicator is that it can be used to evaluate risk factors for 30-day mortality following any SACT and can thus be used prospectively, i.e., to potentially adapt the quality of the treatment in the clinic. In contrast, risk factors calculated following Wallington et al.'s approach can only be used adequately for the last SACT, which is only known in hindsight. By considering every SACT, this indicator also allows us to investigate the effect of the line number and the type of treatment used on the risk of treating too close to death without conditioning on future events and thus avoids selection bias. It is nevertheless important to note that the risk factors are only usable once the SACT is started and thus cannot be used to decide to start a SACT or not. Instead, it is intended to help clinicians better assess the risk of early mortality to stop an already started treatment in time.

An additional benefit of the adapted 30-day mortality indicator is that it, in contrast to Wallington's indicator, remains unbiased across different choices of observation interval; for example, the 30-day mortality calculated for a quarter, or a month can be directly compared to the 30-day mortality calculated over a year or a decade (see Figure 2B).

4.3. Comparison with other studies

Older studies have reported an increase[23,24] in late chemotherapy administration in cancer patients. However, in line with recent studies[25,26], we report a decrease in 30-day mortality over time, notably for palliative SACTs.

Compared to Wallington et al.'s original results, we found similar results for breast cancer, with values of 0.2% compared to 0.3% for curative SACTs and 9.4% compared to 7.5% for palliative SACTs. For lung cancer, we found larger differences, with values for 30-day mortality for Wallington's indicator of 1.3% compared to 2.9% in Wallington et al.'s study for curative SACTs and 17.8% compared to 10.0% for palliative SACTs. This can be partially explained by the difference in the inclusion criteria. This could also be due to recent developments in the treatment of lung cancer patients, notably the introduction of protein kinase inhibitor treatments.

Concerning other studies, the differences in inclusion criteria and endpoint definition limit the comparability with our study. This could explain the large variability in the results reported[23,26] and illustrate the need for more standardized definitions, as proposed here.

4.4. Perspectives

The improved indicator can be used to properly identify risk factors for high 30-day mortalities, with the objective of potentially improving the quality of life near the end of life and better utilizing the available resources in the health care system. This indicator for 30-day mortality following SACT should ideally be more focused on specific cancer diagnoses and treatment regimens in order to define recommendations and potential

prognostic models to support the work of clinicians in daily clinical practice. More complex models allowing dynamic prediction and leveraging more extensive clinical data should also be built to help clinicians to decide when to stop an ongoing treatment.

4.5. Conclusions

We defined an improved quality indicator based on the approach followed by Wallington et al.[9] to evaluate the 30-day mortality following SACT. This indicator can be used to identify clinical risk factors for increased 30-day mortality and stays consistent across different choices of observation intervals. Using this indicator, we noted a significant downward trend in 30-day mortality following palliative SACT over an 11-year period. A multicenter study should be performed to define a more reliable benchmark for this improved indicator.

Acknowledgements

We would like to thank the System Administrator of MedOnc, Annette Juul Madsen, Dept. of Oncology, and Special Consultant Thomas Mulvad Larsen, Business Intelligence Unit, North Denmark Region, for their help in obtaining and understanding the data needed for this study.

Ethical approval and registration

According to Danish legislation, registry projects do not require patient consent and ethical approval, they must only be registered at the data responsible host institution. This study was part of a project registered at the Research Project Inventory of the North Denmark Region (reg. number 2019–41).

Funding sources

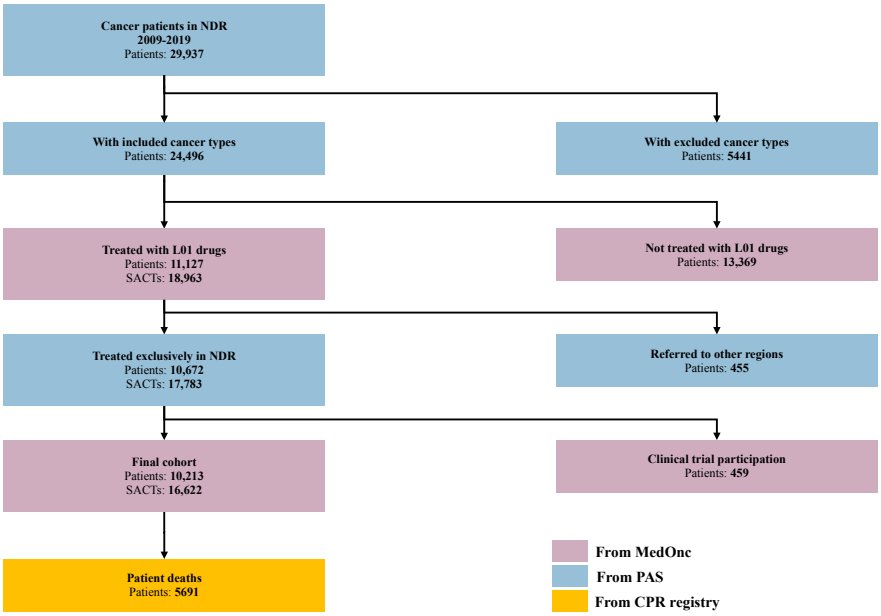
This work was supported by the Department of Oncology, Aalborg University Hospital and the Regional Research Fund of North Denmark Region. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
2. Taylor CW, Kirby AM. Cardiac Side-effects From Breast Cancer Radiotherapy. *Clin Oncol (Royal Coll Radiol*. 2015;27(11):621-629. doi:10.1016/j.clon.2015.06.007
3. Kappers MH, van Esch JHM, Sleijfer S, Danser AHJ, van den Meiracker AH. Cardiovascular and renal toxicity during angiogenesis inhibition: clinical and mechanistic aspects. *J Hypertens*. 2009;27(12):2297-2309. doi:10.1097/HJH.0b013e3283309b59
4. Pabla N, Dong Z. Curtailing side effects in chemotherapy: a tale of PKCδ in cisplatin treatment. *Oncotarget*. 2012;3(1):107-111. doi:10.18632/oncotarget.439
5. Davis C. Drugs, cancer and end-of-life care: a case study of pharmaceuticalization? *Soc Sci Med*. 2015;131:207-214. doi:10.1016/j.socscimed.2014.12.007
6. Earle CC, Neville BA, Landrum MB, et al. Evaluating claims-based indicators of the intensity of end-of-life cancer care. *Int J Qual Heal Care*. 2005;17(6):505-509. doi:10.1093/intqhc/mzi061
7. Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end-of-life cancer care from administrative data. *J Clin Oncol*. 2003;21(6):1133-1138. doi:10.1200/JCO.2003.03.059
8. Lund JL, Horváth-Puhó E, Szépligeti SK, et al. Conditioning on future exposure to define study cohorts can induce bias: The case of low-dose acetylsalicylic acid and risk of major bleeding. *Clin Epidemiol*. 2017;9:611-626. doi:10.2147/CLEP.S147175
9. Wallington M, Saxon EB, Bomb M, et al. 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study. *Lancet Oncol*. 2016;17(9):1203-1216. doi:10.1016/S1470-2045(16)30383-7
10. WHO Collaborating Centre for Drug Statistics Methodology. Anatomical Therapeutic Chemical (ATC) classification system. https://www.whocc.no/atc/structure_and_principles/. Accessed March 3, 2021.
11. Sundhedsdatastyrelsen. Disease Classification System - SKS (in Danish). https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer/klassifikation-sygdomme. Accessed March 3, 2021.
12. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
13. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. *Position Power Acad Publ Play Agents Agendas - Proc 20th Int Conf Electron Publ ELPUB 2016*. 2016:87-90. doi:10.3233/978-1-61499-649-1-87
14. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. *Proceeding 9th Python Sci Conf*. 2010;(Scipy):57, 61.
15. Yap TA, Macklin-Doherty A, Popat S. Continuing EGFR inhibition beyond progression in advanced non-small cell lung cancer. *Eur J Cancer*. 2017;70:12-21. doi:10.1016/j.ejca.2016.10.014

16. Smith TJ, Hanna N, Johnson D, et al. Case for stopping targeted therapy when lung cancer progresses on treatment in hospice- eligible patients. *J Oncol Pract*. 2017;13(12):780-783. doi:10.1200/JOP.2017.027367
17. Iyer S, Roughley A, Rider A, Taylor-Stokes G. The symptom burden of non-small cell lung cancer in the USA: A real-world cross-sectional study. *Support Care Cancer*. 2014;22(1):181-187. doi:10.1007/s00520-013-1959-4
18. Skau Rasmussen L, Vittrup B, Ladekarl M, et al. The effect of postoperative gemcitabine on overall survival in patients with resected pancreatic cancer: A nationwide population-based Danish register study. *Acta Oncol (Madr)*. 2019;58(6):864-871. doi:10.1080/0284186X.2019.1581374
19. Mattsson TO, Pottegård A, Jørgensen TL, Green A, Bliddal M. End-of-life anticancer treatment – a nationwide registry-based study of trends in the use of chemo-, endocrine, immune-, and targeted therapies. *Acta Oncol (Madr)*. 2021;0(0):1-7. doi:10.1080/0284186x.2021.1890332
20. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res*. 2018;20(5):e185. doi:10.2196/jmir.9134
21. Phelan M, Bhavsar NA, Goldstein BA. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference. *eGEMs*. 2017;5(1):22. doi:10.5334/egems.243
22. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National patient registry: A review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-490. doi:10.2147/CLEP.S91125
23. Ho TH, Barbera L, Saskin R, Lu H, Neville BA, Earle CC. Trends in the aggressiveness of end-of-life cancer care in the universal health care system of Ontario, Canada. *J Clin Oncol*. 2011;29(12):1587-1591. doi:10.1200/JCO.2010.31.9897
24. Earle CC, Neville BA, Landrum MB, Ayanian JZ, Block SD, Weeks JC. Trends in the aggressiveness of cancer care near the end of life. *J Clin Oncol*. 2004;22(2):315-321. doi:10.1200/JCO.2004.08.136
25. Khoja L, McGurk A, O'Hara C, Chow S, Hasan J. Mortality within 30 days following systemic anti-cancer therapy, a review of all cases over a 4 year period in a tertiary cancer centre. *Eur J Cancer*. 2015;51(2):233-240. doi:10.1016/j.ejca.2014.11.011
26. Gibson AJW, Li H, D'Silva A, et al. Factors associated with early mortality in non-small cell lung cancer patients following systemic anti-cancer therapy: A 10 year population-based study. *Lung Cancer*. 2019;134(February):141-146. doi:10.1016/j.lungcan.2019.06.003

Supplementary material



Supplementary figure 1. *Flow chart of patient inclusion in the study and the corresponding SACTs SACTs, the number of SACTs given to the corresponding patients; PAS, the Danish Patients Administrative System; CPR registry, the Danish Civil Registration System registry; L01 drugs, antineoplastic agents as defined by the Anatomical Therapeutic Chemical classification. "Treated with L01 drugs" refers to patients treated exclusively in the North Denmark Region with L01 drugs in the study period.*

Supplementary Table 1. *Malignancy grouping, inclusion status, and exclusion criteria*

ICD10	Malignancies	Inclusion	Reason for exclusion
C00-C14, C30-C33	Head and neck	No	Relatively rare and not primarily treated with SACT
C15-C16	Gastroesophageal	Yes	
C17	Intestine	No	Rare
C18-C20	Colorectal	Yes	
C21	Anal	No	Rare
C22-C24	Hepato-biliary	No	Referred to other hospitals
C25	Pancreas	Yes	
C26, C39, C57, C76, C80	Ill-defined	No	Ill-defined
C34	Lung	Yes	
C37-C38	Thoracic other than lung	No	Rare
C40-C41	Bone and articular cartilage	No	Rare
C43	Melanoma	No	Referred to other hospitals
C44	Skin other than melanoma	No	Not treated with SACT
C45-C49	Connective and soft tissue	No	Rare
C50	Breast	Yes	
C51-C52	Vulva and vagina	No	Rare
C53-C55	Uterine	Yes	
C56	Ovarian	Yes	
C58	Placenta	No	Rare
C61	Prostate	Yes	
C64-C68	Urinary	Yes	
C69-C70	Eye and meninges	No	Rare
C71	Brain	Yes	
C72	Nervous system	No	Rare
C73-C75	Endocrine	No	Rare
C77-C79	Secondary	No	Not primary
C81-C96	Haematological	No	Not solid tumour

Supplementary Table 2. *SKS codes for comorbidities*

Comorbidity	SKS codes
Myocardial infarction	DI21, DI22, DI23
Congestive heart failure (CHF)	DI50, DI110, DI130, DI132
Peripheral vascular disease	DI70, DI71, DI72, DI73, DI74, DI77
Cerebrovascular accident (CVA) or transient ischemic attack (TIA)	DI60, DI61, DI62, DI63, DI64, DI65, DI66, DI67, DI68, DI69, DG45, DG46
Dementia	DF00, DF01, DF02, DF03, DF051, DG30
Chronic obstructive pulmonary disease (COPD)	DJ40, DJ41, DJ42, DJ43, DJ44, DJ45, DJ46, DJ47, DJ60, DJ61, DJ62, DJ63, DJ64, DJ65, DJ66, DJ67, DJ684, DJ701, DJ703, DJ841, DJ920, DJ961, DJ982, DJ983
Connective tissue disease	DM05, DM06, DM08, DM09, DM30, DM31, DM32, DM33, DM34, DM35, DM36, DD86
Peptic ulcer disease	DK221, DK25, DK26, DK27, DK28
Liver disease – Mild	DB18 K700, DK701, DK702, DK703, DK71, DK73, DK74, DK760
Liver disease - Moderate to severe	DB150, DB160, DB162, DB190, DK704, DK72, DK766, DI85
Diabetes mellitus - Uncomplicated	DE100, DE101, DE109, DE110, DE111, DE119
Diabetes mellitus - End-organ damage	DE102, DE103, DE104, DE105, DE106, DE107, DE108
Hemiplegia	DG81, DG82
Moderate to severe chronic kidney disease	DI12, DI13, DN00, DN01, DN02, DN03, DN04, DN05, DN07, DN11, DN14, DN17, DN18, DN19, DQ61
Malignancy - Localized solid tumor	DC00-DC75 not finishing with a 'M', except DC44 (non-melanoma skin cancer)
Malignancy - Metastatic solid tumor	DC76, DC77, DC78, DC79, DC80 or DC00-DC75 finishing with a 'M'
Malignancy – Leukemia	DC91, DC93, DC93, DC94, DC95
Malignancy - Lymphoma	DC81, DC82, DC83, DC84, DC85, DC88, DC90, DC96
AIDS	DB21, DB22, DB23, DB24

Supplementary Table 3. *30-day mortality following palliative SACTs per line number and diagnosis group*

	Line 1	Line 2	Line 3	Line 4	Line 5+
Brain	25/243 (10.3)	5/88 (5.7)	1/36 (2.8)	1/19 (5.3)	2/8 (25.0)
Lung	302/2166 (13.9)	122/808 (15.1)	41/310 (13.2)	17/93 (18.3)	4/33 (12.1)
Breast	41/727 (5.6)	26/389 (6.7)	19/220 (8.6)	7/126 (5.6)	13/115 (11.3)
Gastroesophageal	54/323 (16.7)	16/99 (16.2)	1/34 (2.9)	0/13 (0.0)	0/4 (0.0)
Pancreatic	76/494 (15.4)	21/156 (13.5)	5/34 (14.7)	0/5 (0.0)	None
Colorectal	59/985 (6.0)	44/459 (9.6)	14/222 (6.3)	5/97 (5.2)	1/69 (1.4)
Ovarian	17/265 (6.4)	7/92 (7.6)	1/48 (2.1)	2/27 (7.4)	1/18 (5.6)
Uterine	4/87 (4.6)	4/23 (17.4)	0/3 (0.0)	0/2 (0.0)	0/1 (0.0)
Prostate	18/434 (4.1)	2/107 (1.9)	0/5 (0.0)	0/3 (0.0)	0/1 (0.0)
Urinary	20/194 (10.3)	4/56 (7.1)	2/14 (14.3)	0/4 (0.0)	0/1 (0.0)

The numerator is the number of SACTs followed by the death of the patient within 30 days of the last administration. The denominator is the total number of SACTs given. The value between parentheses is the corresponding 30-day mortality in percentage.

