

Use of healthcare databases for the evaluation of effectiveness of drugs used in patients with non-small cell lung cancer in Italy

Andrea Spini

► To cite this version:

Andrea Spini. Use of healthcare databases for the evaluation of effectiveness of drugs used in patients with non-small cell lung cancer in Italy. Human health and pathology. Université de Bordeaux, 2021. English. NNT: 2021BORD0390. tel-03868169

HAL Id: tel-03868169 https://theses.hal.science/tel-03868169

Submitted on 23 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CO-SUPERVISED THESIS PRESENTED TO OBTAIN THE QUALIFICATION OF

DOCTOR OF THE UNIVERSITY OF BORDEAUX

DOCTORAL SCHOOL UBX

École Doctorale Sociétés, Politique, Santé Publique Pharmacologie option Pharmaco-épidémiologie, Pharmacovigilance

Andrea SPINI

Use of healthcare databases for the evaluation of effectiveness of drugs used in patients with non-small cell lung cancer in Italy

Under the supervision of Francesco SALVO and the co-supervision of Marina ZICHE and Rosa GINI On 17 December 2021

Members of the examination panel:

Mr PARIENTE, Antoine	Professor, University of Bordeaux	President
Mrs EHRENSTEIN, Vera	Professor, University of Aharus	Reviewer
Mr MAZZAGLIA, Giampiero	Associate Professor, HDR, University of Milano	Reviewer
Mrs DONNINI, Sandra	Associate Professor, HDR, Univeristy of Siena	Membre
Mrs LUCENTEFORTE, Ersilia	Associate Professor, University of Pisa	Membre
Mr SALVO, Francesco	Associate Professor, University of Bordeaux	Directeur

Titre

Utilisation de bases de données de santé pour l'évaluation de l'effectivité des médicaments utilisés pour le cancer du poumon non à petites cellules en Italie.

Résumé

Le traitement du cancer du poumon avancé et métastatique a connu une transformation rapide au cours de la dernière décennie, en particulier pour l'histotype le plus fréquent, le cancer du poumon non à petites cellules (CPNPC). Après la chimiothérapie à doublet de platine, le traitement a rapidement évolué pour inclure les thérapies ciblées et l'immunothérapie. Les autorisations de mise sur le marché des thérapies anticancéreuses sont généralement basées sur les résultats d'essais contrôlés randomisés portant sur un faible nombre de patients, avec une courte durée d'exposition et des critères de sélection stricts. Il est essentiel de surveiller le profil d'utilisation des médicaments antinéoplasiques en vie réelle afin de produire des connaissances à partir de populations plus larges et non sélectionnées, d'identifier les différences d'utilisation entre des populations spécifiques (i.e. hommes et les femmes) et de compléter les informations provenant des essais cliniques randomisés, tant en termes d'efficacité que de sécurité.

Cette thèse avait pour but d'étudier le schéma de traitement des médicaments anticancéreux et la survie des patients atteints de CPNPC non résécable en vie réelle.

Trois études ont été réalisées pendant cette thèse.

Étude 1) Fournir une vue d'ensemble des études sur des données de santé collectées en routine et étudier le profil l'utilisation des immunothérapies et des thérapies ciblées chez les patients atteints de cancer du poumon. Une revue systématique a été menée et elle a mis en évidence un manque d'études menées en Europe (seulement 3/32), et évaluant les immunothérapies en première ligne (2/32). En plus, aucune étude n'a évalué le cancer du poumon à petites cellules. Ce travail servira comme point de départ pour la réalisation d'études en vie réelle tant pour la conception des études que pour l'analyse comparative des résultats.

Étude 2) Développer et valider un algorithme de recherche de cas pour l'identification des patients atteints de CPNPC dans le registre régional de pathologie de la région de Toscane. Un algorithme de recherche de cas a été développé pour l'application dans le registre régional de pathologie de la Toscane et a montré une valeur prédictive positive élevée (93%) et une sensibilité croissante au cours de la période d'étude (jusqu'à 65%). L'algorithme a donc été considéré comme adapté pour la conduite d'études pharmaco-épidémiologiques et épidémiologiques.

Étude 3) Décrire l'utilisation des thérapies anticancéreuses en première ligne et la survie des patients atteints de CPNPC non résécable au cours de la dernière décennie en Toscane. Une étude descriptive et rétrospective a été menée en utilisant l'algorithme mentionné précédemment pour l'identification des patients atteints de CPNPC. Environ 4 400 patients atteints d'un CPNPC primaire non résécable ont été inclus et l'utilisation des médicaments et la survie ont été évaluées sur la base du sexe et de l'histologie. L'étude a mis en évidence une utilisation différente des thérapies ciblées entre les sexes en fonction du profil biomoléculaire des femmes et des hommes. La survie s'est légèrement améliorée avec le temps et les patients ont également présenté un profil de survie différent par rapport au sexe et les histotypes de CPNPC.

En perspective, la cohorte continuera à être suivie et inclura davantage autres patients afin : i) d'analyser les changements de thérapie et les traitements en deuxième intention et ii) d'évaluer le profil de sécurité des nouvelles thérapies chez les patients atteints de CBNPC non résécable.

Mots clés : Cancer du poumon non à petit cellule; validation des données; mortalité

Title

Use of healthcare databases for the evaluation of effectiveness of drugs used in patients with non-small cell lung cancer in Italy.

Abstract

The treatment of advanced and metastatic lung cancer has undergone a rapid transformation over the last decade in particular for the most frequent histotype Non-Small Cell Lung Cancer (NSCLC). After the advent of platinum doublet chemotherapy, treatment has rapidly evolved to include targeted therapies and immunotherapies. The marketing authorizations of antineoplastic drugs are usually based on the results of randomized controlled trials with a low number of patients, a short duration of exposure and strict selection criteria.

Monitoring the real-world pattern of utilization of antineoplastic drugs is crucial to produce evidence from large and unselected populations, identify and address difference in the utilization between specific populations and to complete information derived from randomized clinical trials, both in terms of efficacy and safety.

This thesis aimed to study the pattern of treatment of antineoplastic drugs and the survival of patients with non-resectable NSCLC over the last year in real-life setting.

Three different studies were performed during the thesis period.

Study 1) To provide an overview of studies based on routinely collected electronical healthcare data studying drug utilization of immunotherapies and target therapies in patients with lung cancer. A scoping review was performed and highlighted a paucity of studies that were conducted in Europe (only 3/32) and assessing the use of a first-line immunotherapy (2/32). No studies evaluated small cell lung cancer. This work will serve as a starting point for future real-world studies on innovative lung cancer pharmacotherapies.

Study 2) To develop and validate a case-finding algorithm for the identification of NSCLC patients in the regional pathology registry of the Tuscany region. A case-finding algorithm was developed for the application in the regional pathology registry of Tuscany and showed a high positive predictive (93%) value and growing sensitivity over the study period (up to 65%). The algorithm was considered suitable for case identification in conduction pharmaco-epidemiological and epidemiological studies.