Supplementary Table 4. *30-day mortality following palliative SACTs with the top four regimens of each diagnosis group in the 1st and 2nd lines*

Brain	Bevacizumab + Irinotecan	Temozolomide	Lomu. + Vinc. + Proc.	Lomustine + Bevacizumab	Others
1. line	11/79 (13.9)	10/114 (8.8)	0/20 (0.0)	3/19 (15.8)	1/11 (9.1)
2. line	3/49 (6.1)	0/4 (0.0)	1/10 (10.0)	0/8 (0.0)	1/17 (5.9)
Lung	Vinorelbine + Carboplatin	Etoposide + Carboplatin	Pemetrexed	Erlotinib	Others
1. line	116/1014 (11.4)	124/703 (17.6)	5/27 (18.5)	28/113 (24.8)	29/309 (9.4)
2. line	5/76 (6.6)	8/68 (11.8)	28/236 (11.9)	31/111 (27.9)	50/317 (15.8)
Breast	Paclitaxel	Fluorouracil	Docetaxel	Palbociclib	Others
1. line	15/215 (7.0)	6/76 (7.9)	2/107 (1.9)	1/74 (1.4)	17/255 (6.7)
2. line	2/34 (5.9)	5/75 (6.7)	0/5 (0.0)	2/30 (6.7)	17/245 (6.9)
Gastroesophageal	Fluorouracil + Oxaliplatin	Fluo. + Epir. + Oxal.	Fluo. + Oxal. + Tras.	Paclitaxel	Others
1. line	27/133 (20.3)	9/62 (14.5)	5/33 (15.2)	5/15 (33.3)	8/80 (10.0)
2. line	2/17 (11.8)	0/3 (0.0)	0/4 (0.0)	2/19 (10.5)	12/56 (21.4)
Pancreatic	Gemcitabine	Fluo. + Oxal. + Irin.	Gemcitabine + Paclitaxel	Fluorouracil + Oxaliplatin	Others
1. line	61/239 (25.5)	7/138 (5.1)	4/58 (6.9)	3/16 (18.8)	1/43 (2.3)
2. line	12/51 (23.5)	0/10 (0.0)	0/35 (0.0)	7/38 (18.4)	2/22 (9.1)
Colorectal	Fluo. + Beva. + Irin.	Fluorouracil + Irinotecan	Fluo. + Oxal. + Beva.	Fluorouracil + Bevacizumab	Others
1. line	5/170 (2.9)	13/170 (7.6)	4/144 (2.8)	9/108 (8.3)	28/393 (7.1)
2. line	3/64 (4.7)	4/54 (7.4)	5/39 (12.8)	3/54 (5.6)	29/248 (11.7)
Ovarian	Pacl. + Carb. + Beva.	Doxorubicin	Carboplatin + Bevacizumab	Doxorubicin + Carboplatin	Others
1. line	2/63 (3.2)	5/34 (14.7)	1/36 (2.8)	0/33 (0.0)	9/99 (9.1)
2. line	0/2 (0.0)	2/17 (11.8)	1/10 (10.0)	0/7 (0.0)	4/56 (7.1)
Uterine	Doxorubicin	Carboplatin	Doxorubicin + Carboplatin	Epirubicin	Others
1. line	2/25 (8.0)	1/12 (8.3)	0/11 (0.0)	0/6 (0.0)	1/33 (3.0)
2. line	1/3 (33.3)	0/1 (0.0)	None	1/2 (50.0)	2/17 (11.8)
Prostatic	Docetaxel	Cabazitaxel	Etoposide + Carboplatin	Topotecan	Others
1. line	17/415 (4.1)	None	0/10 (0.0)	0/1 (0.0)	1/8 (12.5)
2. line	0/2 (0.0)	1/96 (1.0)	1/1 (100.0)	0/2 (0.0)	0/6 (0.0)
Urinary	Gemcitabine + Cisplatin	Gemcitabine	Gemcitabine + Carboplatin	Gemcitabine + Paclitaxel	Others
1. line	9/82 (11.0)	7/29 (24.1)	1/23 (4.3)	2/18 (11.1)	1/42 (2.4)
2. line	0/3 (0.0)	1/10 (10.0)	1/8 (12.5)	0/3 (0.0)	2/32 (6.2)

The numerator is the number of SACTs followed by the death of the patient within 30 days of the last administration. The denominator is the total number of SACTs given. The value between parentheses is the corresponding 30-day mortality in percentage.

Paper II: High validity of the Danish National Patient Registry for systemic anticancer treatment registration from 2009 to 2019

Published in Clinical Epidemiology 2021:13 1085–1094

Authors

Charles Vesteghem[1,2,3], Rasmus Froberg Brøndum[1,2,3], Ursula G. Falkmer[1,3,4], Anton Pottegård[5], Laurids Østergaard Poulsen[1,3,4], Martin Bøgsted[1,2,3]

Affiliations

[1] Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[2] Department of Haematology, Aalborg University Hospital, Aalborg, Denmark
[3] Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark
[4] Department of Oncology, Aalborg University Hospital, Aalborg, Denmark
[5] Department of Public Health, University of Southern Denmark, Odense, Denmark

Description

In this paper, we evaluated the validity of the registration of the SACT procedures in the Danish National Patient Registry based on data from the prescription software MedOnc. The MedOnc dataset was used as the gold standard, as it is a clinical tool, whereas the Danish National Patient Registry is primarily used for administrative purposes. The aim was to assess whether the DNPR can be used for research and quality monitoring.

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

High Validity of the Danish National Patient Registry for Systemic Anticancer Treatment Registration from 2009 to 2019

Charles Vesteghem[1–3]
Rasmus Froberg Brøndum[1–3]
Ursula G Falkmer[1,3,4]
Anton Pottegård[5]
Laurids Østergaard Poulsen[1,3,4]
Martin Bøgsted[1–3]

[1]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; [2]Department of Hematology, Aalborg University Hospital, Aalborg, Denmark; [3]Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark; [4]Department of Oncology, Aalborg University Hospital, Aalborg, Denmark; [5]Department of Public Health, University of Southern Denmark, Odense, Denmark

Background: The Danish National Patient Registry is a major resource for Danish epidemiology. Only a few studies have been conducted to check the validity of the reporting of systemic anticancer treatments. In this study, we assessed this validity for a range of cancer types over a long period of time.

Patients and Methods: We extracted systemic anticancer treatment procedures from the Danish National Patient Registry for patients with solid malignant tumors treated at the Department of Oncology at Aalborg University Hospital between 2009 and 2019 (12,014 patients with 215,293 drug records). These data were compared to records obtained from the antineoplastic prescription database used at the department. We estimated the sensitivity, positive predictive value (PPV), and F1-score defined as the harmonic mean of the sensitivity and the PPV.

Results: There was an overall high concordance between the two datasets with a sensitivity and a PPV >92%. Treatments for brain, ovarian and endometrial cancers displayed lower concordance (81–89%). The validity was stable over the study period, with a slight drop during 2016–2017. Most drugs had a high validity with F1-scores above 90%. Fluorouracil, gemcitabine, pemetrexed, pembrolizumab, and nivolumab had F1-scores above 97%. Drugs that were introduced in the study period, such as lapatinib, palbociclib, erlotinib, pertuzumab, and panitumumab, yielded lower F1-scores due to the absence of specific registry codes early after introduction.

Conclusion: The Danish National Patient Registry can be used to reliably obtain information about systemic anticancer treatments, keeping in mind limitations for recently introduced drugs and for some types of cancer.

Keywords: antineoplastic agents, registries, Danish National Patient Registry, epidemiology, sensitivity and specificity, validity

Background

Nordic countries have extensive nationwide healthcare registries.[1] These registries are notably used for epidemiological studies.[2] One of the main data sources used to conduct these studies is the Danish National Patient Registry (DNPR) which has been shown to have a high validity for cancer diagnoses.[3] While most of these studies use the diagnoses recorded in the DNPR to analyze patients' trajectories,[4,5] other types of data are available, such as treatment procedure codes. It is of special interest in oncology to study for example the real-world efficacy of systemic anticancer treatments.[6] However, one of the main concerns of studies using the

Correspondence: Charles Vesteghem
Department of Clinical Medicine, Aalborg University, Søndre Skovvej 15, Aalborg, 9000, Denmark
Tel +45 97 66 38 72
Fax +45 97 66 63 23
Email charles.vesteghem@rn.dk

Received: 13 August 2021
Accepted: 28 October 2021
Published: 24 November 2021

DNPR data is the validity of the registration. Some work has already been published to address this concern for these treatments,[7,8] reporting high validity in terms of positive predictive value and sensitivity, but these studies were focused on colorectal cancers and included less than 500 patients. Thus, it remains unknown whether this high validity could be extrapolated to other solid malignant tumor types.

The aim of this study was to investigate the validity, using the same metrics, of systemic anticancer treatment procedure registration over a wide range of solid malignancies and over a long period of time.

Patients and Methods

A retrospective cohort study was conducted on patients with solid malignant tumors treated in the North Denmark Region.

Data Sources

The DNPR is encoded using the Danish Health Care Classification System (SKS)[9] and was used to obtain primary diagnoses and procedure information for both in- and outpatients containing the patient identifier, the admission and discharge dates, and the diagnosis or procedure code. For category-level diagnoses, the SKS encoding is identical to the ICD-10 classification.[10]

The second main data source was the database from the ARIA OIS for Medical Oncology v13.7 prescription software[11] (MedOnc) used at the Department of Oncology, Aalborg University Hospital. The corresponding data include the patient identifier, the start of treatment date, the duration, the drug name, and the dose given for each prescription and are only available for patients treated in the Region North Denmark. The MedOnc dataset was used as the gold standard to evaluate the validity of the DNPR dataset.

Data Extraction

Our focus is on anti-neoplastic agents as defined by the Anatomical Therapeutic Chemical (ATC) classification,[12] ie, drugs with an ATC code starting with "L01". These drugs are referred to here as L01 drugs. The corresponding data were extracted from the DNPR using SKS codes looking at the procedures: "Special medical treatments and treatment principles" (codes starting with "BWH") and "Treatment with antibodies and immunomodulatory therapy" (codes starting with "BOHJ"). These procedures were mapped to ATC codes. Procedures corresponding to drug combinations, ie, multiple ATC codes, in the DNPR data were split into individual drug entries. Drugs administered over consecutive days were grouped into one drug entry with a duration equal to the number of consecutive days. These drug entries are referred to here as drug cycles (see Figure 1).

For MedOnc, the drug names were mapped to ATC codes. The MedOnc prescriptions with no dose given, corresponding to non-administered treatments, were removed from the dataset. The drug entries were grouped in drug cycles, where applicable, in a similar manner to the DNPR dataset.

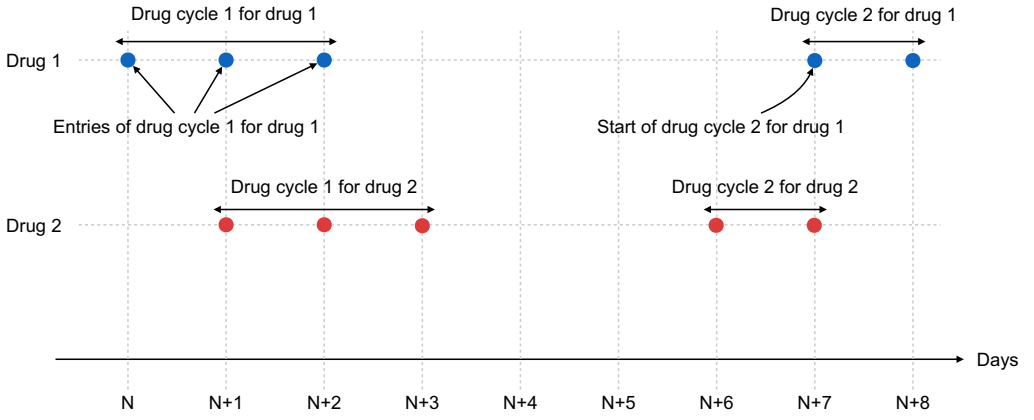


Figure 1 Grouping of drug entries into drug cycles.

Inclusion Criteria

The patients included in this study were identified using the cancer diagnosis codes (ICD-10 codes starting with C) found in the DNPR data as primary diagnosis. The diagnoses were grouped into common cancer types (see Supplementary Table 1). Only patients with a listed cancer type and at least one L01 drug cycle record in either the DNPR or MedOnc were included (see Supplementary Figure 1).

For the DNPR, we considered only L01 drug cycles from procedures performed at the Department of Oncology, Aalborg University Hospital between 2009 and 2019 (11 years). These data cover all systemic anticancer treatments given in the North Denmark Region. For MedOnc, we similarly only considered L01 drug cycles given over the same period.

In Denmark, each citizen is assigned an ID number from the Danish Civil Registration System.[13] The data sets were pseudonymized and linked at the patient level using an encoded version of this number.

Analysis

The comparisons of the two datasets were performed both for patients and for L01 drug cycles. For the patients, matching was performed using the patient identifier and the analyses were stratified by diagnosis. For L01 drug cycles, the ATC code and the start of treatment date were additionally considered for matching and the analyses were stratified by diagnosis, year, and drug.

Following an approach similar to Broe et al[8] the concordance of the datasets was measured using the positive predictive value (PPV) and the sensitivity. The MedOnc data were the gold standard, and the DNPR dataset was the predictive dataset. PPV was defined as the ratio of drug cycles in the intersection between both datasets and in the DNPR dataset, and the sensitivity was defined as the ratio of drug cycles in the intersection between both datasets and in the MedOnc dataset. Additionally, the F_1 score, defined as the harmonic mean of the PPV and sensitivity, was also used as an overall metric for concordance. As a sensitivity analysis, we considered a margin of 1 day for matching on the start date, as used by Broe et al.[8]

The data management and statistical analyses were performed using SAS Enterprise Guide 8.3 (SAS Institute Inc., Cary, NC, USA) and Python 3.8 in Jupyter notebooks,[14] respectively.

Table 1 Study Population Characteristics

Category	Variable	Count	Ratio
Overall	Patients	12,155	100%
Sex	Male	5113	42%
	Female	7042	58%
Age at diagnosis	18–44	878	7%
	45–59	3617	30%
	60–74	6089	50%
	75+	1571	13%
Cancer Diagnosis	Brain	462	4%
	Lung	2621	22%
	Breast	2968	24%
	Gastro-esophageal	620	5%
	Pancreatic	573	5%
	Colorectal	2306	19%
	Ovarian	557	5%
	Endometrial	226	3%
	Prostatic	514	4%
	Urinary	284	2%
	Other	1024	7%

Ethical Approval and Study Registration

According to Danish legislation, ethical approval and patient consent for purely registry-based projects is not required, only registration at the data responsible host institution is needed. The study protocol was registered in the North Denmark Region's research project inventory under the number 2019–41 and thereby complies with relevant data protection and privacy regulations.

Results

Study Population

This study included patients with a broad range of solid malignant tumors, the largest groups being lung, breast, and colorectal cancers, representing two-thirds of the cohort (see Table 1). Female patients accounted for the majority of the patients (58%). Ninety-three percent of the patients were >45 years old at diagnosis.

Matching Patients and Drug Cycles

Almost all patients are present in the intersection between MedOnc and the DNPR, which translates into a large concordance between the two datasets at the patient level, with a PPV and a sensitivity of 98.8% and 98.4%, respectively (see Table 2). However, the matching of brain tumor patients led to a lower sensitivity of 90%.