Study 3) To describe the utilization of anticancer therapies in first-line and the survival of patients with non-resectable NSCLC over the last decade in Tuscany. A descriptive, retrospective study using previous mentioned algorithm for the identification of NSCLC patients was conducted. About 4400 patients with a non-resectable, primary NSCLC were included and drug utilization and survival were evaluated on the basis of gender and histology. The study highlighted a different use of target therapies between sexes according to the known biomolecular profile of women and men. Survival slightly improved over time, but patients also presented gender difference in survival when diagnosed with different histotypes of NSCLC.

In perspective, the cohort will be followed and will include more patients in order to: i) analyze switches and second-line treatment of antineoplastic drugs in patients and ii) assess the safety profile of new therapies in patients with non-resectable NSCLC.

Keywords: Non-small cell lung cancer; data validation; overall survival

Synopsis (en français)

Introduction

En 2020, 2,2 millions de nouveaux cas de cancer du poumon ont été diagnostiqués dans le monde. L'incidence du cancer du poumon est deux fois plus élevée chez les hommes (environ 1,4 million de cas) que chez les femmes (environ 800 000 cas) et la plupart des personnes diagnostiquées avec un cancer du poumon avaient 65 ans ou plus. Le cancer du poumon est l'une des causes principales de décès liés au cancer dans les pays développés : la survie à 5 ans est inférieure à 18 %.

Le cancer du poumon peut être séparé en deux catégories histologiques : le cancer du poumon non à petites cellules (CPNPC), et le cancer du poumon à petites cellules (CPPC). La différence entre ces deux types de cancers est liée aux différentes caractéristiques morphologiques des cellules. Historiquement, les deux types de tumeurs étaient considérés comme des entités uniques ; aujourd'hui, grâce à l'avancée de nos connaissances sur leur biologie, nous savons que les deux classes comprennent des entités très hétérogènes. Le CPNPC représente 85 à 90 % des cancers du poumon et se compose de deux principaux sous-types histologiques : 1) le carcinome squameux et 2) le carcinome non-squameux.

L'évaluation du stade de la tumeur est cruciale pour le choix du traitement. Aux premiers stades (I, II et IIIa), la chirurgie est le traitement de choix du CPNPC. Au stade avancé/métastatique de la maladie (IIIb et IV), la chirurgie n'est plus une option et le traitement pharmacologique joue un rôle fondamental dans la prise en charge de ces patients.

Le traitement pharmacologique du cancer du poumon avancé et métastatique a connu une transformation rapide au cours de la dernière décennie, en particulier pour le CPNPC. Après la chimiothérapie à doublet de platine, le traitement a rapidement évolué pour inclure les thérapies ciblées et les immunothérapies.

Les thérapies ciblées sont des traitements pharmacologiques ciblant des protéines spécifiques (ex. récepteurs des tyrosines kinases), des gènes ou l'environnement tumoral. Les principales catégories de thérapies ciblées approuvées sont les anti-EGFR, anti-ALK, anti-BRAF et les anti-VEGF.

L'immunothérapie comprend des médicaments qui ciblent la protéine de mort programmée 1 (PD-1), une protéine de membrane cellulaire des cellules tumorales qui ralentit la réponse immunitaire et permettre ainsi à la tumeur de progresser. L'objectif de l'immunothérapie est d'améliorer/rétablir la capacité du système immunitaire à détecter et à détruire les cellules cancéreuses bloquant cette protéine (anti PD-1) ou son ligand (anti PD-L1).

Les autorisations de mise sur le marché des médicaments antinéoplasiques sont généralement fondées sur les résultats d'essais contrôlés randomisés portant sur un faible nombre de patients, une courte durée d'exposition et des critères de sélection stricts.

Par conséquent il est essentiel de comprendre le profil d'utilisation de ces médicaments en vie réelle afin de produire des connaissances à partir de populations non sélectionnées, et compléter les informations dérivées des essais cliniques randomisés, tant en termes d'efficacité quant en termes de sécurité.

Hypothèses et objectifs de la thèse

Nous avons répondu aux questions de recherche suivantes :

- Quelles sont les évidences disponibles dans la littérature concernant les études observationnelles basées sur des données électroniques de santé collectées de façon routinière (*routinely collected electronic healthcare data* - rcEHD) sur le cancer du poumon ?

- Est-il possible de mener une étude sur l'utilisation des médicaments chez les patients atteints de cancer du poumon non à petites cellules en utilisant les données italiennes disponibles ?

- Comment la pharmacothérapie du CPNPC a-t-elle évolué au cours de la dernière décennie ?

- La survie de ces patients est-elle améliorée dans un contexte de vie réel ?

Les objectifs de cette thèse étaient : 1) valider les données médico-administratives d'une base de données Italienne, 2) décrire l'évolution de l'utilisation des médicaments du CPNPC au cours de la dernière décennie, et 3) évaluer l'impact de l'introduction de ces médicaments sur la survie des patients atteints de CPNPC.

Études réalisées pendant la période de l'étude

Trois études différentes ont été réalisées pendant la période de la thèse.

Étude 1) Revue de la littérature

Les rcEHDs ont un potentiel énorme pour enrichir les informations sur les thérapies cibles et les immunothérapies utilisées pour traiter le cancer du poumon. L'objectif de ce travail était de fournir une vue d'ensemble des études basées sur les rcEHDs qui concernaient l'évaluation du profil d'utilisation des thérapies ciblées et des immunothérapies chez les patients atteints de cancer du poumon.

Les études éligibles publiées entre 2016 et 2020 dans PubMed et ISI Web of Science ont été recherchées. La source des données et les caractéristiques des études, ainsi que les données sur l'utilisation des médicaments et la survie ont été extraites.

Trente-deux études ont été incluses. Treize études ont lié plusieurs sources de données, notamment les données administratives, les registres du cancer, et les dossiers médicaux. Vingt-six études ont utilisé des données nord-américaines, tandis que trois ont utilisé uniquement des données européennes. Dans 29 études les informations relatives au cancer provenaient des dossiers médicaux/registres de cancer ; ainsi, dans 31 études les informations sur l'utilisation des médicaments ou la survie des patients provenaient ou des dossiers médicaux ou des bases de données administratives. La plupart des études concernaient des patients atteints de CPNCP (29 sur 32). Seules deux études décrivaient l'utilisation en première ligne des immunothérapies.

Cette revue de la littérature a fourni une vue d'ensemble structurée des études publiées basées sur le rcEHD qui ont examiné l'utilisation en vie réelle des thérapies ciblées et des immunothérapies chez les patients atteints de cancer du poumon. Les caractéristiques des études incluses dans cette revue ont montré que le *data linkage* de différentes sources de rcEHD est souvent nécessaire, voir indispensable. Les informations relatives au cancer (ex. histologie, stade) ont été principalement extraites des dossiers médicaux ou des registres du cancer, tandis que les informations sur l'utilisation des médicaments ou le statut vital ont été extraites dans la plupart des cas des dossiers médicaux ou des données administratives. En ce qui concerne les évidences sur l'utilisation de nouveaux médicaments, cette

revue a mis en évidence la rareté des études réalisées en Europe ainsi que celles concernant les immunothérapies, en particulier lorsque utilisées en première intention. Ce travail servira comme point de départ pour la réalisation d'études en vie réelle tant pour la conception des études que pour l'analyse comparative des résultats.

Cette revue de littérature a mis en évidence que les registres du cancer, lorsqu'ils sont liés à d'autres rcEHD, sont une source de données précieuse pour la réalisation d'études de pharmacoépidémiologie sur le cancer du poumon. Même si la région de Toscane dispose d'un registre régional du cancer, ces données ne sont joignables à d'autres sources de données, car l'identifiant du patient est différent de celui utilisé dans la base de données administrative de la région Toscane, rendant impossible sont utilisation pour les besoins de la thèse.

Néanmoins, la base de données de la région Toscane comprend plusieurs registres qui peuvent être liés entre eux par un code d'identification régional unique et pseudo-anonymisé. En plus des registres administratifs, la région Toscane possède également un registre régional d'anatomie pathologique (RRAP). Ce dernier couvre l'ensemble de la population de la toscane et contient un champ de texte libre ainsi que les codes SNOMED utiles pour l'identification de l'histologie du cancer du poumon. Une étude de validation du RRAP a donc été menée.

Étude 2) Étude de validation

L'objectif de cette étude était de développer et de valider un algorithme pour l'identification de cas de CPNCP dans le registre régional d'anatomie pathologique (RRAP) de la région Toscane en Italie.

Trois sources de données ont été utilisées : 1) le registre d'anatomie pathologique de l'hôpital universitaire de Sienne, 2) la base de données de la pharmacie hospitalier de Sienne et 3) la base de données administrative de santé de la région de Toscane.

Une première version de l'algorithme d'identification de cas CPNPC a été créée par un groupe de travail comprenant du personnel médical du département de pathologie de l'hôpital Universitaire de Sienne, des oncologues, des épidémiologistes, et des informaticiens. L'algorithme devait identifier les cas à partir d'une composante morphologique et une composante topographique. La composante morphologique a été développé pour identifier les histologies correspondantes au CPNPC, tandis que

la composante topographique visait à identifier le site pulmonaire. Pour ces deux composantes les codes SNOMED, ainsi que les mots-clés appliqués au champ de diagnostic en texte libre ont été utilisés. La première version de l'algorithme a été testé dans le registre d'anatomie pathologique de l'hôpital universitaire de Sienne entre le 01/01/2009 et le 31/08/2017.

Une fois les identifiants des patients anonymisés par l'autorité compétente, la valeur prédictive positive (VPP) de l'algorithme d'identification a été calculée sur la base des résultats d'un examen manuel fait en insu par deux évaluateurs sur les dossiers provenant d'anatomie pathologique de l'hôpital universitaire de Sienne. Tous les cas identifiés comme CPNPC par l'algorithme ont été ainsi liés à la base de données de la pharmacie hospitalière de Sienne, qui contient l'indication d'utilisation attribué par les cliniciens prenant en charge les patients, et qui a été utilisé pour estimer la sensibilité (SE) de l'algorithme.

Des modifications de l'algorithme ont ensuite été testées pour en améliorer les performances : la VPP a été calculée par rapport à un ensemble de données validées provenant du PR de Sienne ; la SE [minmax] a été estimée dans le RRAP de Toscane à l'aide de formules analytiques qui supposaient une incidence de CPNPC égale à 80 % ou 90 % de l'incidence globale du cancer du poumon en Toscane, à partir des données du registre des cancers. La version de l'algorithme présentant les meilleures performances a été choisie comme version finale et elle a été testé sur un échantillon aléatoire de 200 cas a été extrait du RRAP de Toscane pour une validation manuelle finale.

Dans le RRAP de Toscane, la version finale de l'algorithme présentait une SE qui augmentait avec le temps : entre 24,7 % et 28,0 % en 2009 et entre 57,9 % et 65,1 % en 2017. La VPP était par contre stable au cours du temps avec un valeur totale de 93 %.

L'algorithme final de détection des CPNPC a montré une VPP très élevée. La SE, même si avec des valeurs plus basses, a été considérée satisfaisante. Dans la période d'étude nous avons assisté à une amélioration de la transmission des données concernant les cancers des hôpitaux à la région Toscane au cours du temps, avec une tendance claire à l'augmentation au cours du temps calendaire. Étant donné la large utilisation de la terminologie SNOMED dans les rapports d'anatomie pathologique, l'algorithme proposé pourrait être directement adapté à d'autres bases registres d'anatomie pathologique.

Compte tenu de sa validité prometteuse, l'algorithme proposé pour la détection des CPNPC a été considéré comme adapté pour conduire la troisième étude de la thèse.

Étude 3) Prise en charge des patients avec CPNPC avancée par rapport au sexe des patients : étude d'utilisation des médicaments en première intention et de survie des patients.

La plupart des études retrouvées grâce à la revue de la littérature ont examiné l'utilisation des médicaments anticancéreux en fonction de l'histologie, seules quelques études rapportait des résultats stratifiés par sexe.

Les femmes et les hommes ont un profil biomoléculaire différent qui pourrait avoir un impact sur l'utilisation des médicaments et sur la survie des patients atteints de CPNPC. L'objectif de cette étude était de décrire la survie et l'utilisation des médicaments anticancéreux de première intention chez une cohorte de patients atteints de CPNPC avancé en fonction du sexe.

Une cohorte de patients avec un nouveau diagnostique de CPNPC avancé (non résécable) entre 2009 et 2019 (entrée de la cohorte) A partir du RRAP de la régionale Toscane a été constituée. Les antinéoplasiques délivrées jusqu'à 4 mois après l'entrée de la cohorte ont été classés en chimiothérapie standard, thérapies ciblées, immunothérapies, anticorps monoclonaux non définis ont été retrouvée dans la base de données administrative de la région.

Le traitement de première ligne et la survie des patients recevant un traitement médicamenteux ont été décrits. Les analyses ont été stratifiées en fonction de l'histologie, du sexe et de l'année d'entrée dans la cohorte.

Un total de 4 393 cas incidents de CPNPC avancé a été inclus. Les femmes atteintes d'un CPNPC non résécable ont reçu une thérapie ciblée plus fréquemment que les hommes (14,9 % contre 6,5 %). La fréquence d'utilisation de l'immunothérapie a varié entre 3,8 % en 2017 et 9,1 % en 2019. La proportion de patients n'ayant reçu aucun traitement médicamenteux dans les quatre mois suivant l'entrée dans la cohorte était stable pendant toute la période d'étude, et concernait environ 1/3 des patients atteints de CPNPC (36 % de squameux et 29 % de non-squameux).

La survie à deux ans des patients inclus dans la cohorte a augmenté au cours du temps : pour les CPNPC non squameux, il était de 22,3 % entre 2009 et 2011, et de 30,6 % entre 2018 et 2019, tandis

que pour les CPNPC squameux, elle était respectivement de 13,5 % et de 22,5 %. Parmi les CPNPC non squameux, la survie médiane était plus longue chez les femmes que chez les hommes (389 jours et 276 jours). Une augmentation de la survie globale et à 5 ans comparable a été ainsi retrouvé.