Table 2 PPV, Sensitivity, and F1-Score for Patients and L01 Drug Cycles per Diagnosis

Cancer Diagnosis	Type	MedOnc	DNPR	Intersection	PPV	Sensitivity	F1-Score
Overall	Patients	12,014	11,965	11,824	98.8%	98.4%	98.6%
	Drug cycles	215,293	216,074	198,888	92.0%	92.4%	92.2%
	With a 1-day margin	215,293	216,074	200,301	92.7%	93.0%	92.9%
Brain	Patients	440	419	397	94.7%	90.2%	92.4%
	Drug cycles	6671	6804	5546	81.5%	83.1%	82.3%
	With a 1-day margin	6671	6804	5610	82.5%	84.1%	83.3%
Lung	Patients	2613	2571	2563	99.7%	98.1%	98.9%
	Drug cycles	34,628	33,240	31,774	95.6%	91.8%	93.6%
	With a 1-day margin	34,628	33,240	32,066	96.5%	92.6%	94.5%
Breast	Patients	2937	2953	2922	99.0%	99.5%	99.2%
	Drug cycles	62,637	63,644	57,198	89.9%	91.3%	90.6%
	With a 1-day margin	62,637	63,644	57,582	90.5%	91.9%	91.2%
Gastro-oesophageal	Patients	619	611	610	99.8%	98.5%	99.2%
	Drug cycles	10,811	10,558	9863	93.4%	91.2%	92.3%
	With a 1-day margin	10,811	10,558	9998	94.7%	92.5%	93.6%
Pancreatic	Patients	573	565	565	100.0%	98.6%	99.3%
	Drug cycles	11,255	11,086	10,693	96.5%	95.0%	95.7%
	With a 1-day margin	11,255	11,086	10,726	96.8%	95.3%	96.0%
Colorectal	Patients	2255	2296	2245	97.8%	99.6%	98.7%
	Drug cycles	55,580	56,928	53,434	93.9%	96.1%	95.0%
	With a 1-day margin	55,580	56,928	53,688	94.3%	96.6%	95.4%
Ovarian	Patients	554	544	541	99.4%	97.7%	98.5%
	Drug cycles	11,901	12,252	10,586	86.4%	89.0%	87.7%
	With a 1-day margin	11,901	12,252	10,681	87.2%	89.7%	88.4%
Endometrial	Patients	225	222	221	99.5%	998.2%	98.8%
	Drug cycles	2800	3003	2511	83.6%	89.7%	86.5%
	With a 1-day margin	2800	3003	2527	84.1%	90.2%	87.1%
Prostatic	Patients	509	512	507	99.0%	99.6%	99.3%
	Drug cycles	3766	3861	3589	93.0%	95.3%	94.1%
	With a 1-day margin	3766	3861	3600	93.2%	95.6%	94.4%
Urinary	Patients	283	275	274	99.6%	96.8%	98.2%
	Drug cycles	3562	3541	3331	94.1%	93.5%	93.8%
	With a 1-day margin	3562	3541	3344	94.4%	93.9%	94.2%
Other	Patients	1006	997	979	98.2%	97.3%	97.8%
	Drug cycles	11,682	11,157	10,363	92.9%	88.7%	90.7%
	With a 1-day margin	11,682	11,157	10,479	93.9%	89.7%	91.8%

Notes: Patients are matched on encrypted CPR number, drug cycles on start date and ATC code. The 1-day margin is on the start date for drug cycles allowing additional matching if the start dates of unmatched drug cycles in MedOnc and the DNPR are 1 day or less from each other.

Matching the drug cycles using the patient identifier, the ATC code, and the start of treatment date generated a PPV and a sensitivity above 92%. Treatments within all diagnoses except brain, ovarian, and endometrial cancers have a sensitivity and a PPV above 89%, with treatments for pancreatic cancer above 95% (see Figure 2). Adding a 1-day margin for the start date improves the performance with a gain of 0.7% for PPV, 0.6% for sensitivity and 0.7% for F1-score.

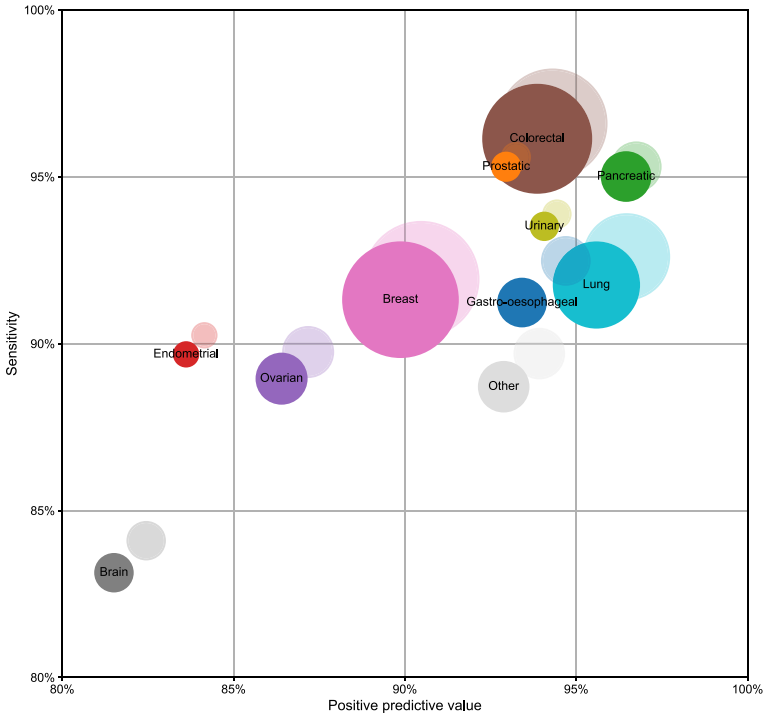


Figure 2 Positive predictive value vs sensitivity for the matching of drug cycles per cancer diagnosis. The area of the circle is proportional to the number of corresponding drug cycles. The lighter circles in the background correspond to the performances with a 1-day margin.

Evolution Over Time

The validity of the registered drug cycles is mostly stable over the 2009–2019 period (11 years) (see Figure 3). Nevertheless, a drop in PPV can be seen for 2016 and 2017. The sensitivity was also negatively impacted in 2012 and 2016. The effect of the 1-day margin, shown as lighter surfaces above both lines in Figure 3, seems to be stable over the period.

Validity per Drug

Looking at the most frequently administered drugs there is a more detailed picture, with most drugs having F1-scores above 90% (see Table 3). Some drugs (fluorouracil, gemcitabine, pemetrexed, pembrolizumab, and nivolumab) have high validity with F1-scores above 97%, while others (temozolomide, pertuzumab, palbociclib, erlotinib and lapatinib) have F1-scores below 80%. The low validity is typically due to a low sensitivity with values below 70%, ie, many entries in MedOnc cannot be matched with corresponding data in the DNPR (see Figure 4). As shown in Table 3, there is a strong correlation between drugs and diagnoses, for example temozolomide and cyclophosphamide are almost exclusively used for brain and breast cancer, respectively.

Discussion

Main Results

The DNPR data can be used as a good proxy for L01 drug cycles when matching the ATC code and start of treatment date. The reporting of drug cycles appears to be reliable across diagnoses, especially for colorectal and pancreatic cancers, but historically not for brain cancers, even though improvements have occurred. Looking at specific drugs, only a few have limited validity among frequently used drugs, including temozolomide.

Using the Start of Treatment Date Only

The duration of the cycle was not considered because the DNPR does not contain this information. However, in the context of a specific treatment for a specific cancer type,

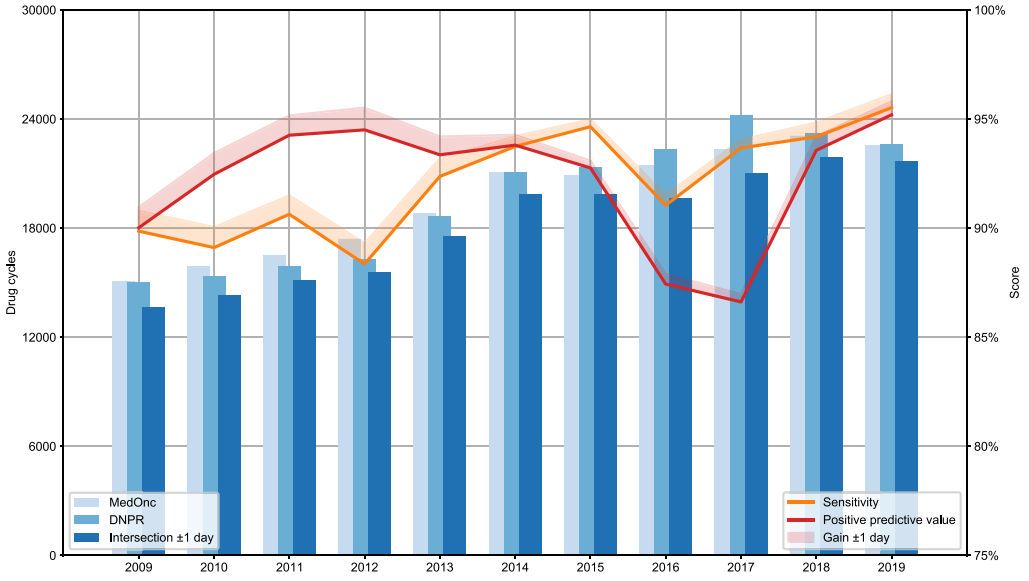


Figure 3 Evolution over time of the validity of the DNPR registrations for L01 drug cycles for systemic anticancer treatments. The lighter surface above each line represents the gain in performance by adding a 1-day margin.

the durations of cycles would be known, especially for adjuvant and neoadjuvant treatments and, to a lesser extent, for palliative treatments. Thus, the whole history of patients could be reconstructed, as a cycle is typically not stopped in the middle but instead cancelled or postponed altogether if the patient is not fit for it.

Temozolomide and Brain Cancer

Temozolomide cycles from the DNPR have a good PPV but a low sensitivity, ie, a significant proportion of these cycles do not seem to have been registered in the DNPR up to 2014 (see Figure 4). This is due to historically poor reporting in the DNPR by administrative personnel. This could be explained by the complexity of the treatment regimen used for glioblastoma[15] and thus point toward reporting issues at the diagnosis level. This poor reporting mechanically impacts the concordance at the patient level, as seen in Table 2.

Recent Drugs

Similar to temozolomide, other drugs, such as pertuzumab, palbociclib, erlotinib, lapatinib, and panitumumab, also display a good PPV with a low sensitivity but for a different reason. Indeed, these are recently introduced drugs for which specific national registry codes were not available when first used, leading to a suboptimal registration at the drug level. For example, pertuzumab was first used in 2012 according to the MedOnc dataset but was only registered in the DNPR with a specific code in 2015.

Cyclophosphamide and Epirubicin

Cyclophosphamide and epirubicin display a low PPV but a high sensitivity. This is due to an error in the registration in 2016 and 2017. These two drugs are administered to breast cancer patients in an adjuvant regimen composed of three cycles of these two drugs followed by three cycles of docetaxel. They were nevertheless registered in the DNPR as given for all six cycles until the registration error was discovered. This can also explain the drop in PPV seen for these years, since they are frequently used drugs to treat breast cancers which is the largest sub-cohort of the study and thus have a significant impact on the overall performance. Outside of these years, the performances are nevertheless good with sensitivities and PPVs above 90%.

Limitations and Strengths

Limitations

MedOnc was used as a reference, but some manual curation was nevertheless needed. We considered MedOnc to

Table 3 Matching Performances for Drug Cycles per Drug Type for Drugs with More Than 500 Cycles in MedOnc

ATC Code	Drug Name	Top 2 Diagnoses	MedOnc	DNPR	Matching	PPV	Sensitivity	F1-Score
L01BC02	Fluorouracil	Colorectal (19,508), Pancreatic (1552)	23,272	22,935	22,644	98.7%	97.3%	98.0%
L01CD01	Paclitaxel	Breast (12,263), Ovarian (2834)	19,001	17,793	17,208	96.7%	90.6%	93.5%
L01CA04	Vinorelbine	Lung (11,098), Breast (5751)	17,779	16,541	16,233	98.1%	91.3%	94.6%
L01XA02	Carboplatin	Lung (8794), Ovarian (3973)	16,099	15,956	15,446	96.8%	95.9%	96.4%
L01XX19	Irinotecan	Colorectal (10,924), Brain (1421)	14,981	15,051	14,500	96.3%	96.8%	96.6%
L01XC03	Trastuzumab	Breast (14,103), Gastro-esophageal (736)	15,094	14,491	14,011	96.7%	92.8%	94.7%
L01XC07	Bevacizumab	Colorectal (7140), Ovarian (2995)	13,924	13,703	13,209	96.4%	94.9%	95.6%
L01XA03	Oxaliplatin	Colorectal (8136), Gastro-esophageal (2826)	12,840	13,063	12,388	94.8%	96.5%	95.6%
L01BC06	Capecitabine	Colorectal (5163), Breast (2785)	11,173	12,394	10,460	84.4%	93.6%	88.8%
L01CD02	Docetaxel	Breast (5378), Prostatic (2703)	9964	9937	9581	96.4%	96.2%	96.3%
L01BC05	Gemcitabine	Pancreatic (5753), Urinary (1768)	9624	9475	9330	98.5%	96.9%	97.7%
L01DB03	Epirubicin	Breast (6925), Gastro-esophageal (1269)	8361	9951	8088	81.3%	96.7%	88.3%
L01AA01	Cyclophosphamide	Breast (6278), Lung (72)	6435	8053	6283	78.0%	97.6%	86.7%
L01XC06	Cetuximab	Colo-rectal (3058), Other (1415)	4754	4710	4483	95.2%	94.3%	94.7%
L01CB01	Etoposide	Lung (3894), Other (273)	4556	4467	4351	97.4%	95.5%	96.4%
L01XA01	Cisplatin	Other (1734), Lung (942)	3785	3652	3497	95.8%	92.4%	94.0%
L01BA04	Pemetrexed	Lung (1936), Other (266)	2244	2253	2216	98.4%	98.8%	98.6%
L01XC18	Pembrolizumab	Lung (1142), Urinary (265)	1688	1640	1623	99.0%	96.1%	97.5%
L01XC17	Nivolumab	Lung (1115), Other (420)	1616	1616	1600	99.0%	99.0%	99.0%
L01AX03	Temozolomide	Brain (2145), Other (78)	2318	1620	1495	92.3%	64.5%	75.9%
L01XX41	Eribulin	Breast (1475), Ovarian (6)	1481	1457	1408	96.6%	95.1%	95.8%
L01DB01	Doxorubicin	Ovarian (774), Endometrial (199)	1260	1246	1202	96.5%	95.4%	95.9%
L01XC13	Pertuzumab	Breast (1554), Other (6)	1560	1073	1035	96.5%	66.3%	78.6%
L01XE33	Palbociclib	Breast (1544), Endometrial (46)	1590	1109	998	90.0%	62.8%	74.0%
L01XX17	Topotecan	Lung (758), Ovarian (184)	1072	1056	979	92.7%	91.3%	92.0%
L01XC08	Panitumumab	Colo-rectal (1220), Ovarian (74)	1350	979	951	97.1%	70.4%	81.7%
L01XE03	Erlotinib	Lung (1223), Breast (87)	1334	1043	889	85.2%	66.6%	74.8%
L01XC14	Trastuzumab emtansine	Breast (746), Colo-rectal (3)	749	736	716	97.3%	95.6%	96.4%
L01CD04	Cabazitaxel	Prostatic (525), Other (3)	528	541	517	95.6%	97.9%	96.7%
L01XE07	Lapatinib	Breast (639), Colo-rectal (9)	648	391	336	85.9%	51.9%	64.7%

https://doi.org/10.2147/CLEP.S312776

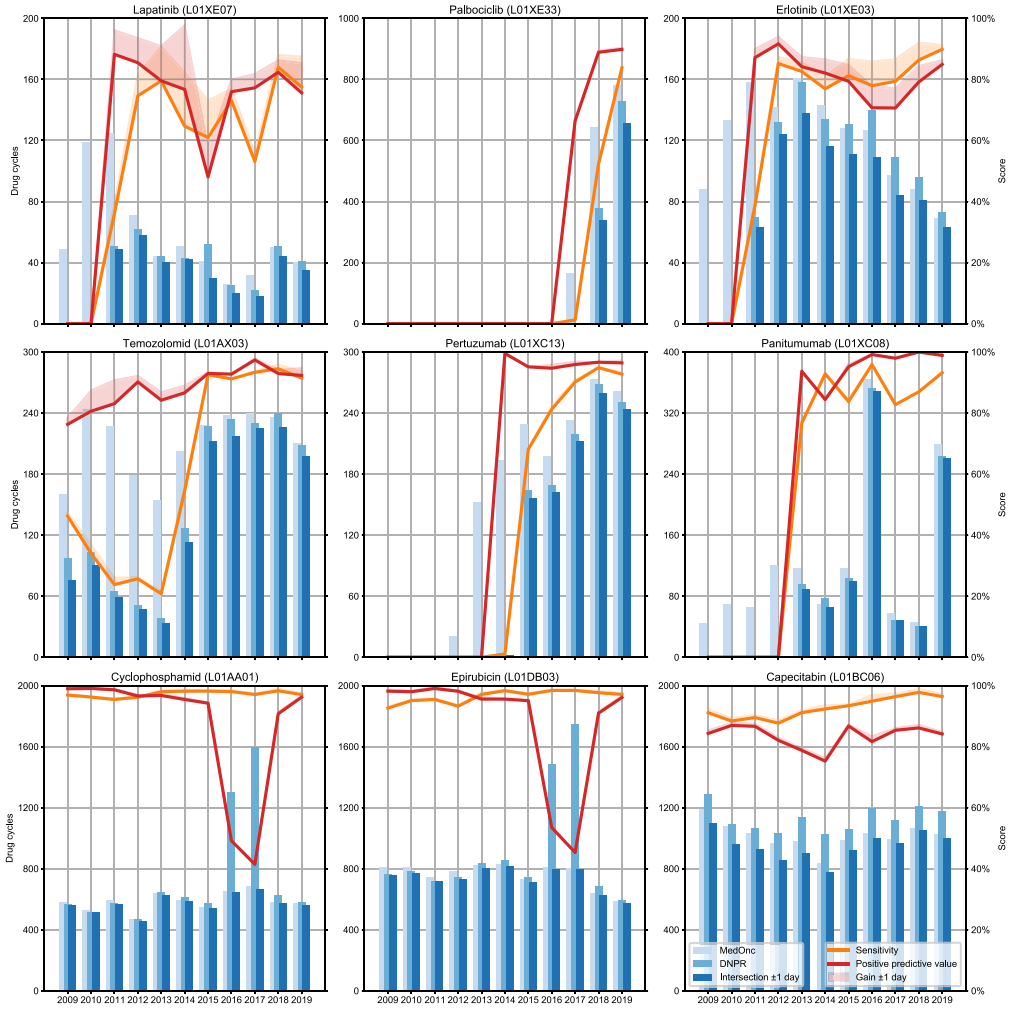


Figure 4 Evolution over time of the validity of the DNPR registrations for bottom 9 performing L01 drugs. Only drugs with more than 500 cycles were considered. The lighter surface above each line represents the gain in performance by adding a 1-day margin.

be a reliable source because it is used in clinical practice to plan, prescribe, and administer treatment; therefore, data entry is expected to be done by doctors and nurses with much more care than in the DNPR, which is an administrative tool filled in by secretaries. However, the DNPR is used for reimbursement of procedures which is a strong incentive to avoid underreporting in this system. The validity of MedOnc compared to patient journals remains unknown but is expected to be similar.

Also, the results shown here might be specific to the North Denmark Region since there might be some spatial and temporal differences across Denmark and Scandinavia in terms of clinical tools and reporting practices. Indeed, Broe et al have reported slight discrepancies between university hospitals and other hospitals,[8] but this study only included data from one university hospital.

We report issues in the DNPR data. However, these issues only affect a limited number of drugs and seem to

have been resolved in recent years. The fact that they are consistent with previously reported results suggests the generalizability of these results.

Strengths

The main strength of this study is its large time span and broad range of cancer diagnoses with low variability in the results, which should guarantee a high level of consistency in the data reported in the DNPR.

Comparison to Other Studies

Only a few articles[7,8] analyzing registration practices are available, and they focus exclusively on colorectal cancers with much smaller cohorts. Broe et al's work[8] is the more directly comparable with ours. For individual drug cycles to colorectal cancer patients, we report a PPV of 94% and a sensitivity of 97% compared to a PPV of 95% and a sensitivity of 90% in Broe et al's study, illustrating the reliability of the MedOnc dataset. Lund et al's study,[7] similarly to our work, reports high validity of the DNPR for fluorouracil, oxaliplatin, and bevacizumab.

Conclusions

This study confirms the validity of the registration of DNPR drug cycles for a large variety of cancer types and antineoplastic drugs, with some limitations for brain cancer and recently introduced drugs. Identified reporting issues, notably for temozolomide, cyclophosphamide, and epirubicin, seem to have been resolved in the latter years of the study period. Therefore, these data can be used for retrospective studies on antineoplastic agent usage across the country.

Acknowledgments

We would like to thank System Administrator of MedOnc, Annette Juul Madsen and Special Consultant Thomas Mulvad Larsen for their help in obtaining and understanding the data needed for this study.

Disclosure

This work was supported by grants from Department of Oncology, Aalborg University Hospital, The Regional Research Fund of North Denmark Region, and from "Det Obelske Familie Fond", no. 50.62 to Ursula G Falkmer. The authors report no other conflicts of interest in this work.

References

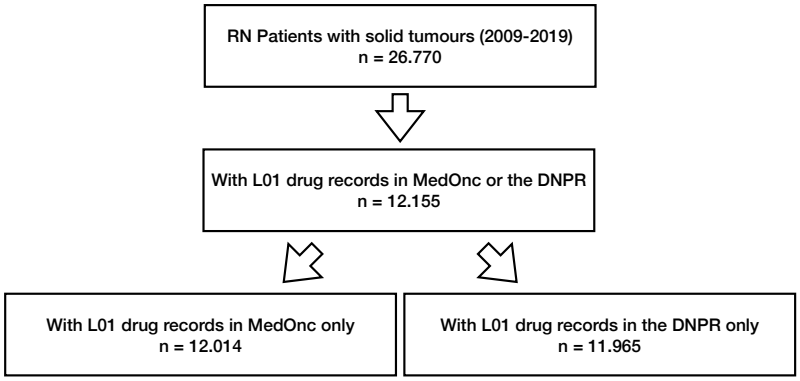
1. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol*. 2010;106(2):86–94. doi:10.1111/j.1742-7843.2009.00494.x
2. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National patient registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
3. Thygesen SK, Christiansen CF, Lash TL, Christensen S, Sørensen HT. Predictive value of coding of diagnoses in the Charlson comorbidity index in the Danish national registry of patients. *BMC Med Res Methodol*. 2011;11(83):2–7. doi:10.1186/1471-2288-11-83
4. Beck MK, Westergaard D, Jensen AB, Groop L, Brunak S. Temporal order of disease pairs affects subsequent disease trajectories: the case of diabetes and sleep apnea. *Biocomput*. 2017;2017:380–389. doi:10.1142/9789813207813_0036
5. Beck MK, Jensen AB, Nielsen AB, Perner A, Moseley PL, Brunak S. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Sci Rep*. 2016;6(July):1–9. doi:10.1038/srep36624
6. Skau Rasmussen L, Vittrup B, Ladekarl M, et al. The effect of postoperative gemcitabine on overall survival in patients with resected pancreatic cancer: a nationwide population-based Danish register study. *Acta Oncol*. 2019;58(6):864–871. doi:10.1080/0284186X.2019.1581374
7. Lund JL, Frøslev T, Deleuran T, et al. Validity of the Danish National Registry of patients for chemotherapy reporting among colorectal cancer patients is high. *Clin Epidemiol*. 2013;5(1):327–334. doi:10.2147/CLEP.S49773
8. Broe MO, Jensen PB, Mattsson TO, Pottegård A. Validity of antineoplastic procedure codes in the danish national patient registry: the case of colorectal cancer. *Epidemiology*. 2020;31(4):599–603. doi:10.1097/EDE.0000000000001208
9. Sundhedsdatastyrelsen. Disease classification system - SKS (in Danish). Available from: https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer/klassifikation-sygdomme. Accessed March 3, 2021.
10. World Health Organization. ICD-10 Version:2016; 2016. Available from: https://icd.who.int/browse10/2016/en. Accessed March 26, 2020.
11. Varian Medical Systems Inc. ARIA OIS for medical oncology. Available from: https://www.siemens-healthineers.com/dk/radiotherapy/software-solutions/ariaois. Accessed November 17, 2021.
12. WHO Collaborating Centre for Drug Statistics Methodology. Anatomical Therapeutic Chemical (ATC) classification system. Available from: https://www.whocc.no/atc/structure_and_principles/. Accessed March 3, 2021.
13. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
14. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. Position Power Acad Publ Play Agents Agendas - Proc 20th Int Conf Electron Publ ELPUB. 2016;2016:87–90. doi:10.3233/978-1-61499-649-1-87.
15. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987–996. doi:10.1056/NEJMoa043330

Supplementary Material

Supplementary table 1 - Cancer type grouping, inclusion status, and exclusion criteria.

ICD10	Cancer type	Specific group	Reason
C00-C14, C30-C33	Head and neck	No	Relatively rare and not primarily treated with systemic anticancer treatment
C15-C16	Gastro-oesophageal	Yes	
C17	Intestine	No	Rare
C18-C20	Colo-rectal	Yes	
C21	Anal	No	Rare
C22-C24	Hepato-biliary	No	Referred to other hospitals
C25	Pancreatic	Yes	
C26, C39, C55, C57, C76, C80	Ill-defined	No	Ill-defined
C34	Lung	Yes	
C37-C38	Thoracic other than lung	No	Rare
C40-C41	Bone and articular cartilage	No	Rare
C43	Melanoma	No	Referred to other hospitals
C44	Skin other than melanoma	No	Not treated with systemic anticancer treatment
C45-C49	Connective and soft tissue	No	Rare
C50	Breast	Yes	
C51-C52	Vulva and vagina	No	Rare
C53	Cervical	No	Referred to other hospitals
C54	Endometrial	Yes	
C56	Ovarian	Yes	
C58	Placenta	No	Rare
C61	Prostate	Yes	
C64-C68	Urinary	Yes	
C69-C70	Eye and meninges	No	Rare
C71	Brain	Yes	
C72	Nervous system	No	Rare
C73-C75	Endocrine	No	Rare
C77-C79	Secondary	No	Not primary

"Specific group" refers to the fact that the corresponding diagnosis groups were considered individually, while other diagnoses were grouped into an "Other" group.



Supplementary figure 1 – Inclusion flow chart. RN stands for North Denmark Region. L01 drugs refer to drugs whose codes, according to the ATC classification, start with L01, namely antineoplastic agents.

Paper III: Dynamic risk prediction of 30-day mortality of patients with advanced lung cancer: Comparing five machine learning approaches

Submitted to Artificial Intelligence in Medicine, November 2021

Authors

Charles Vesteghem[1,2,3], Weronika Maria Szejniuk[1,3,4], Rasmus Froberg Brøndum[1,2,3], Ursula G Falkmer[1,3,4], Chloé-Agathe Azencott[5,6,7], Martin Bøgsted[1,2,3]

Affiliations

[1] Department of Clinical Medicine, Aalborg University, Denmark
[2] Department of Haematology, Aalborg University Hospital, Denmark
[3] Clinical Cancer Research Centre, Aalborg University Hospital, Denmark
[4] Department of Oncology, Aalborg University Hospital, Denmark
[5] CBIO Mines ParisTech, PSL Research University, France
[6] Institut Curie, France
[7] INSERM U900, France

Description

The goal of this paper was to build a predictive model for dynamic risk prediction of 30-day mortality for cancer patients using extensive health data. We decided to focus on patients with advanced lung cancer to obtain a more homogeneous cohort. To find the best suited approach, we compared five different machine learning methods: logistic regression with elastic net regularisation, random forest, gradient tree boosting, multilayer perceptron, and long short-term memory (LSTM) architecture. The LSTM architecture was considered for its ability to handle the sequence of events, as the trajectory of the patients could play a role in the short-term mortality of patients and summary variables may not capture all of the available information.

1. Abstract

1.1. Background

Administering systemic anticancer treatment (SACT) to patients near death can negatively impact their health-related quality of life, often with limited clinical benefits. Therefore, late SACT administrations should be avoided in these cases. The availability of extensive registry data suggests exploring machine learning techniques to build decision support tools for clinicians to limit late SACT administration.

1.2. Material & methods

Patients with advanced lung cancer who were treated at the Department of Oncology, Aalborg University Hospital and died between 2010 and 2019 were included (n=2368). Their diagnoses, treatments, biochemical data, and histopathological results were collected, and corresponding summary variables were generated. The data were used to train predictive models of 30-day mortality using five different machine learning approaches, logistic regression with elastic net penalty, random forest, gradient tree boosting, two artificial neural networks, a multilayer perceptron, and a long short-term memory network. The importance of the variables in each model was estimated using Shapley additive explanation values. Clinical utility was evaluated by estimating the number of preventable SACT administrations in the last 30 days while avoiding treatment cessation before 90 days of death.

1.3. Results

The random forest and gradient tree boosting models outperformed other models, while the artificial neural network models underperformed. Adding summary variables had a modest effect on performance with an increase in average precision from 0.500 to 0.505 and from 0.498 to 0.509 for the gradient tree boosting and random forest models, respectively. Most of the top variables selected in each model were biochemical results, notably albumin, lactate dehydrogenase, leukocytes, neutrophils, and carbamide values.
Biochemical results alone contained most of the information with a limited degradation of the performances when fitting models with only these variables. The average precision decreased from 0.509 to 0.493 and from 0.505 to 0.487 for the gradient tree boosting and random forest models, respectively.
The utility analysis showed that by applying a simple threshold to the predicted risk of 30-day mortality, 44% of late SACT administrations could have been prevented at the cost of 3% of patients stopping their treatment 90 days before death.

1.4. Conclusion

This study demonstrates the potential of a decision support tool to limit late SACT administration in cancer patients. Further work is warranted to refine the model, build an easy-to-use prototype, and conduct a prospective validation study.

2. Background

Systemic anticancer therapy (SACT) includes chemotherapy, targeted therapy, immunotherapy, and hormonal therapy. SACT should only be considered in patients with an adequate benefit from the treatment since SACTs often have a short-term negative impact on health-related quality of life[1–7]. An accepted threshold for late SACT administration is 30 days before death[8]. However, clinicians' experience in predicting the remaining lifetime of patients may be inadequate[9], leading to prescription of SACT too late to achieve a clinical benefit[8]. Furthermore, death from advanced cancer often has a multifactorial background where acute complications, such as infections, venous thromboembolisms or myocardial infarctions, could lead to patient death.

Lung cancer is a frequently occurring cancer type with poor prognosis and high mortality, particularly in advanced stages. Thus, patients with lung cancer are at higher risk of receiving SACT close to death than other cancer types with a better prognosis.

There is a need for decision support tools to assist the work of clinicians to minimize the risk of decreasing health-related quality of life due to SACT of lung cancer patients receiving palliative treatment in advanced stages. Patient health might promptly deteriorate during treatment, requiring frequent use of dynamic predictive tools to assess their situation. To the best of our knowledge, existing studies addressing this issue 1) are based on a limited number of clinical variables, 2) do not consider artificial neural network-based models, 3) are based on different endpoints, for example, 6-month mortality, or 4) are not suitable for dynamic risk prediction[10–17].

The aim of this study was to investigate the potential use of machine learning approaches on electronic health registers and administrative data to limit late SACT by building dynamic predictive models for the 30-day mortality of patients with advanced lung cancer. It is based on the hypothesis that extensive medical data can improve the performances of the predictive models and on the hypothesis that artificial neural network-based machine learning techniques can outperform other methods.

3. Materials and methods

3.1. Data sources and data management

This study was based on five data sources from the North Denmark Region (see Table 1). The data were merged at the patient level in patients with advanced lung cancer treated at the Department of Oncology, Aalborg University Hospital between 2010 and 2019, with the Danish civil registration number and were subsequently pseudonymized. Data management and analysis followed a protocol similar to that proposed by Tomašev et al.[18]. For each patient, a sequence of records was generated. A record was defined as a day where a diagnosis, a procedure, a drug prescription, a result, or contact with the Department of Oncology, Aalborg University Hospital was made. All variables were present for each record (Table 1). All data available for a given day were grouped into one record, the latest value being retained in cases of multiple measurements on the same day. The dataset contained three different types of variables: baseline, cumulative, and status. Baseline variables represented information up to 30 days before diagnosis and were constant for each patient across the sequence of records (Figure 1).