Cette étude a démontré que les données administratives de la région Toscane peuvent constituer un outil précieux pour étudier l'utilisation des médicaments anticancéreux et pour évaluer la survie des patients dans le domaine du cancer du poumon. En particulier, cette étude décrit les traitements en première intention des thérapies ciblées et des immunothérapies chez des patients atteints de CPNPC non résécable en vie réelle. Nos résultats mettent en évidence une utilisation différente des thérapies cibles entre les sexes : chez les femmes atteintes de CPNPC non squameux, les médicaments anti-EGFR étaient les thérapies ciblées les plus fréquemment utilisées. La survie des patients de la cohorte s'est améliorée au cours de la période d'étude, et les patients masculins et féminins présentent également un profil de survie différent selon qu'ils sont diagnostiqués avec un CPNPC squameux ou non squameux.

Perspectives

L'algorithme d'identification des cas de CPNCP dans le RRAP de la région Toscane s'est démontré prometteur et adaptée à la conduite d'études épidémiologiques et pharmacoépidémiologiques. Autres RRAP disponibles dans d'autres régions italiennes devraient être étudiés pour comprendre la reproductibilité de cet algorithme et ses mesures de validité. Cela permettra de collecter un plus grand nombre de patients atteints de NSCLC, et donc d'étendre l'étude d'utilisation à une cohorte plus grande.

Cette thèse a produit une connaissance actualisée de l'impact des nouvelles pharmacothérapies au cours de la dernière décennie sur la survie dès les patients atteints de CPNCP recevant au moins un traitement pharmacologique de première intention.

Quant à l'utilisation des médicaments, l'étude est, à notre connaissance, la première étude évaluant l'incidence des pharmacothérapies de première intention chez les patients atteints de CPNCP non résécable en Italie. Elle pourrait fournir des informations importantes à la communauté scientifique pour des futures analyses comparatives des résultats, et aux autorités de santé pour comprendre le profil d'utilisation de ces médicaments en vie réelle, dans un contexte d'évolution continue de la prise en charge des patients atteints de CPNPC. Cette étude pourrait être reproduite au cours des prochains années pour caractériser les utilisateurs, découvrir les différentes utilisations de ces médicaments en fonction de variables, telles que le sexe et étudier également les médicaments donnés en deuxième intention et suivantes.

En ce qui concerne la survie, la cohorte pourrait être suivie dans les années à venir pour inclure plus de patients. Étant donné que les bénéfices de l'immunothérapie sont plus évidents à long terme et que le pembrolizumab n'a été approuvé qu'en 2017 en première intention, une mise à jour de cette étude dans les prochaines années pourrait être utile pour comprendre l'évolution de la survie de ces patients. Une autre prospective de cette thèse pourrait comprendre l'évaluation de l'incidence des effets indésirables de ces médicaments, qui n'a pas encore été prise en compte. Cela pourrait être utile pour comprendre le profil de sécurité en vie réelle.

Research units

Host

University of Bordeaux, Centre de recherche Bordeaux Population Health, Inserm UMR 1219, Team of Pharmacoepidemiology and population impact of drugs, 156 Rue Leo Saignat, 33076 Bordeaux Cedex, France

Co-tutoring

University of Siena, Department of Life Sciences, Via Aldo Moro 2, Siena, Italy

Agenzia Regionale di Sanità della Toscana, Osservatorio di Epidemiologia, Via Pietro Dazzi, Firenze, Italy

Acknowledgements

First, I am extremely grateful to my supervisor, Prof. Francesco Salvo, for his continuous support and patience during these past three years. He gave me the passion and enthusiasm needed to pursue my research. It was his kind help and support that have made my study and life in France a wonderful time.

I would also like to thank my co-supervisors, Prof. Marina Ziche and Dr. Rosa Gini.

To Prof. Marina Ziche for her valuable guidance throughout my studies. Thank you for all the opportunities I was given to further my research. I'm lucky to have you as a mentor. To Dr. Rosa Gini whose expertise was priceless in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

Thank you for the opportunity you gave to me. I could not have imagined having better supervisors for my PhD study in these hard times.

I also want to thank Prof. Sandra Donnini, whose experience have encouraged me in my academic research and daily life. Your support has been precious. Thank you also to Dr. Giuseppe Roberto, for his scientific advice and knowledge. You provided many insightful suggestions. Your expertise has been fundamental in shaping my methods and discussions. Thank you also to Prof. Cristiana Bellan. Without your support it would not have been possible to complete most of my thesis.

Thank you for teaching me so much over these three years.

My sincere thanks also go to Prof. Antoine Pariente, who welcomed me in the research team of INSERM at the University of Bordeaux. Thank you for being part of my monitoring committee during these three years and for being part of the thesis jury.

I would like also to thank the thesis reviewers, Prof. Giampiero Mazzaglia and Prof. Vera Ehrenstein, for making me the honor of evaluating my thesis. And thanks also to Prof. Ersilia Lucenteforte, who has accepted to be part of the jury.

I would like to thank my colleagues and research team (Pietro, Jordan, Clement, Theo, Allison, Valerio, Arianna, Shirley and Prof. Lucia Morbidelli) for a cherished time spent together at the University of Bordeaux and at the University of Siena and in social settings.

Last but not least, to my family.

To my mum and dad who always stand by me, no matter what. You are always there for me. To Valentina, who holds my hand in every situation. I want to thank her for her critical sense and for share with me the little trivial moments of everyday life. To Edoardo who taught me in the last year more than I would admit. For his patience and for letting me play with him all the evenings with the balls of wool.

Without your tremendous understanding and encouragement in the past three years, it would be impossible for me to complete my study.

Finally, I could not have completed this thesis without the support of my whole family (Anna, Roberto, Tommaso, Manuele, Giuliana, Emma, Ivo, Mariarosa, Ivano and Andrea) and friends (Lorenzo, Stefano, Cosimo, Fausto, Luca, Gianluca, Giovanni, Sara), who put a smile on my face even when I was tired or homesick.

"Strana mania di volere il doppione di ogni cosa: del corpo l'anima, del passato il ricordo, dell'opera d'arte la valutazione, di sé stesso il figlio... Altrimenti, i primi termini ci parrebbero sprecati, vani. E i secondi allora? È perché tutto è imperfetto? O perché "si vedono le cose solo la seconda volta"?