Table 1. *Datasets and corresponding variables*

Dataset	Variable group	Inclusion criteria	Variable type	Variables	Baseline	Patients	Events per patient
CPR	Sex, age		Baseline	2 (2)	2	2368 [2368-2368]	N/A
	Date of death	Died between 2010 and 2019	Outcome	1 (1)	N/A	2368 [2368-2368]	1 [1-1]
PAS	Cancer diagnoses	Lung cancer patients (advanced stage)	Baseline and cumulative	10 (73)	5	307 [24-2368]	1.8 [1.1-3.7]
	Comorbidities		Baseline and cumulative	8 (20)	11	107 [26-360]	1.6 [1.2-2.1]
	Symptoms and side effects		Baseline and cumulative	9 (20)	6	180 [38-514]	1.3 [1.1-1.6]
	Surgeries		Cumulative	10 (405)	0	96 [29-206]	1.1 [1.0-1.5]
	Radiotherapies		Cumulative	2 (7)	0	824 [701-946]	12.2 [8.2-16.2]
MedOnc	Drug administrations	Received palliative treatment	Cumulative	25 (64)	0	394 [25-1890]	15.1 [1.8-161.2]
	"If needed" drug prescriptions		Cumulative	3 (6)	0	838 [314-1857]	7.1 [1.5-11.9]
	BMI		Status	1 (1)	0	1677 [1677-1677]	4.7 [4.7-4.7]
LABKA	Biochemical data		Status	45 (62)	0	1648 [220-2367]	10.8 [2.2-31.6]
Patobank	Morphology and biomarkers		Status	56 (1170)	0	160 [22-929]	1.1 [1.0-1.5]

For the "Variable type" column, "Baseline" indicates which variable groups were included as baseline variables, i.e., which variables were present before diagnosis (with a margin of 30 days). The "status" and "cumulative" types describe the method used for filling missing values (see paragraph 2.3). The "Variables" column indicates the number of variables included in each variable group, and the value in parentheses is the number of variables before filtering. The "Baseline" column informs on the number of variables included as baseline variables. The "Patients" column indicates the mean number of patients for each variable with at least one event for this variable. "Events per patient" is the mean number of events per patient with at least one event and per variable. For the last two columns, the values between brackets show the range across the variables of the group.

The status and cumulative variables differed by the method used for filling missing values. For status variables, the last value was carried forward, while for the cumulative variables, empty values were designated with zeros. Status variables represented a potentially variable state, such as BMI, blood tests, diagnoses, or biomarkers. Cumulative variables included those that could be counted or summed and were treatment related, e.g., the number of a certain type of surgery or cumulative dose of a certain drug. The latest value for each variable was carried forward in both cases when working with models not handling sequences of records. If no value was still available, the mean value was used. Data management was performed using SAS Enterprise Guide 8.3 (SAS Institute Inc., Cary, NC, USA) and Python 3.8 with Jupyter[19] notebooks.

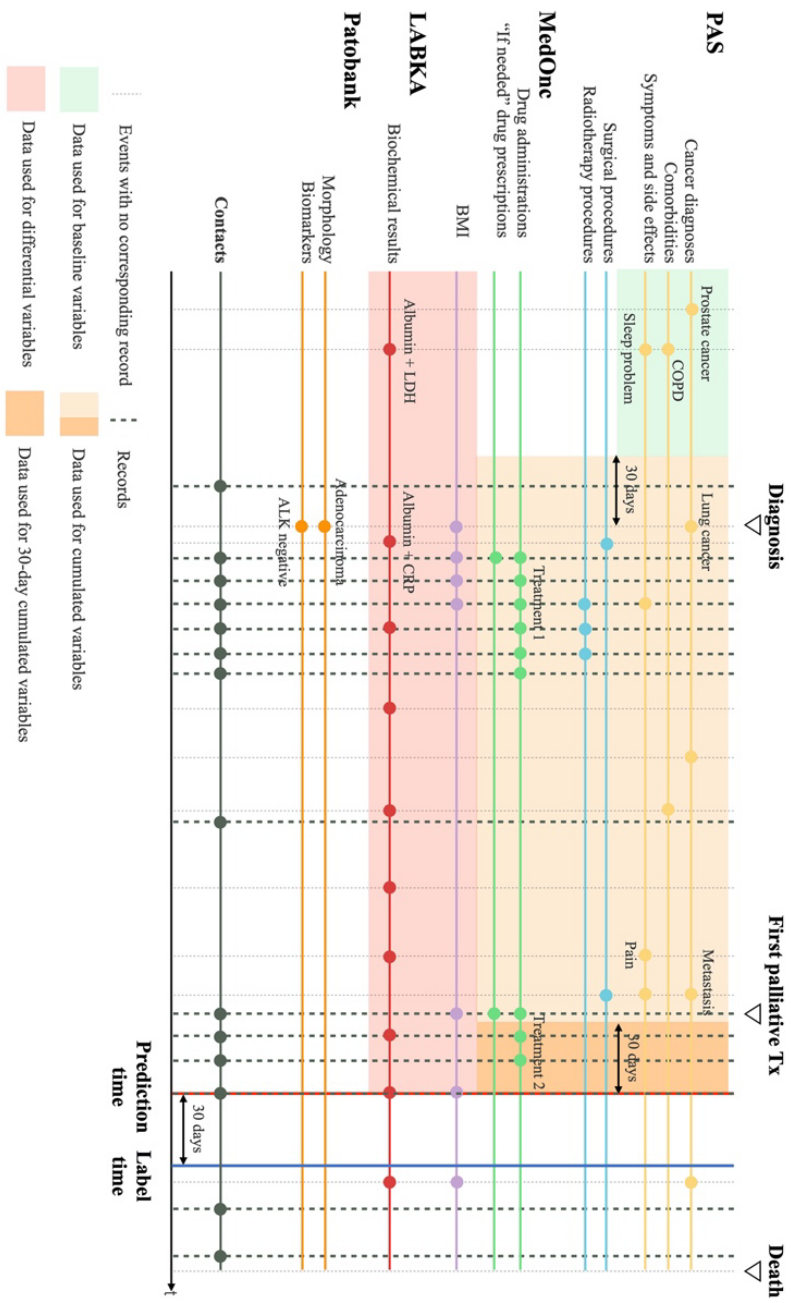


Figure 1. *Example of records for a patient. The prediction time corresponds to the timepoint where the prediction is made, the prediction being whether the patient is dead or alive at the label time. Each dot on the figure represents a data point.*

3.1.1. PAS

The primary data source was the North Denmark Region's Patients Administrative System (PAS). PAS includes all diagnoses and procedures from hospital inpatient and outpatient visits coded following the Danish Health care Classification System[20] (SKS), which is similar to the ICD-10 classification[21] for diagnoses. All other cancer diagnoses before or after the lung cancer diagnosis, comorbidity diagnoses according to the Charlson's Comorbidity Index[22] (CCI), symptom diagnoses, and side effect diagnoses (Supplementary Table 1) were considered. Diagnoses for the localization of metastases were included individually. These diagnostic variables were used to calculate corresponding baseline variables and were used directly as status variables containing binary values.

PAS was also used to extract procedures for surgery (the SKS codes between KA and KQ) and radiotherapy (the SKS codes starting with BWG but not BWGA) performed in relation to a cancer diagnosis. These procedure variables were also used as cumulative variables with binary values. We excluded minor surgical procedures (SKS codes starting with KT), endoscopies, unknown operations, and procedures related to transplantation due to either dependency on local practice or lack of relevance in the study context.

3.1.2. The CPR registry

The Danish Civil Registration System (CPR) registry contains information on sex, date of birth, and date of death for patients in contact with the Department of Oncology between 01/01/2008 and 31/12/2019. This information was used to compute the "sex" and "age" baseline variables. Furthermore, the date of death was used to label the binary outcome variable, i.e., 30-day mortality from the record of which the prediction was made (Figure 1).

3.1.3. MedOnc

Data from the prescription software ARIA OIS for Medical Oncology v13.7 (Varian Medical Systems Inc., Palo Alto, CA, USA) (MedOnc), used at the Department of Oncology, contains information about drug prescriptions as well as body mass index (BMI). The drug prescriptions were characterized by date, Anatomical Therapeutic Chemical classification[23] (ATC) code, dose in mg, and RN (Pro Re Nata, if needed) status. The dose for non-PRN drugs records whether the drug was given, with a dose equal to zero for non-administered drugs. Additionally, a regimen name variable was available to infer the intent of the treatment, i.e., neoadjuvant, adjuvant, or palliative. A cumulative variable containing information about the dose administered was created for each combination of ATC code and PRN status. The BMI was included as a status variable.

3.1.4. Patobank

The Danish National Pathology Registry (Patobank) contains histopathological data, including morphology, determining the subtype of lung cancer, and genetic biomarkers, such as the presence of significant mutations in the epidermal growth factor receptor gene or the programmed death ligand 1 protein expression level. These data were used as binary status variables.

3.1.5. LABKA

Biochemical results data from the Clinical Laboratory Information System (LABKA) used at Aalborg University Hospital were included as status variables. This includes mostly biochemical tests. These data were coded using the Nomenclature, Properties, and Units[24] (NPU) classification.

3.1.6. Summary variables

Variables generated from these five data sources are collectively referred to as base variables. To capture information on the trajectory of the patients for models that do not support time series, summary variables were created from the base variables. For each cumulative variable, two additional cumulative variables were created, one adding the values accumulated from diagnosis to the current record and one for the last 30 days before the current record as reported by Elfiky et al.[17].

Additionally, binary baseline variables were created for the cancer diagnoses, comorbidities, side effects, and symptoms from the PAS dataset based on the presence of corresponding diagnoses before the initial lung cancer diagnosis.

For biochemical results and BMI data, mean values for the past records, including data before the start of the first palliative treatment, was computed. A differential variable was calculated as the difference between the value for the current record and the mean of the previous measures.

For the machine learning model designed to handle the sequence of records (see Models section), datasets containing these variables were not included.

3.1.7. Dataset generation and feature selection

The overall dataset included records only after the start of the first palliative treatment or, if no palliative treatment was initiated, the first diagnosis of metastatic disease. Only records or sequences of records associated with contact with the Department of Oncology were retained.

Two versions of the dataset were created, one with only the base variables, referred to as the base dataset, and one with both the base and summary variables. For both datasets, a nonspecific feature filtering step was performed to allow the convergence of all models. This consisted of filtering out variables found in less than 1% of patients and highly correlated or colinear variables with thresholds of 0.99 for Pearson's correlation coefficient and 20 in variance inflation factor[25] (VIF) for multicollinearity. In cases with a high correlation between two variables, the first variable alphabetically was removed. Variables with the highest VIF were removed first. The correlation was calculated using the pandas library[26], and the VIF was calculated using linear regression. If a simple logistic regression could still not be trained, the VIF threshold was lowered by one unit until convergence was possible. Both the linear and logistic regressions were trained using the Python scikit-learn library[27].

In a second phase, an additional dataset, referred to as the biochemical results dataset, was generated keeping only the variables from the LABKA and CPR data sources to compare performance to the base dataset. Two versions of this dataset were also created and filtered as detailed above.

3.2. Study population

Patients in contact with the Department of Oncology between 01/01/2008 and 31/12/2019 and who died between 01/01/2010 and 31/12/2019 were identified from PAS (n=14,902). Among these patients, 3,856 were diagnosed with lung cancer. Only those who received SACT for advanced or metastatic lung cancer or who were diagnosed with metastatic lung cancer were included in the final dataset (n=2368, Supplementary Figure 1). The patients were split into three cohorts: the training cohort with patients who died between 2010 and 2017, the validation cohort with patients who died in 2018, and the test cohort for patients who died in 2019. Since only patients in contact with the Department of Oncology between 2008 and 2019 were accessible, including patients who died in the same period would exclude patients who died in this period but were in contact with the Department of Oncology only before 2008. To avoid missing this type of patient, only patients who died after 2010 were included. A 2-year margin was considered sufficient in this context.

3.3. Models

In addition to a logistic regression model without regularization, which was used as a baseline model, five popular machine learning models were considered in this study: a logistic regression with elastic net regularization[28] (LRENR), a random forest classifier[29] (RF), a gradient tree boosting classifier[30] (GB), a multilayer perceptron (MLP), and a long-short term memory model[31] (LSTM), as proposed in the literature[13,14,32]. The architectures of the MLP and the LSTM are shown in Supplementary Figure 2.

Hyperparameters were optimized using records from the training cohort, referred to as the training records, as the training set, and the records from the validation cohort, referred to as the validation records, as the validation set. In practical terms, the models were trained using the training set with various values for the hyperparameters, and the performance was evaluated on the validation set to select the best set of hyperparameters. To limit the complexity of the tree-based ensemble models (RF and GB), the lowest values among those resulting in performance within 1% of the best performance were selected. The 1% threshold was arbitrarily set. No cross-validation was performed to maintain the temporal structure of the data. Once the optimal hyperparameter values were found, the models were retrained with these values on the training and validation records combined. The final performance was evaluated on the records using the held-out test cohort, referred to as the test records[33].

To assess the variability of the performance for all models, nine additional training sets of the same length were generated by bootstrapping the combined training and validation records. Using the optimal hyperparameters, the models were fitted using these ten sets. In each case, performance was evaluated on the test records. This process was also performed using the training records as the training set and the validation records as the validation set.

LRENR, RF, and GB were fitted using the scikit-learn library, and hyperparameters were optimized using a grid search. The hyperparameters optimized for LRENR were the inverse of regularization strength and the elastic net mixing parameter with values between 10^{-5} and 10^{-2} and 0.1 and 1, respectively. For RF and GB, optimization was performed on the number of trees, with values between 1 and 2000 and the maximum

depth of these trees. This depth was set between 2 and 200 for RF and between 1 and 50 for GB (Supplementary Figure 3). The two artificial neural network-based models were trained using the Python Keras library[34], and their best hyperparameters were found by Bayesian optimization using the Python Keras-Tunner library[35] (Supplementary Table 2).

The average precision (AP) and receiver operating characteristic area under the curve (ROC AUC), as calculated by scikit-learn, were used to evaluate model performance. The AP was the primary performance metric used as recommended for imbalanced datasets[18]. This metric is equivalent to the area under the precision-recall curve.

The importance of each variable was estimated from their SHapley Additive exPlanation (SHAP) values[36] on a random sample of 1000 records from the combined training and validation records. The SHAP approach was chosen due to its ability to generate comparable results between all models. To measure the overall importance of each variable, the effect of each variable was calculated by summing the absolute values of these effects across the sampled records. For models using summary variables, the effects of potential summary variables were added to the effect of the corresponding base variables at the record level before the summation of absolute values across records.

3.4. Evaluation of utility

To assess the usefulness of a predictive model, the potential effect of limiting late SACT administrations based on a simple rule was investigated. Given a threshold on the 30-day mortality risk, SACT should be administered if the predicted risk is below that threshold. Conversely, if the predicted risk is above the threshold, SACT should not be given. In cases where the risk is above the threshold at a given time point but becomes below the threshold at a later stage, SACT is considered administrable at that later timepoint and is therefore only considered delayed. An administration was considered preventable if the risk prediction at the time of administration, as well as for all subsequent contacts, were above the threshold (Supplementary Figure 4).

To avoid stopping treatment too early, no administration should be considered preventable more than 90 days before death in more than 1% of the patients, putting a constraint on the value of the threshold. The threshold value was determined from a model trained on the training records and used to predict risks for each validation record. The lowest value for the 30-day mortality risk prediction fulfilling the above constraint was selected as the threshold.

Risk was predicted for each test record using the best model with respect to AP trained on the combined training and validation records of the base dataset with summary variables. The threshold-based rule was applied to these predictions to identify preventable SACT administrations within 30 days, 90 days and more than 90 days from death and the corresponding number of patients.

To compare utilities, the F_1-score (harmonic mean of precision and recall) was calculated, where SACT administrations within 30 days of death were considered positive events and SACT administrations more than 90 days from death negative events. SACT administrations between 30 days and 90 days were not considered in the calculation. Preventable SACT administrations according to the threshold-based rule were

considered as predicted as positive, true positive events being preventable SACT administrations with 30 days of death.

3.5. Ethical approval and registration

According to Danish legislation, health registry projects do not require patient consent or ethical approval but should be registered by the legal entity responsible for the data. The study was registered at the North Denmark Region's research project inventory (reg. number 2019-41).

4. Results

4.1. Study population

The population characteristics are described in Table 2. The overall cohort was well balanced regarding sex (52% male and 48% female). A majority of patients (56%) died between the age of 60 and 74, and most patients (87%) received palliative SACT, among whom 65% died less than 12 months from the initiation of palliative treatment. The most prevalent histopathological subtype was adenocarcinoma (43%), followed by small cell carcinoma (25%). An increasing proportion of adenocarcinoma and decreasing incidence of small cell lung cancer were observed between the training, validation, and test cohorts. The number of patients surviving more than 12 months was increased in the test (50.5%) cohort compared to the training (32.6%) and validation cohorts (35.4%).

4.2. Performances

4.2.1. Comparing models

The AP and ROC AUC of the five models were computed for all datasets (Figure 2). First, all values from the base dataset, with or without summary variables, were considered. There were limited differences in the validation set, with mean values for the AP varying between 0.486 and 0.544. The inclusion of summary variables had a beneficial effect on the performance. The differences were larger on the test set where values between 0.342 and 0.509 were observed, with the MLP and LSTM underperforming. The effect of the summary variables was beneficial on the performances of all applicable models except the MLP model but was modest for the GB and RF models, changing from 0.500 to 0.505 and from 0.498 to 0.509, respectively.

Regarding performance variability, using the bootstrapped datasets, the LRENR model exhibited the least variability on the validation set. The variability increased for the LRENR and LSTM models on the test set and remained similar for the other models.

The patterns were similar using the ROC AUC as a performance metric, with values between 0.794 and 0.846 in the validation set and between 0.766 and 0.868 in the test set, with the MLP and LSTM models performing poorly on the test set. For both AP and ROC AUC, the best-performing approaches remained the same on the validation and test sets, suggesting robust results.

Table 2. *Study population characteristics*

	Training set	Validation set	Test set	Overall
Patients	1 819 (100.0%)	309 (100.0%)	240 (100.0%)	2 368 (100.0%)
Sex				
Male	937 (51.5%)	168 (54.4%)	117 (48.8%)	1 222 (51.6%)
Female	882 (48.5%)	141 (45.6%)	123 (51.3%)	1 146 (48.4%)
Histopathology				
Adenocarcinoma	753 (41.4%)	143 (46.3%)	122 (50.8%)	1 018 (43.0%)
Small cell carcinoma	493 (27.1%)	64 (20.7%)	43 (17.9%)	600 (25.3%)
Large cell carcinoma	264 (14.5%)	30 (9.7%)	27 (11.3%)	321 (13.6%)
Squamous cell carcinoma	223 (12.3%)	57 (18.4%)	40 (16.7%)	320 (13.5%)
Other	86 (4.7%)	15 (4.9%)	8 (3.3%)	109 (4.6%)
Age at death				
18-44	35 (1.9%)	3 (1.0%)	5 (2.1%)	43 (1.8%)
45-59	416 (22.9%)	54 (17.5%)	47 (19.6%)	517 (21.8%)
60-74	1 071 (58.9%)	165 (53.4%)	132 (55.0%)	1 368 (57.8%)
75+	297 (16.3%)	87 (28.2%)	56 (23.3%)	440 (18.6%)
Palliative treatment				
Yes	1 657 (91.1%)	226 (73.1%)	194 (80.8%)	2 077 (87.7%)
No	162 (8.9%)	83 (26.9%)	46 (19.2%)	291 (12.3%)
Survival from start of palliative treatment				
0-1 months	105 (6.3%)	9 (4.0%)	9 (4.6%)	123 (5.9%)
1-6 months	526 (31.7%)	70 (31.0%)	46 (23.7%)	642 (30.9%)
6-12 months	486 (29.3%)	67 (29.6%)	41 (21.1%)	594 (28.6%)
12+ months	540 (32.6%)	80 (35.4%)	98 (50.5%)	718 (34.6%)
Contacts				
All contacts	68 876 (100.0%)	12 205 (100.0%)	12 209 (100.0%)	93 290 (100.0%)
Within 30 days of death	10 783 (15.7%)	2 190 (17.9%)	1 526 (12.5%)	14 499 (15.5%)

The percentages in parentheses present the proportion of corresponding patients in the cohort, except for the survival data, where the value is the proportion of corresponding patients among patients who received palliative treatment, and the contact data, which represents the proportion of contacts.

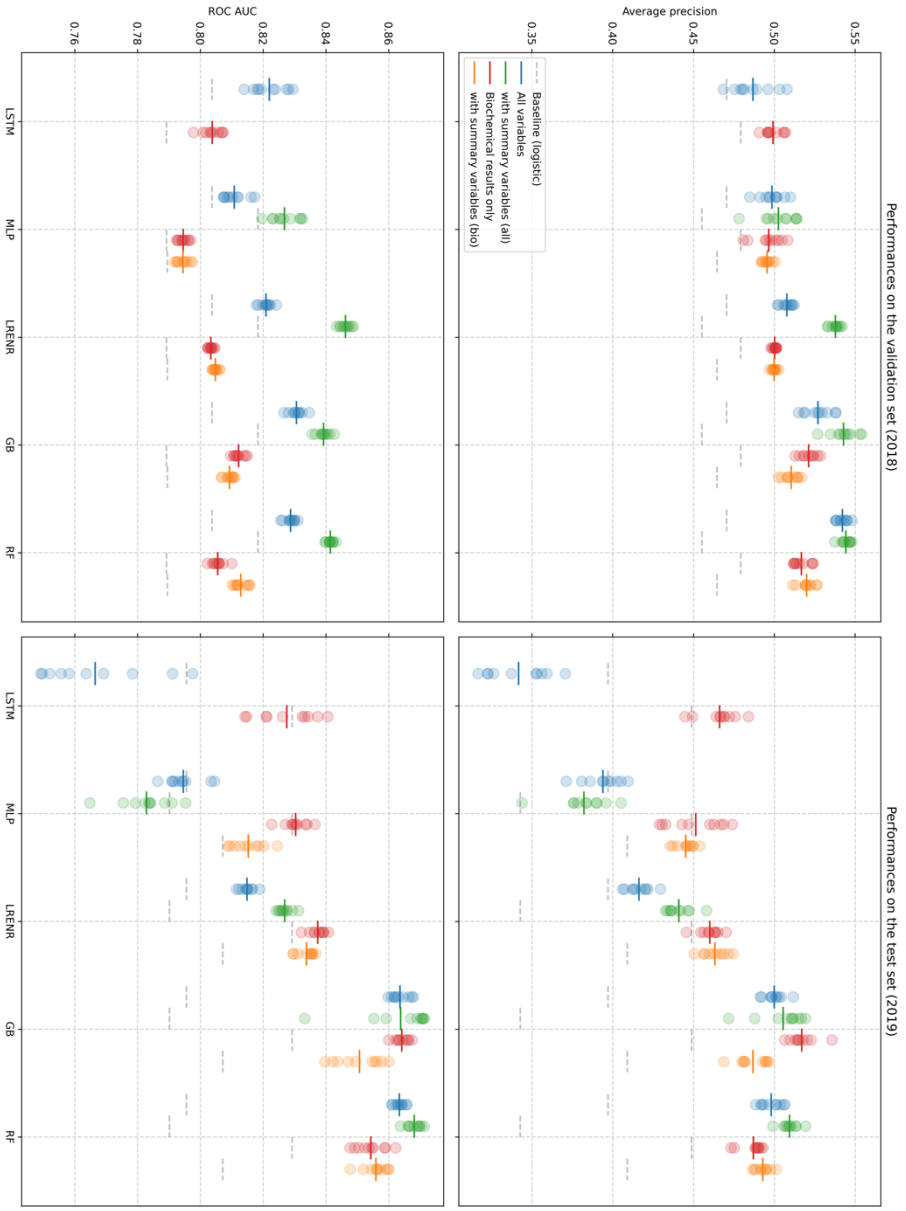


Figure 2. *Average precision and ROC AUC per model and dataset on the validation and test sets. The baseline values show the performance of a logistic regression without regularization. The horizontal lines represent the mean value of the corresponding metric. Each circle represents the performance on a bootstrapped dataset.*

4.2.2. Top variables

For the complete dataset, most of the top variables by importance, without (Figure 3A) or with summary variables (Figure 3B), were the biochemical results across all models,

especially in the RF model (10 and 8 were in the top 10, respectively) and the GB model (8 were in the top 10 in both cases). In particular, albumin, leukocytes, carbamide, and lactate dehydrogenase were all in the top 10 for all models with or without summary variables, and albumin and leukocytes were in the top 4 variable for all models. Considering only the best performing models, RF and GB, with summary variables, creatinine and neutrophils were additionally in the top 10. Concerning nonbiochemical variables, treatment data were important only when summary variables were included with prednisolone and carboplatin in the top 10 for both the RF and GB models.

4.2.3. Performances using only biochemical results

As most of the important variables were biochemical results, the performance of models trained exclusively on these variables was investigated (Figure 2). This had a limited negative impact on the performance of RF and GB. The mean AP values between the base dataset and the biochemical results dataset, both with the summary variables, went from 0.509 to 0.493 and from 0.505 to 0.487 in the RF and GB models, respectively. Notably, the optimum performance for the GB model was obtained for the biochemical results dataset without the summary variables with a mean AP of 0.517.

4.1. Utility

The RF model was chosen as the best model in terms of AP on the base dataset with summary variables to assess the utility of a decision support tool in preventing late SACT administrations.

In the test cohort of the 195 patients who received palliative SACT, 16% (n=32) received, on average, 3.2 (103/32) administrations within 30 days of death (Table 3).

The threshold identified by the validation set was 34.9%. Using this threshold, 44% (14/32) of patients from the test cohort could have had preventable SACT administrations within 30 days of death, corresponding to 44% (44/103) of the late SACT administrations (see Figure 4). However, this threshold led to preventable SACT administrations before the 90-day landmark for 3% (6/195) of patients. The 44 preventable late SACT administrations were primarily for osimertinib, etoposide, alectinib, and carboplatin.

The F_1-scores for the GB and RF models and all datasets were calculated both at the patient and administration levels in Table 4 with heterogeneous results. The best F_1-score was observed for the RF model with the base dataset without summary variables, while for the GB model, the best F_1-score was for the biochemical data without summary variables.

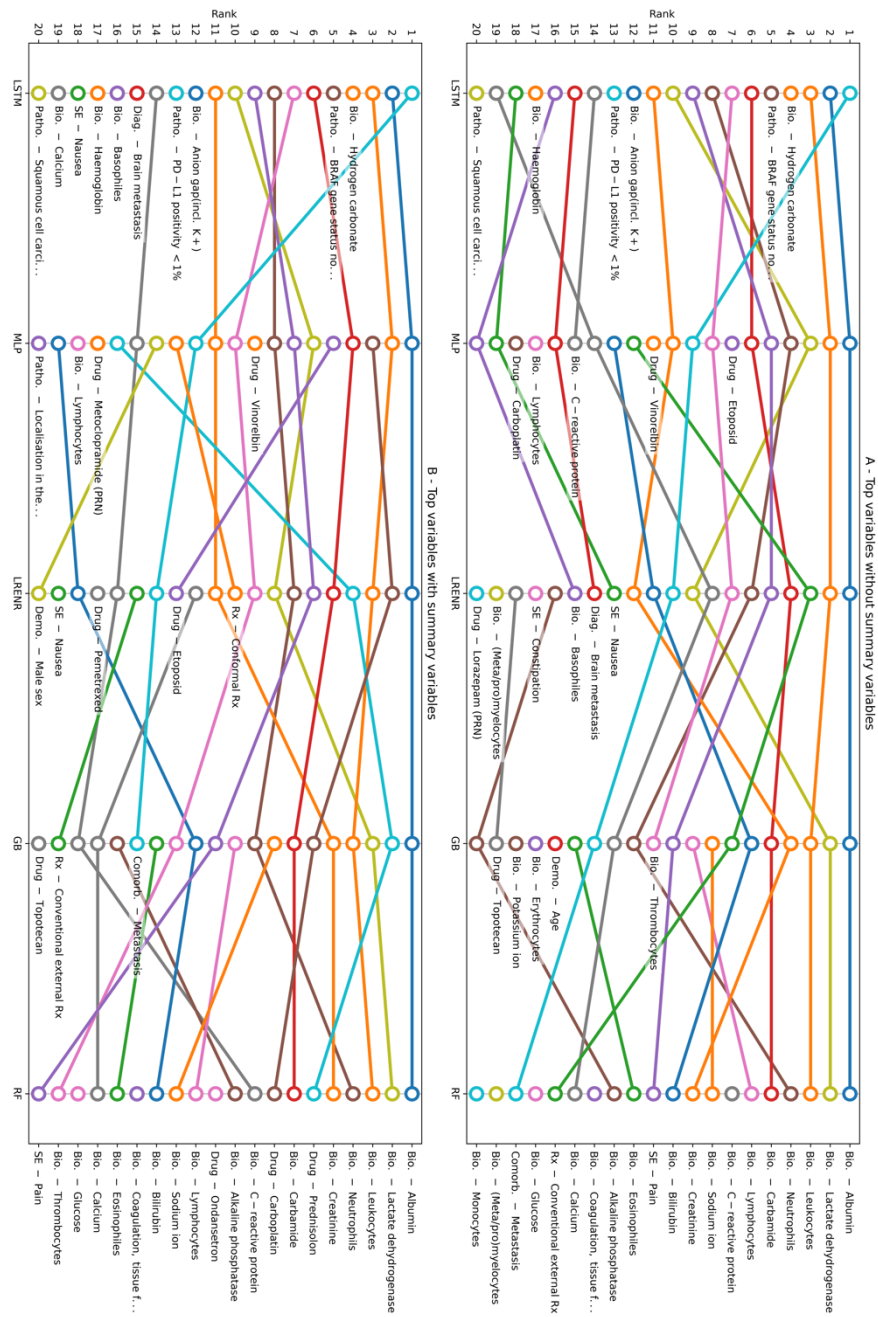


Figure 3. *Top variables per importance for the full dataset with and without summary variables. "Bio." stands for biochemical results, "Rx" for radiotherapy, "Drug" for drug administration, "SE" for side effects, "Comorb." for comorbidities, "Patho." for results from pathological analysis, and "Demo." for demographic variables. The same LSTM model was compared to the other models both without and with summary variables.*

Table 3. *Utility evaluation for the prevention of SACT administration within 30 days of death in the RF model on the base dataset with summary variables.*

	Administered		Preventable						
	Total	Within 30 days	Total	Within 30 days	Within 90 days	Before 90 days	Pr.	Re.	F1-score
Treated patients	195	32 (16.4%)	23 (11.8%)	14 (7.2%)	6 (3.1%)	6 (3.1%)	0.7	0.4	0.5
Administrations	4154	103 (2.5%)	73 (1.8%)	44 (1.1%)	18 (0.4%)	11 (0.3%)	0.8	0.4	0.6
Etoposide	500-1000	4%	3%	1%	1%	0%	0.8	0.3	0.4
Vinorelbine	500-1000	1%	0%	0%	0%	0%	1.0	0.3	0.4
Carboplatin	500-1000	3%	1%	1%	0%	0%	0.8	0.3	0.4
Nivolumab	200-500	1%	0%	0%	0%	0%	NA	0.0	NA
Pemetrexed	200-500	2%	2%	0%	0%	1%	0.0	0.0	NA
Osimertinib	200-500	8%	6%	6%	0%	0%	1.0	0.8	0.9
Gemcitabine	50-200	0%	1%	0%	1%	1%	0.0	NA	NA
Docetaxel	50-200	2%	2%	1%	1%	0%	1.0	0.7	0.8
Pembrolizumab	50-200	1%	3%	1%	0%	2%	0.4	1.0	0.6
Topotecan	50-200	8%	3%	3%	0%	0%	1.0	0.4	0.6
Alectinib	50-200	14%	14%	14%	0%	0%	1.0	1.0	1.0

The "Administered" columns inform on the number of SACT administrations given. The corresponding "Within 30 days" column informs on the number of SACT administrations given within 30 days of death. "Preventable" columns are for the numbers of SACT administrations that would have been prevented using the threshold as mentioned in Section 2.4. The corresponding "Within 30 days" informs on the number of SACT drug administrations that would have been preventable within 30 days of death. The "Within 90 days" column provides the number of SACT drug administrations that would have been preventable between 90 days and 30 days to death. The "Before 90 days" column informs on the number of SACT drug administrations that would have been preventable more than 90 days from death. The "Treated patients" row provides information on the number of corresponding patients for each column. For individual drugs, only the range of total administrations is shown, and the percentages shown are in relation to the total number of administrations or patients and are rounded to the closest percent for anonymisation. "Pr.", "Re." and "NA" stand for Precision, Recall and Non-applicable, respectively.