C. Pavese

Abbreviations

ALK	Anaplastic lymphoma kinase
AHD	Administrative healthcare data
CTLA-4	Cytotoxic T-lymphocyte-4
BSC	Best supportive care
CI	Confidence interval
DFS	Disease free survival
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMA	European medicines agency
FDA	Food and drug administration
HDR	Hospital discharge records
HPS	Hospital Pharmacy database of Siena
ICIs	Immune check point inhibitors
IQR	Interquartile range
KRAS	Kirsten ras sarcoma virus (gene)
NSCC	Non squamous cell carcinoma
NSCLC	Non-small cell lung cancer
ODD	Outpatient drug dispensings
OED	Outpatient encounter and diagnostic procedure data
OS	Overall survival
PD-1	Programmed death protein 1
PD-L1	Programmed death ligand 1
PPV	Positive predictive value
PS	Performance status
rcEHD	Routinely collected electronic healthcare data
SCC	Squamous cell carcinoma
SCLC	Small-cell lung cancer
SD	Standard deviation
SE	Sensitivity
SNOMED	Systematized Nomenclature of Medicine
SNOMED CT	Systematized Nomenclature of Medicine – Clinical terms
TKIs	Tyrosine kinase inhibitors
ТМВ	Tumor mutational burden
uMAB	Unspecified monoclonal antibodies
US	United States
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor

SUMMARY

Résumé (FR) / Abstract (EN) Synopsis (FR) Research units Acknowledgements Abbreviations

CHAPTER 1. Thesis introduction and objectives

1.1	Incidence, mortality and risk factor of lung cancer	1
1.2	Histological classification	2
1.3	Staging and diagnosis	4
1.4	Pharmacological treatments in lung cancer	6
	1.4.1 Standard chemotherapy	6
	1.4.2 Target therapies	7
	1.4.3 Immunotherapy	14
1.5	Therapeutic approach	17
	1.5.1 Early stages: NSCLC and SCLC	17
	1.5.2 Advanced/metastatic stages (IIIB/IV): NSCLC	18
1.6	1.5.3 Extensive/metastatic stages (III/IV): SCLC	21
1.6	Value of conducting pharmacoepidemiology studies in lung cancers	23
1./	Datasources to conduct pharmacoepidemiology studies in oncology	25
		• •
CHAPTER	2. Scoping review of the literature: using routinely collected electro	nical
healthcare	data to study drug utilization in patients with lung cancer	17
2.1	Introduction	29
2.2	Study objectives	29
2.3	Publication	30
CHAPTER	3. Identification and validation of a case-finding algorithm to extra	ct
NSCLC cas	ses from the regional pathology registry of Tuscany	53
3.1	Introduction	54
	3.1.1 Pathology registries to conduct observational studies in oncology	54
	3.1.2 SNOMED coding in pathology registries	55
3.2	Study objectives	56
3.3	Publication	57
CHAPTER	A 4. Drug utilization and survival of patients with non-resectable NS	CLC: a
real-word s	study with gender perspective	70
4.1	Introduction	71
	4.1.1 Gender difference in cancer	71
	4.1.2 Gender difference in NSCLC	73
4.2	Study objectives	74
4.3	Publication	75
CHAPTER	3. 5. Thesis conclusion and perspectives	96
5.1	Review of the existing literature	97
5.2	Validation study	98
5.3	Study using the administrative database of the Tuscany region	99
	5.3.1 Drug utilization study	99
	5.3.2 Survival	100
6. BIBLIO	GRAPHY	102
7. PUBLIC	ATIONS	119
7.1	Publications and congress communications related to thesis	120
7.2	Other publications and communications	123

CHAPTER 1. Thesis introduction and objectives

1.1 Incidence, mortality and risk factor of lung cancer

In 2020, about 2.2 million of new cases of lung cancer were registered worldwide (11.4% of all cancer sites).¹ The incidence of lung cancer is roughly two times higher in men (about 1.4 millions of cases) than in women (about 800k cases) and most people diagnosed with lung cancer are 65 or older.

Figure 1.1 shows the age standardized incidence rates by sex for lung cancer among men and women in 2020 worldwide. The tobacco epidemic deeply affected the variation among countries in lung cancer rates and trends: its consumption caused an increasing incidence, first in men and later in women.^{1,2} The Italian National Institute of Statistic reported between 1970-2011 a progressive reduction of tobacco consumption among males (from 55% to 28%) with a constant increase among female (from 12% to 17%).² As a result of this sex-specific trend, incidence rates among women are approaching or equaling those among men in several industrialized countries in Europe and Northern America.¹

As for mortality, lung cancer is one of the leading causes of cancer-related death in developed countries: more than half of people die within a year of diagnosis and its five-year survival is less than 18%. The EUROCARE-5 study reported a five-year survival of 13% for all lung cancer patients diagnosed in 2000–2007, with a range from 9% in the UK to 15% in central Europe.³ Another study conducted in the United States (US) reported that the five-year survival was lower than 21% in 2015.⁴ Smoking is the main environmental risk factor. According to the guidelines of the European Society of Molecular Oncology, quitting smoking before the age of 40 reduces the risk of lung cancer in the next 10-15 years.⁵ Environmental risk factors also include passive smoking, occupational exposures (e.g. radon and asbestos),⁶ pre-existing lung diseases such as pulmonary fibrosis,⁷ Hodgkin's lymphoma,⁸ human immunodeficiency virus infection and alcohol consumption.^{9,10} Residential and

occupational exposure to radon (radioactive noble gas) is the second leading cause of lung cancer after cigarette smoking.¹⁰



countries. CA Cancer J Clin. 2021: 71: 209- 249

1.2 Histological classification

Lung cancer can be divided into two histological categories (**Figure 1.2**): non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). The difference between the two tumor types lies in the different morphological characteristics of the cells.



Both types of tumors, until recently, were considered as single entities; today, thanks to an advance in our knowledge of its biology, we know that the two classes comprise very heterogeneous entities.¹¹ Advances in genomics facilitated the understanding of intra-tumor heterogeneity and genetic alterations specific to these two classes of lung cancer.^{12,13} SCLC has neuroendocrine and epithelial features and is predominantly centrally located tumor. It accounts for approximately 10-15% of lung cancers and given is low prevalence is considered a rare disease (prevalence: 1-5 cases out of 10.000).

NSCLC represents 85-90% of the lung cancers and consists of two main histological subtypes:¹⁴

<u>Squamous cell carcinoma</u>: accounts for 25-30% of lung cancers and the main risk factor is tobacco smoke. Squamous cell lung cancers usually occur in the central part of the lung or in the main airways, such as the left or right bronchus.

Non-Squamous cell carcinoma, which could be further distinguished in:

- Adenocarcinoma: approximately 40% of all lung cancers are adenocarcinomas and is the most common subtype diagnosed in people who have never smoked. Adenocarcinoma arises from cells lining the mucus-secreting airways and usually occurs in the peripheral region of the lung.
- *Large cell carcinoma*: this type accounts for 10-15% of lung cancer, and it is named for the size of the cells observed under the microscope. It has poor differentiation and cannot be further investigated by immunohistochemistry or electron microscopy.