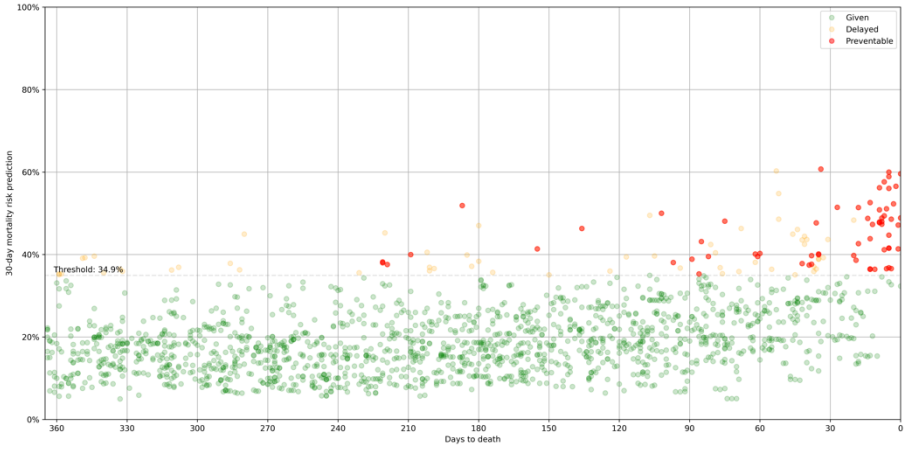


Figure 4. *30-day mortality risk predictions in the last 365 days before death for patients included in the test dataset using the RF model with summary variables. Each circle represents a SACT drug administration. The trajectories are not represented, and the days to death values have been randomly shifted within 2 days of the actual values for anonymisation.*

Table 4. *Comparison of the RF and GB models on all 4 datasets in terms of F1-score for utility evaluation.*

		Patients				Admin.			
	Thres.	Within 30 days	Within 90 days	Before 90 days	F1-score	Within 30 days	Within 90 days	Before 90 days	F1-score
RF									
Base dataset	39.5%	15 (7.7%)	7 (3.6%)	4 (2.1%)	0.59	47 (1.1%)	19 (0.5%)	10 (0.2%)	0.59
With sum. variables	34.9%	14 (7.2%)	6 (3.1%)	6 (3.1%)	0.54	44 (1.1%)	18 (0.4%)	11 (0.3%)	0.56
biochemical results dataset	35.8%	13 (6.7%)	7 (3.6%)	5 (2.6%)	0.52	43 (1.0%)	22 (0.5%)	10 (0.2%)	0.55
With sum. variables (bio)	34.1%	14 (7.2%)	11 (5.6%)	8 (4.1%)	0.52	49 (1.2%)	33 (0.8%)	18 (0.4%)	0.58
GB									
Base dataset	65.8%	9 (4.6%)	5 (2.6%)	3 (1.5%)	0.41	25 (0.6%)	18 (0.4%)	7 (0.2%)	0.37
With sum. variables	44.6%	11 (5.6%)	5 (2.6%)	4 (2.1%)	0.47	27 (0.6%)	9 (0.2%)	4 (0.1%)	0.40
biochemical results dataset	42.9%	14 (7.2%)	8 (4.1%)	6 (3.1%)	0.54	47 (1.1%)	36 (0.9%)	13 (0.3%)	0.58
With sum. variables (bio)	34.8%	14 (7.2%)	9 (4.6%)	5 (2.6%)	0.55	40 (1.0%)	31 (0.7%)	10 (0.2%)	0.52

In the "Admin." section referring to administrations, the "Thres." column shows the value of the threshold used and the "Within 30 days" column informs on the number of preventable SACT administrations given within 30 days of death. The "Within 90 days" column provides the number of SACT drug administrations that would have been preventable between 90 days and 30 days to death. The "Before 90 days" column informs on the number of SACT drug administrations that would have been preventable more than 90 days from death. The value in parentheses is the percent of corresponding administrations among the total number of administrations (Table 3). In the "Patients" section, the presented values inform on the number of corresponding patients in relation to the drug administrations, and values in parentheses are the percentage of corresponding patients among treated patients. "sum." stands for summary.

5. Discussion

5.1. Main findings

Five different machine learning models were compared for dynamic prediction of 30-day mortality in advanced and metastatic lung cancer patients, including two artificial neural network-based models and two tree-based ensemble models. The two tree-based ensemble models performed best and exhibited similar performances in terms of average precision. When used on the final test set, for the two models RF and GB, the ROC AUC increased, while the negative impact on AP was limited.

The performances of all models were only marginally impacted, even increasing in some cases, when using only the biochemical results. Inclusion of summary variables had no clear benefit.

A utility analysis was performed on the tree-based ensemble models, indicating that some late SACT administrations could be prevented by implementing the predictive model with a simple threshold-based rule while maintaining almost all treatments before the 90-day landmark.

5.2. Critical assessment

5.2.1. Study population

One of the main strengths of this study is the long inclusion timeframe of patients, allowing us to thoroughly investigate temporal performance of the models.

Additionally, the single centre study guaranties a high level of consistency and allows us to access detailed clinical information. On the one hand, this leads to questions regarding the generalizability of these results outside this centre since treatment implementations might vary among hospitals. On the other hand, there is high homogeneity in the Danish health care system that should allow an extension of these results to the entire country.

5.2.2. Biochemical results and generalizability

Biochemical results are potentially highly variable and capture information that can change significantly between the periods before and after the 30-day threshold. The fact that using only biochemical results did not significantly alter the performance would tend to indicate that limited information for this outcome is contained in the rest of the data and that there is no strong interaction with other variables. A simpler model using only these data should therefore be encouraged. It also indicates that these results can be broadly extended because they are less dependent on local practices. Additionally, since the most important biochemical results were not specific to lung cancer, these results could also be extended to other diagnoses sharing the same type of progression, such as pancreatic cancer. This is consistent with many studies investigating the prediction of mortality in cancer patients that primarily use biochemical results in their models[11,14,15,37].

Another major advantage of the biochemical results is the objectivity of the corresponding values, as they do not depend on the interpretation of clinicians.

Another key aspect is the validity of the model over time. Clinical practice changes over the years, notably with the introduction of new treatment options and better management of side effects so that a model fitted on old data might not be applicable to

current data in lung cancer patients. This was the reasoning for using patients who died in 2018 and 2019 as validation and test sets, respectively. For the tree-based ensemble models, AP was moderately impacted, while the ROC AUC was even improved. This improvement could be explained by more consistent data in recent years, i.e., on average, the 2010-2018 data were more similar to the 2019 data than the 2010-2017 data were to the 2018 data. The introduction of immune checkpoint inhibitors in Denmark between 2015 and 2017 could explain, at least in part, this higher similarity even though most of the information seems to lie in biochemical results, notably albumin and leukocytes, that should not have been impacted by this change. Another explanation is that more data in the training set helped the model to better capture the information available to make better predictions. Additionally, one reason the ROC AUC improved could have been the decrease in the frequency of 30-day mortality in the 2019 data (Table 2), which tends to improve the ROC AUC by lowering the false-positive rate at a specific sensitivity.

The stability of the prediction over time was also the reason behind the approach followed to evaluate the variability of the performance by bootstrapping the training set. Indeed, due to the changing nature of treatment procedures over time, we expect some variations in the training data as the predictive model is updated. Our goal was to avoid models whose performances are at risk of being largely impacted by small changes in the training dataset.

5.2.3. Using health care data registries

The primary interest in using electronic health records (EHRs) and administrative data is that they can be easily leveraged to build decision support tools. There are some issues using such data, notably informed presence bias[38,39], but in the current study, the patients were actively followed, and thus, limited differences in data availability were expected.

A potential limitation of using EHRs is the lack of reporting for potentially relevant data, i.e., data that cannot be used in the training phase. This was the case for performance status, which is a major resource in evaluating the survival time of patients in clinical practice. However, previous studies have shown that performance status is well correlated with certain biochemical results, such as C-reactive protein and albumin[40]. This correlation should nevertheless be confirmed in a more recent observational cohort study.

5.2.4. Summary variables and overfitting

Inclusion of summary variables to inform patient trajectories yielded mixed results. Some models benefitted from it, notably the RF model, while it had varying impact on other models, such as the GB model. Not including these variables could prevent overfitting in some cases. The same mechanism was potentially observed for nonbiochemical variables that could cause overfitting. An important aspect is how well these summary variables inform the trajectory of the patients. Indeed, each variable probably exhibits different dynamics near the end of life, which would require the design of specific summary variables.

5.2.5. Feature engineering vs. architecture optimization

The artificial neural network models did not perform well in this study, but it could be speculated that this was due to a poor choice of architecture, notably for the LSTM. While better performances could have likely been achieved by fine tuning the architecture

of the model, this was not realistic within the time constraints of this project. Furthermore, this extra tuning time could have also been used to perform much more extensive feature engineering for the other models, for example, by including interactions, designing specific summary variables that would better represent the trajectory of the patients, or optimizing other hyperparameters such as the learning rate for the GB model. Therefore, we decided to maintain relatively simple architectures for the LSTM and MLP to allow for a fair comparison to identify in which direction more effort should be placed.

5.2.6. Explainability and usability

SHAP values were used to assess the importance of each variable in the predictions from each model. However, the interpretability of these values is limited compared to that of the coefficient in a linear model such as LRENR. The explainability of nonlinear models such as tree-based models or artificial neural networks is the topic of ongoing research[41]. If explainability is of critical importance, LRENR should be prioritized.

The primary aim of this study was to introduce a predictive model capable of supporting clinical decision-making to ordinate SACT and thus potentially limit the risk of unnecessary harm. The crucial aspect of the models is their utility in practice. As shown in Table 4, the thresholds found across models and datasets for the considered rule changed extensively from 34.1% to 65.8%, implying that providing only predictions would leave much room for subjective assessment and might thus be challenging for clinicians to use. Therefore, effort should be concentrated into making the results as comprehensive as possible. Nevertheless, this study demonstrates that using a simple rule alongside the predictive model could limit the amount of SACT given too close to death. The rule cannot be implemented as is in a clinical context but could guide oncologists in their decision-making. Indeed, many parameters should be considered, notably the type of SACT. For example, protein kinase inhibitors such as osimertinib and alectinib typically have milder side effects while potentially avoiding flaring of the tumour, limiting the interest of stopping the treatment, even close to death. Conversely, etoposide and carboplatin often have much more severe side effects; therefore, a predictive model could help limit their use too close to death. These drugs are at high risk of being given close to death since they are used to treat small cell lung cancer, which is a rapidly progressing form of lung cancer. They are frequently given even in patients with late diagnosis and poor performance status, as sensitivity to treatment is usually high, resulting in good symptom control and long-lasting palliation; therefore, better selection of patients could have tangible results.

5.3. Comparison to other studies

Concerning model selection, other studies have also reported the typically good performance of gradient boosting in a similar context, as well as the often poor performances of artificial neural network-based models using tabular data[14,17,32,42]. Indeed, artificial neural network models are difficult to tune due to the number of hyperparameters and the instability of the optimization procedure and rarely outperform other approaches on structured data. Additionally, tree-based ensemble models outperform linear models by handling potential interactions between variables.

With respect to mortality prediction in cancer patients, work has already been done but in different contexts and with different endpoints. This includes simpler models[13,16] and more long-term endpoints, such as 6-month mortality[14]. Studies using 30-day mortality as endpoints typically take a more classical survival approach, i.e., predicting 30-day mortality from inclusion[11–13,17]. In the two studies implementing the RF approach and/or the GB approach[14,17], only a few biochemical variables were selected as important variables, probably due to the difference in endpoint and time of prediction. As opposed to most of the aforementioned studies, age was not selected as an important variable. This could be explained by the short-term nature of the prediction, where age might be of less importance than age for the prediction of more long-term survival.

5.4. Perspectives

The goal of this study was to investigate the possibility of constructing a decision support tool to avoid administering unnecessary SACTs in lung cancer patients. Additional work is needed to develop a prototype applicable in a clinical context and to conduct a prospective validation study. In practice, we envision a web server with a live connection to the EHRs and administrative data with a user-friendly web interface where clinicians can acquire an assessment of the risk of individual patient 30-day mortality by explicitly providing an identifier. The most recent data would be automatically retrieved from the relevant registries and used by the predictive model with an evaluation of the most important variables in that specific case. Such a solution could also be validated by a two-armed prospective validation study, randomized between active guidance of the where the clinician in the treatment arm will based on the predictive assessment of the patient compared and the control arm will be a standard decision without access to the data.
Once the solution is validated, long-term support and maintenance will be required to retrain the model to maintain an acceptable level of performance. This might require additional validation studies but is considered reasonable in terms of the potential benefit for clinical practice.

5.5. Conclusion

Prediction of 30-day mortality in patients with advanced lung cancer was most accurate using tree-based machine learning models on EHRs and administrative data. Most of the information was contained in biochemical parameters, limiting the interest of using other datasets for the prediction, such as comorbidities, disease trajectories, or histopathology. Using predictive modelling may potentially help to limit late SACT, reducing the risk of causing unnecessary harm to patients in the late stage of life.

Acknowledgements

We would like to thank the System Administrator of MedOnc, Annette Juul Madsen, Dept. of Oncology, and Special Consultant Thomas Mulvad Larsen, Business Intelligence Unit, North Denmark Region, for their help in obtaining and understanding the data needed for this study.

Funding sources and conflict of interest

This work was supported by the Department of Oncology, Aalborg University Hospital, the North Denmark Region's Health Innovation Foundation no. 2019-046332, and the Danish Cancer Research Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare, they have no other conflict of interest.

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
2. Taylor, C. W. & Kirby, A. M. Cardiac Side-effects From Breast Cancer Radiotherapy. *Clin. Oncol. (Royal Coll. Radiol.* **27**, 621–629 (2015).
3. Kappers, M. H., van Esch, J. H. M., Sleijfer, S., Danser, A. H. J. & van den Meiracker, A. H. Cardiovascular and renal toxicity during angiogenesis inhibition: clinical and mechanistic aspects. *J. Hypertens.* **27**, 2297–309 (2009).
4. Pabla, N. & Dong, Z. Curtailing side effects in chemotherapy: a tale of PKCδ in cisplatin treatment. *Oncotarget* **3**, 107–111 (2012).
5. Davis, C. Drugs, cancer and end-of-life care: a case study of pharmaceuticalization? *Soc. Sci. Med.* **131**, 207–14 (2015).
6. Li, T., Mizrahi, D., Goldstein, D., Kiernan, M. C. & Park, S. B. Chemotherapy and peripheral neuropathy. *Neurol. Sci.* **42**, 4109–4121 (2021).
7. Earle, C. C. *et al.* Evaluating claims-based indicators of the intensity of end-of-life cancer care. *Int. J. Qual. Heal. Care* **17**, 505–509 (2005).
8. Wallington, M. *et al.* 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study. *Lancet Oncol.* **17**, 1203–1216 (2016).
9. Christakis, N. A. & Lamont, E. B. Extent and determinants of error in doctors' prognoses in terminally ill patients: Prospective cohort study. *Br. Med. J.* **320**, 469–472 (2000).
10. Simmons, C. P. L. L. *et al.* Prognostic Tools in Patients With Advanced Cancer: A Systematic Review. *J. Pain Symptom Manage.* **53**, 962-970.e10 (2017).
11. Hamano, J. *et al.* A combination of routine laboratory findings and vital signs can predict survival of advanced cancer patients without physician evaluation: a fractional polynomial model. *Eur. J. Cancer* **105**, 50–60 (2018).
12. Adelson, K. *et al.* Development of imminent mortality predictor for advanced cancer (IMPAC), a tool to predict short-term mortality in hospitalized patients with advanced cancer. *J. Oncol. Pract.* **14**, e168–e175 (2018).
13. Renfro, L. A. *et al.* Clinical Calculator for Early Mortality in Metastatic Colorectal Cancer: An Analysis of Patients From 28 Clinical Trials in the Aide et Recherche en Cancérologie Digestive Database. *J. Clin. Oncol.* **35**, 1929–1937 (2017).
14. Parikh, R. B. *et al.* Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer. *JAMA Netw. open* **2**, e1915997 (2019).
15. Uneno, Y. *et al.* Development and validation of a set of six adaptable prognosis prediction (SAP) models based on time-series real-world big data analysis for patients with cancer receiving chemotherapy: A multicenter case crossover study. *PLoS One* **12**, 1–13 (2017).
16. Sjoquist, K. M. *et al.* Personalizing Survival Predictions in Advanced Colorectal Cancer: The ARCAD Nomogram Project. *J. Natl. Cancer Inst.* **110**, 1–11 (2017).
17. Elfiky, A. A., Pany, M. J., Parikh, R. B. & Obermeyer, Z. Development and Application of a Machine Learning Approach to Assess Short-term Mortality Risk Among Patients With Cancer Starting Chemotherapy. *JAMA Netw. open* **1**,

e180926 (2018).
18. Tomašev, N. *et al.* Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nat. Protoc.* (2021). doi:10.1038/s41596-021-00513-5
19. Kluyver, T. *et al.* Jupyter Notebooks—a publishing format for reproducible computational workflows. *Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016* 87–90 (2016). doi:10.3233/978-1-61499-649-1-87
20. Sundhedsdatastyrelsen. Disease Classification System - SKS (in Danish). Available at: https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer/klassifikation-sygdomme. (Accessed: 3rd March 2021)
21. WHO. ICD-10 Version:2016. (2016). Available at: https://icd.who.int/browse10/2016/en. (Accessed: 26th March 2020)
22. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* **40**, 373–83 (1987).
23. WHO Collaborating Centre for Drug Statistics Methodology. Anatomical Therapeutic Chemical (ATC) classification system. Available at: https://www.whocc.no/atc/structure_and_principles/. (Accessed: 3rd March 2021)
24. Joint Committee on Nomenclature, P. and U. (C-S.-N. of the I. and I. *et al.* Clinical laboratory sciences data transmission: the NPU coding system. *Stud. Health Technol. Inform.* **150**, 265–9 (2009).
25. O'Brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **41**, 673–690 (2007).
26. McKinney, W. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* **1**, 56–61 (2010).
27. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
28. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
29. Ho, T. K. Random decision forests. *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR* **1**, 278–282 (1995).
30. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, (2001).
31. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997).
32. Lauritsen, S. M. *et al.* Early detection of sepsis utilizing deep learning on electronic health record event sequences. (2019).
33. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning. Springer Series in Statistics* **26**, (Springer New York, 2009).
34. Chollet, F. & Others. Keras. (2015). Available at: https://keras.io/.
35. O'Malley, T. *et al.* Keras-tuner. (2019). Available at: https://github.com/keras-

team/keras-tuner.

36. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *NeurIPS Proc.* (2017).
37. Zhu, L. *et al.* A new prognostic score based on the systemic inflammatory response in patients with inoperable non-small-cell lung cancer. *Onco. Targets. Ther.* **9**, 4879–4886 (2016).
38. Verheij, R. A., Curcin, V., Delaney, B. C. & McGilchrist, M. M. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J. Med. Internet Res.* **20**, e185 (2018).
39. Phelan, M., Bhavsar, N. A. & Goldstein, B. A. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference. *eGEMs* **5**, 22 (2017).
40. Forrest, L. M., McMillan, D. C., McArdle, C. S., Angerson, W. J. & Dunlop, D. J. Comparison of an inflammation-based prognostic score (GPS) with performance status (ECOG) in patients receiving platinum-based chemotherapy for inoperable non-small-cell lung cancer. *Br. J. Cancer* **90**, 1704–1706 (2004).
41. Belle, V. & Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front. Big Data* **4**, 1–25 (2021).
42. Shwartz-Ziv, R. & Armon, A. Tabular Data: Deep Learning is Not All You Need. 1–11 (2021).

Supplementary materials

Supplementary Table 1. *SKS codes for comorbidities, symptoms, and side effects.*

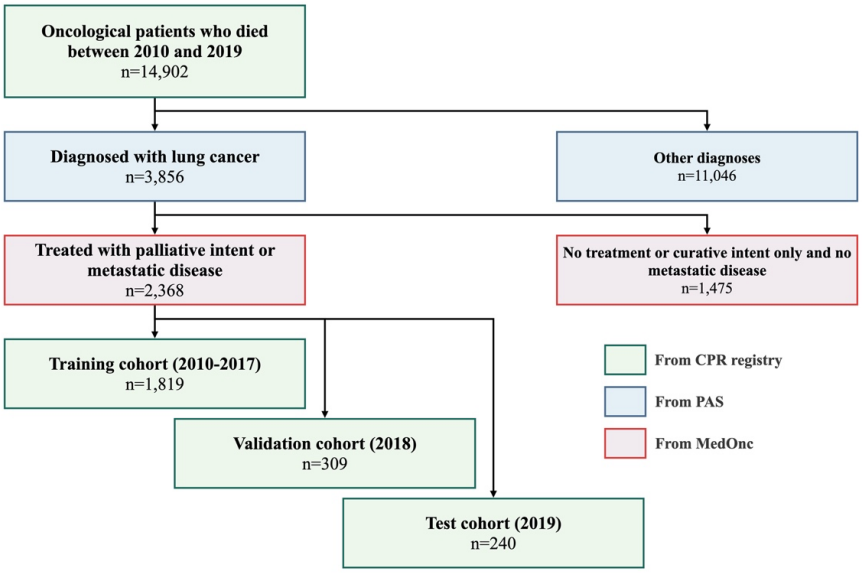
Comorbidity	SKS codes
Myocardial infarction	DI21, DI22, DI23
Congestive Heart Failure (CHF)	DI50, DI110, DI130, DI132
Peripheral vascular disease	DI70, DI71, DI72, DI73, DI74, DI77
Cerebrovascular Accident (CVA) or Transient Ischaemic Attack (TIA)	DI60, DI61, DI62, DI63, DI64, DI65, DI66, DI67, DI68, DI69, DG45, DG46
Dementia	DF00, DF01, DF02, DF03, DF051, DG30
Chronic Obstructive Pulmonary Disease (COPD)	DJ40, DJ41, DJ42, DJ43, DJ44, DJ45, DJ46, DJ47, DJ60, DJ61, DJ62, DJ63, DJ64, DJ65, DJ66, DJ67, DJ684, DJ701, DJ703, DJ841, DJ920, DJ961, DJ982, DJ983
Connective tissue disease	DM05, DM06, DM08, DM09, DM30, DM31, DM32, DM33, DM34, DM35, DM36, DD86
Peptic ulcer disease	DK221, DK25, DK26, DK27, DK28
Liver disease - Mild	DB18 K700, DK701, DK702, DK703, DK71, DK73, DK74, DK760
Liver disease - Moderate to severe	DB150, DB160, DB162, DB190, DK704, DK72, DK766, DI85
Diabetes mellitus - Uncomplicated	DE100, DE101, DE109, DE110, DE111, DE119
Diabetes mellitus - End-organ damage	DE102, DE103, DE104, DE105, DE106, DE107, DE108
Hemiplegia	DG81, DG82
Moderate to severe chronic kidney disease (CKD)	DI12, DI13, DN00, DN01, DN02, DN03, DN04, DN05, DN07, DN11, DN14, DN17, DN18, DN19, DQ61
Malignancy - Localized solid tumour	DC00-DC75 not finishing with a 'M', except DC44 (nonmelanoma skin cancer)
Malignancy - Metastatic solid tumour	DC76, DC77, DC78, DC79, DC80 or DC00-DC75 finishing with a 'M'
Malignancy - Leukaemia	DC91, DC93, DC93, DC94, DC95
Malignancy - Lymphoma	DC81, DC82, DC83, DC84, DC85, DC88, DC90, DC96
AIDS	DB21, DB22, DB23, DB24
Symptoms and side effects	**SKS codes**
Anaemia	Between DD50 and DD64
Appetite Loss	DR63, DF50
Bleeding and Bruising (Thrombocytopenia)	DD68, DD69
Constipation	DK590
Delirium	DF05
Diarrhoea	DK529, DK591
Oedema (Swelling)	DR60
Fatigue	DR53

Hair Loss (Alopecia)	DL65
Neutropenia	DD70
Bacterial infection	Between DA00 and DA79
Viral infection	Between A80 and B34
Lymphedema	DI89
Memory or Concentration Problems	DR41
Mouth and Throat Problems	DR682, DR070
Nausea and Vomiting	DR11
Nerve Problems (Peripheral Neuropathy)	DG900
Pain	DR52
Sexual Health Issues	DF52
Skin and Nail Changes	Between DL58 and DL60, DL62
Sleep Problems	DG47
Urinary and Bladder Problems	DN32

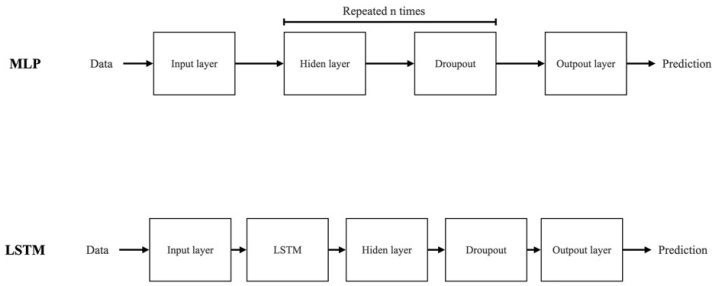
Supplementary Table 2. *Hyperparameter tuning for the artificial neural network-based models and values for the best trial from the Bayesian optimization.*

Hyperparameter	Domain	Best trial
MLP – Base dataset without summary variables		
Hidden layers with dropout	1 to 5	5
Neurons per hidden layer	10 to 400	310
Activation method for the hidden layers	softmax or relu or tanh or hard_sigmoid	tanh
Dropout ratio	0 to 0.9	0.318
Activation method for the result layer	softmax or sigmoid or hard_sigmoid	sigmoid
Optimizer	SGD or Adam or Adamax	Adam
Learning rate of the optimizer	10^{-6} to 10^{-2}	9.4×10^{-5}
MLP – Base dataset with summary variables		
Hidden layers with dropout	1 to 5	1
Neurons per hidden layer	10 to 400	290
Activation method for the hidden layers	softmax or relu or tanh or hard_sigmoid	tanh
Dropout ratio	0 to 0.9	0.305
Activation method for the result layer	softmax or sigmoid or hard_sigmoid	sigmoid

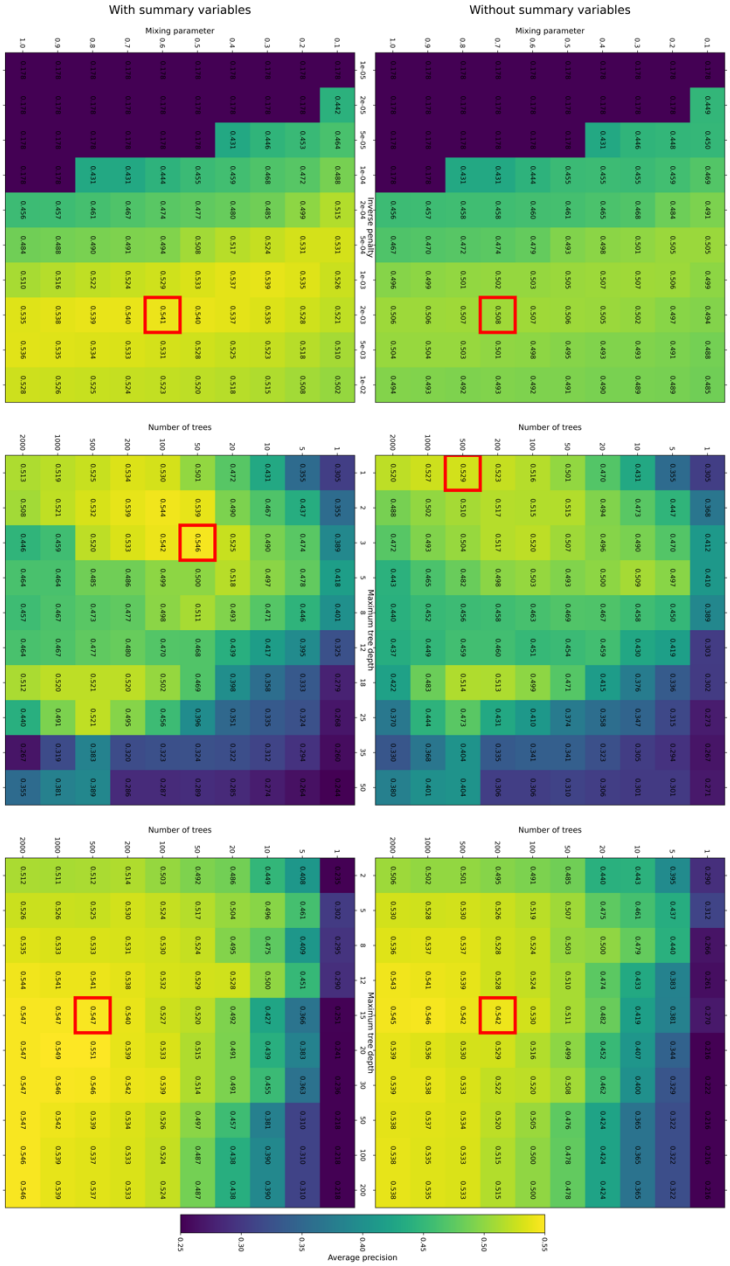
Optimizer	SGD or Adam or Adamax	SGD
Learning rate of the optimizer	10^{-6} to 10^{-2}	$1x10^{-2}$
LSTM		
Neurons for the LSTM part	10 to 400	380
Dropout for the LSTM part	0 to 0.9	0.0
Activation method for the LSTM part	tanh or softmax or relu or sigmoid or hard_sigmoid	tanh
Neurons for the hidden layer	10 to 400	250
Dropout for the hidden layer	0 to 0.9	0.116
Activation method for the hidden layer	hard_sigmoid or softmax or relu or tanh	tanh
Activation method for the result layer	softmax or sigmoid or hard_sigmoid or linear	sigmoid
Optimizer	Adamax or SGD or Adam	SGD
Learning rate of the optimizer	10^{-6} to 10^{-2}	$1.05x10^{-3}$



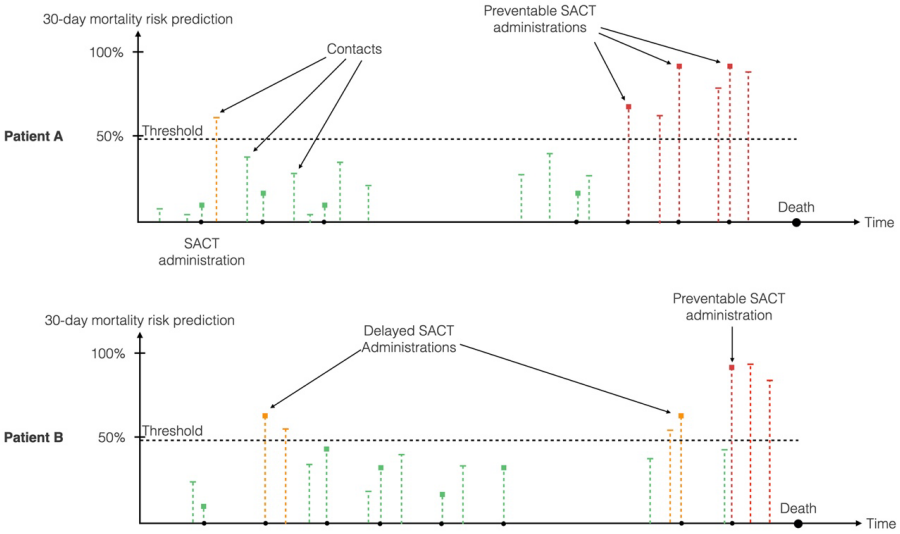
Supplementary Figure 1. *Inclusion workflow.*



Supplementary Figure 2. *Architecture of the MLP and LSTM models.*



Supplementary Figure 3. *Grid search results for the logistic regression with elastic net regularization (LRENR), the gradient tree boosting classifier (GB), and the random forest classifier (RF) models using the validation set (2018) after fitting on the training set (2010-2017). The values in the grids show the average precision of the fitted model using the corresponding hyperparameters, and a higher average precision corresponds to better performance. The red square indicates which hyperparameters were selected.*



Supplementary Figure 4. *Method to determine when SACT administration is preventable. Contacts for SACT administration are represented with dashed lines ending in squares, while other contacts are represented with dashed lines ending in lines. Green contacts are for contacts below the threshold, orange contacts are for contacts above the threshold but followed by contacts below the threshold, and red contacts are above the threshold with no later contacts below the threshold.*

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-955-4

AALBORG UNIVERSITY PRESS