1.3 Staging and diagnosis

Assessing the stage of the tumor is a multidisciplinary process which involves imaging, endoscopy exams and surgical procedures.¹⁵ The staging system most often used for lung cancer is the **TNM** system, which is based on three axis:

- The size and extent of the main **tumor** (**T**)
- The spread to nearby lymph **nodes** (**N**)
- The spread (metastasis) to distant sites (M)

Once the T, N, and M categories have been determined, the overall stage is assigned (**Figure 1.3**).¹⁶

T / M	Subcategory	NO	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	ШВ
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	Mla	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

Figure 1.3. The eight edition TNM stage classification for lung cancer Source: Detterbeck FC. The eighth edition TNM stage classification for lung cancer: What does it mean on main street? *J Thorac Cardiovasc Surg.* 2018 Jan;155(1):356-359

Assessing the stage of the tumor is crucial for its treatment. Since patients might be totally asymptomatic when lung cancer is not advanced, only 25-30% of lung cancer cases are diagnosed in the early stages (I-IIIA).¹⁷ In stages up to III (IIIA for NSCLC, IIB for SCLC) surgical therapy has a curative intent: patients can undergo surgery and then be evaluated according to the overall situation on lobar and lymph node extension and undergo adjuvant chemotherapy. Starting from III stage (IIIB for NSCLC, IIIA for SCLC), patients are classified as having an advanced/extensive tumor and surgery is no longer a possible option.¹⁴

1.4 Pharmacological treatments in lung cancer

1.4.1 Standard chemotherapy

Standard chemotherapy prevents tumor cell multiplication by inducing their death and by interfering with the mechanisms involved in proliferation. Tumor cells multiplicate more rapidly than normal cells, so the effect of standard chemotherapy is mainly observed in tumors with a fast growth. Until 2005, standard chemotherapy has been the only pharmacological treatment available to treat lung cancer. The most common drugs used to treat lung cancer include carboplatin/cisplatin, docetaxel, gemcitabine, nab-paclitaxel, paclitaxel, pemetrexed and vincristine/vinorelbine.¹⁴

The platin compounds (carboplatin and cisplatin) act directly on the DNA of tumor cells which undergo apoptosis. Gemcitabine is a cytidine nucleoside analogue which act by being incorporated into the DNA (or RNA) of the tumor cell and thus promoting cell death. Also pemetrexed is converted to an active form that blocks the activity of enzymes involved in the production of nucleotides. Docetaxel, paclitaxel as also vincristine and vinorelbine act during mitosis process (cell division), thus blocking its multiplication.¹⁸

These drugs could be used in regimens of two or three drugs, or one single drug based on patient's performance status and tumor stage. Currently, platinum-based compounds have been combined with third-generation cytotoxic agents (paclitaxel, gemcitabine, docetaxel and vinorelbine) and this combination is called *platinum doublet*.¹⁹ The introduction of the pemetrexed-cisplatin doublet into treatment has demonstrated a clinical benefit in non-squamous NSCLC compared to gemcitabine-cisplatin treatment, thus enhancing the importance of histology identification.²⁰ Platinum-based chemotherapy remains the standard of care for the majority of patients with advanced NSCLC.

1.4.2 Target therapies

In the 1980s, discoveries in the fields of cell biology, immunology, and molecular biology enabled researchers to better understand the molecular mechanisms responsible for the neoplastic transformation of cells. As a result, new molecular targets were identified.²¹ A major limitation of standard chemotherapy is its lack of selectivity which indiscriminately affects all rapidly reproducing cells, both neoplastic and normal ones. In contrast, target therapy is more selective, as it targets mainly tumor cells and significantly reducing side effects. The targets identified are mainly cancer's specific proteins (i.e., tyrosine kinase receptors), genes, or tissue environment.

Receptors with tyrosine kinase activity are transmembrane proteins involved in intracytoplasmic signal transduction from the extracellular environment. These receptors play an important role in a variety of cellular processes including growth, motility, differentiation, and metabolism.²² Activation of these receptors can induce tumor cell proliferation and growth, suppress apoptosis and promote angiogenesis and metastatic spread. The self-activation of tyrosine kinases linked to the acquisition of genetic abnormalities during oncogenesis is a frequently observed phenomenon. The most important molecular target identified in lung cancer are the epidermal growth factor receptor (EGFR), the anaplastic lymphoma kinase (ALK) and the Kirsten Ras Sarcoma Virus gene (KRAS). Other oncogenic genes with lower incidence are BRAF and ROS1.²³

Tyrosine kinase inhibitors (TKIs) are target therapies that block the activation of the catalytic sites of these enzymes: by doing so, the downstream cell signaling pathways are disrupted, thereby restoring control over cell proliferation and survival. However, difference between SCLC and NSCLC must be underlined.

While SCLC genotyping still need to be further investigated and no target therapies are available for this type of lung cancer, as for NSCLC the discover of molecular alterations

7

permitted in the last years the approval of a considerable number of target therapies. However, also for NSCLC, effective molecularly targeted therapies have disproportionately impacted non-squamous cell carcinoma compared to squamous cell carcinomas and the most frequent genetic alterations in lung adenocarcinomas are shown in **Figure 1.4**.



Anti-EGFR

EGFR is a well-studied target for target therapy of which there are four different isoforms identified as EGFR1, EGFR2, EGFR3 and EGFR4. EGFR is a member of the family of protein kinases widely distributed in the human body. EGFR was chosen as target because these receptor types are present in all epithelial cells and in tumor cells, where they are over-expressed. The signaling of EGFR, activated by the Epidermal Growth Factor (EGF), is important for cell growth and activated second messengers responsible for the induction of

apoptosis resistance mechanisms and the formation of soluble factors crucial for tumor angiogenesis and metastasis (**Figure 1.5**).



EGFR mutations are more prevalent in females than in males and more frequent in the histological type of adenocarcinoma. Three generation of drugs aimed at inhibiting the kinase activity of EGFR have been designed. For NSCLC patients with EGFR exon 19 deletions or an exon 21 mutation, the standard first-line treatment is first-generation TKIs (gefitinib, erlotinib), which can bind reversibly the binding site, while second-generation TKIs (afatinib, dacomitinib) are irreversible inhibitors. Many patients may develop resistance to this type of treatment, and the most common resistance is the development of the T790M substitution, which leads to mutation of the tyrosine kinase domain in exon 20. Osimertinib has been approved for patients with T790M-positive NSCLC with acquired resistance to EGFR tyrosine kinase inhibitors (**Table 1.1**).²⁴

Drugs	Indication	Posology	Approval date EMA
Erlotinib Tarceva®	EGFR+, Advanced metastatic NSCLC	150 mg, os daily	3 November 2005, second-line 14 November 2011, first-line
Gefitinib Iressa®	EGFR+, Advanced metastatic NSCLC	250 mg, os daily	24 June 2009
Afatinib Giotrif®	EGFR+, Advanced metastatic NSCLC Advanced metastatic NSCLC, squamous histology	40 mg, os daily	25 September 2013
Osimertinib Tagrisso®	EGFR+, Advanced metastatic NSCLC EGFR+, T790M mutation, Advanced metastatic NSCLC	80 mg, os daily	02 February 2016
Dacomitinib Vizimpro®	EGFR+, Advanced metastatic NSCLC	45 mg, os daily	25 June 2019

Table 1.1. Anti-EGFR drugs approved by EMA for NSCLC

Anti-ALK

The ALK gene is located on the short arm of chromosome 2, and the ALK protein is a transmembrane tyrosine kinase receptor. The resulting tyrosine kinase activates several downstream signaling pathways responsible for cell proliferation and survival.²⁵

ALK rearrangements occur in approximately 4-5% of NSCLC cases and this type of rearrangement is more common in NSCLC patients diagnosed with adenocarcinoma.¹⁴ The incidence between women and men of ALK-positive NSCLC is very similar worldwide.

ROS1 is an integral membrane protein with tyrosine kinase activity that promotes signaling pathways involved in cell growth, proliferation, and survival. Rearrangements in the ROS1 gene promote its fusion with various other genes, leading to activation of tyrosine kinase, which in turn activates other signaling pathways.

ROS1 rearrangements account for 1-2% of NSCLC cases and are mainly observed in women, younger than 50 years, those with no smoking history and those with non-squamous

histology.^{26,27} The homology between the tyrosine kinase domains of ROS1 and ALK allows the definition of unique subgroups of patients highly sensitive to TKIs from the crizotinib family.

EMA has approved the following drugs for these targets: crizotinib, alectinib, ceritinib and brigatinib (**Table 1.2**).

Drugs	Indication	Posology	Approval date EMA	
Crizotinib	ALK+, Advanced metastatic NSCLC	250 mg, os two times	23 October 2012	
лакоп®	ROS1+, Advanced metastatic NSCLC	each day		
Ceritinib Zykadia®	ALK+, Advanced metastatic NSCLC previously treated with crizotinib	450 mg, os daily	06 May 2015	
A 1	ALK+, Advanced metastatic NSCLC		16 February 2017	
Alectinib Alecensa®	ALK+, Advanced metastatic NSCLC previously treated with crizotinib	600 mg, os two times each day		
Duinatin'ih	ALK+, Advanced metastatic NSCLC	90 mg, os daily in the	22 November 2018	
Brigatinib Alunbrig®	ALK+, Advanced metastatic NSCLC previously treated with crizotinib	first 7 days, then 180 mg daily		
Lorlatinib Lorviqua®	ALK+, Advanced metastatic NSCLC, second- third- line	100 mg, os daily	17 June 2019	

Table 1.2. Anti-ALK / ROS-1 drugs approved by EMA for NSCLC

Anti-BRAF

The BRAF kinase is a key factor in intracellular signaling that facilitates signal transmission from the cell surface to the nucleus after EGFR activation. BRAF is part of the mitogenactivated protein kinase pathway that plays an important role in the control of cell growth and regulation. In healthy tissues BRAF kinase is deactivated by negative feedback once the signal has passed to the next step in the cascade. BRAF activating mutations behave as oncogenic promoters that can lead to uncontrolled cell growth and proliferation.²⁸ In patients with NSCLC, BRAF mutations occur in approximately 2-4% of patients with lung adenocarcinoma.²⁹ Recently the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) approved the use of a BRAF inhibitor, dabrafenib in combination with trametinib, in patients with NSCLC with a BRAF V600E mutation (**Table 1.3**).¹⁴

Approval date EMA Drugs Indication Posology BRAF V600+, Advanced metastatic Dabrafenib 150 mg, os two times NSCLC in combination with 29 March 2017 each day Tafinlar® Trametinib BRAF V600 +, Advanced metastatic Trametinib NSCLC in combination with 2 mg, os daily 27 March 2017 Mekinist® Dabrafenib

Table 1.3. Anti-BRAF drugs approved by EMA for NSCLC

Anti-VEGF

Tumor cells release growth factors, which act by activating pre-existing vessels and cause these to proliferate new vessels directed towards the tumor mass providing all the necessary elements for growth and survival. Indeed, the formation of these new blood vessels from normal ones also leads to a poorer clinical prognosis, and reduced overall survival.³⁰

The vascular endothelial growth factor (VEGF) is the most specific and powerful mitogenic factor for endothelial cells, being also the most important determinant of cell survival in newly formed vessels. It causes an increase in vessel permeability, a key factor in making it easier for tumor cells to enter the circulatory stream, leading to the phenomenon of metastasis.³¹

The bevacizumab antibody binds VEGF-A and was the first angiogenesis inhibitor approved for the first-line treatment of advanced NSCLC patients in combination with platinum-based chemotherapy. Binding to VEGF results in the inability of VEGF to bind to its receptors (Vascular Endothelia Growth Factor Receptors: VEGFR-1 and VEGFR-2) on the surface of endothelial cells. This interaction of bevacizumab with VEGF causes blockage of the biological activity of VEGF, thus leading to regression of the tumor vascularization process and normalization of the residual neoplastic vascularization. Ramucirumab, on the other hand, is a monoclonal antibody binding the VEGFR, which together with nintedanib, a multi-tyrosine kinase anti-angiogenic inhibitor, is approved for the second-line treatment of NSCLC in combination with docetaxel (**Table 1.4**).^{32,33}

Drugs	Indication	Posology	Approval date EMA	
Bevacizumab	Advanced metastatic NSCLC, non- squamous histology, in addition to platinum chemotherapy	7,5 o 15 mg/kg, IV every three weeks	21 Amount 2007	
Avastin®	EGFR+, Advanced metastatic NSCLC, non- squamous histology, in combination with erlotinib 15 mg/kg, IV every three weeks		21 August 2007	
Nintedanib Vargatef®	Advanced metastatic NSCLC, non- squamous histology, in combination with docetaxel after previous chemotherapy with platinum	200 mg, os two times each day	08 January 2015	
Domusimumsh	EGFR+, Metastatic NSCLC, in combination with erlotinib	10 mg/kg, IV every two weeks	25 January 2016	
Kainucirumao Cyramza®	Advanced metastatic NSCLC, in combination with docetaxel, after previous chemotherapy with platinum	10 mg/kg, IV every three weeks.		

Table 1.4. Anti-VEGF drugs approved by EMA for NSCLC

1.4.3 Immunotherapy

The identification of immune checkpoints (i.e., molecules that allow the tumor to evade the normal control of the immune system) allowed the discovery of immunotherapy. Such pharmacological treatment includes drugs that target the Cytotoxic T-lymphocyte-4 (CTLA-4) and the ligand and the receptor of the programmed death protein-1 (PD-L1 and PD-1, respectively).^{34,35} The drugs included in this pharmacological category are also called *immune check point inhibitors* (ICIs).

The PD-1 receptor is a checkpoint-immune receptor expressed on the surface of activated T cells and acts mainly at the peripheral level, i.e. in the tumor microenvironment, and plays a role in immune regulation: its ligands PD-L1 and PD-L2 are expressed on the surface of dendritic cells or macrophages.^{36,37} Both PD-L1 and PD-L2 are expressed at relatively low levels in healthy tissues. When PD-L1 binds to PD-1 triggers a regulatory mechanism that causes the inhibition of T cells, which are thus inactivated. The healthy cell, in this condition, is not attacked by the immune system because the effector function of the T cells has been inhibited. Thus, the expression of PD-L1 and PD-L2 mediate the natural immune tolerance and prevent autoimmune responses and tissue damage after an inflammatory response. Cancer cells, on the other hand, have high levels of PD-L1 expression on their surface, allowing PD-1 activation on all T cells that have infiltrated the tumor microenvironment, effectively switching off these cells (**Figure 1.6**).

Expression and overexpression of PD-L1 has been documented in several tumor types, including NSCLC. The activation of key oncogenic pathways, including phosphoinositide 3-kinase and MAPK as also EGFR and BRAF activating mutations, can increase PD-L1 expression. The upregulation of PD-L1 expression levels has been shown to be 35% to 95% in NSCLC and such high levels of PD-L1 expression have been linked to poor clinical outcomes.^{38,39}



The aim of immune checkpoint inhibitors is to improve/restore the immune system's ability to detect and destroy cancer cells by overcoming the mechanisms the tumors evade to suppress the immune response.⁴⁰ Immune checkpoint inhibitors which target this interaction, have been introduced into lung cancer therapy. To date, four anti-PD-1/PD-L1 immune checkpoints inhibitors have been approved for advanced non-squamous or squamous NSCLC: pembrolizumab and nivolumab against PD-1 and durvalumab and atezolizumab for PD-L1. Also ipilimumab (an anti-CTLA-4 drug) in combination with nivolumab and two cycles of platinum based-chemotherapy is approved for metastatic NSCLC without activating EGFR mutations or ALK rearrangements (**Table 1.5**).²⁴ In general, about only 20% of NSCLC patients have a durable response to ICIs in monotherapy. Therefore, it is important to identify biomarkers to establish valid prognostic and predictive factors for treatment response.³⁸

Tabl	e 1	.5.	Immune	chec	kpoint	inl	nibito	rs (immunothe	erapy)	approv	ved b	v EM	A for	NSCLC	
								,	(/			

Drugs	Indications	Posology	Approval date EMA	
	Metastatic NSCLC, without activating EGFR mutations or ALK rearrangements, in combination with ipilimumab and 2 cycles of platinum-based chemotherapy, first-line	360 mg, IV every three weeks	15 June 2015	
Nivolumab Opdivo®	Advanced metastatic NSCLC, as monotherapy, second-line	3 mg/kg, IV every two weeks		
	Metastatic SCLC with progression after platinum-based chemotherapy and at least one other line of therapy	240 mg, IV every two weeks		
	PD-L1 \geq 50 %, Advanced or metastatic non-squamous NSCLC, without activating EGFR mutations or ALK rearrangements, in monotherapy, first-line		17 July 2015	
Pembrolizumab	Advanced or metastatic non-squamous NSCLC, without activating EGFR mutations or ALK rearrangements, in combination with pemetrexed and platinum-containing chemotherapy, first-line	200 mg, IV every three weeks		
Keytruda®	Advanced or metastatic NSCLC, squamous, in combination with carboplatin and paclitaxel or nab- paclitaxel, first-line			
	PD-L1 \geq 1 %, Advanced or metastatic NSCLC, in monotherapy, second-line			
	PD-L1 \geq 50%, Metastatic NSCLC, without activating EGFR mutations or ALK rearrangements, first-line		29 September 2017	
Atezolizumab	Advanced metastatic NSCLC, second-line	1200 mg, IV every three weeks		
Tecentriq®	Extensive stage SCLC in combination with etoposide and carboplatin or cisplatin, first-line			
	Advanced metastatic NSCLC, without activating EGFR mutations or ALK rearrangements, in combination with first-line chemotherapy	840 mg, IV every three weeks		
Durvalumab Imfinzi®	PD-L1 \geq 1%, Advanced NSCLC, monotherapy, second-line	10 mg/kg, IV every two weeks		
	Extensive stage SCLC in combination with etoposide and carboplatin or cisplatin, first-line	1500 mg, IV every three weeks	21 September 2018	
Ipilimumab Yervoy®	Metastatic NSCLC, without activating EGFR mutations or ALK rearrangements, in association with nivolumab and with two cycles of platinum-based chemotherapy	1 mg/kg ipilimumab, IV + 360 mg nivolumab, IV every three weeks	17 September 2020	

1.5 Therapeutic approach

1.5.1 Early stages: NSCLC and SCLC

In the early stages surgery is the treatment of choice, while pharmacological treatment, such as neoadjuvant or adjuvant therapy, can be used before or after tumor resection, respectively. If patients are diagnosed at an early stage, surgical resection of lung cancer allows a favorable prognosis; in 2017, five-year survival rates were 70-90% for localized tumors (stage I).⁴¹ Lobectomy is considered the standard surgical method for early stage NSCLC, however not all patients may be candidates. Some patients, even if technically resectable, refuse surgery or are considered medically inoperable due to insufficient respiratory reserve, cardiovascular disease or comorbidities and general frailty.

As for pharmacological treatment, neoadjuvant therapy is administered prior to surgery and increased chances of complete resection. Anyway, the percentage of lung cancer patients who are candidates for surgery has remained around 30% for the last 20 years. The neoadjuvant treatment regimen usually consists of 3-4 cycles of cisplatin combined with one of the cytotoxic drugs.¹⁴

After complete resection, a substantial proportion of these patients eventually develop local recurrence or distant metastases. Systemic relapses occur due to the presence of micro-metastatic disease at the time of surgery.¹⁷

Therefore, sometimes an adjuvant treatment is needed to reduce lung cancer recurrence rates.⁴²
1.5.2 Advanced/metastatic stages (IIIB/IV): NSCLC

In the advanced/metastatic stage of the disease, surgery is no longer an available option. The pharmacological treatment (see section 1.4) of advanced and metastatic NSCLC has undergone a rapid transformation within a short period of time (**Figure 1.7**).⁴³ The treatment strategy should take into account patients age, performance status, comorbidities, histology and molecular mutations.⁴⁴



Figure 1.7. Treatments for metastatic and resectable NSCLC from 2004 to 2020 approved by FDA

Source: Chaft JE, Rimner A, Weder W, Azzoli CG, Kris MG, Cascone T. Evolution of systemic therapy for stages I-III non-metastatic non-small-cell lung cancer. *Nat Rev Clin Oncol.* 2021 Sep;18(9):547-557 **DFS**: Disease free survival

Non squamous NSCLC

In all patients with advanced/metastatic stage non squamous NSCLC, it is recommended to complement the morphological diagnosis with molecular characterization of the tumor, in light of the possibility of prescribing targeted treatments.^{45,46} If a patients is found with a molecular mutation it is likely to receive a target therapy as a first-line treatment (**Figure 1.8**). Here we report the guidelines published by the European Society for Medical Oncology

committee for the metastatic stage of NSCLC (the most part of the drugs approved for the stage IV of NSCLC are also approved for the stage IIIB - see section 1.4).



Otherwise, if no molecular mutations are present, the patient with metastatic stage (as also advanced stage) non squamous NSCLC could receive immunotherapy or standard chemotherapy based on PD-L1 expression and performance status (**Figure 1.9**).

Pembrolizumab is considered a standard first-line option for patients with advanced NSCLC and PD-L1 expression \geq 50% who do not have contraindications to use of immunotherapy. Also performance status, patients age, tumor mutational burden are taken into account for the treatment algorithm of this patients. In patients with low performance status and with any expression of PD-L1 also the Tumor Mutational Burden (TMB) is considered in the treatment algorithm. The TNB open great perspectives in the personalized treatment of cancer. This molecular biomarker measures the number of mutations in the tumor and makes it possible to take a complete "picture" of the tumor's molecular alterations. It is a valuable tool because it can help to identify patients who will respond better to immunotherapy.⁴⁷



Figure 1.9. Treatment algorithm for stage IV Non squamous NSCLC, molecular tests negative

Source: <u>https://www.esmo.org/guidelines/lung-and-chest-tumours/clinical-practice-living-guidelines-metastatic-non-small-cell-lung-cancer</u>

Cht: Chemotherapy; **NSCC**: Non Squamous Cell Carcinoma; **BSC**: Best supportive care; **PS**: Performance Status; []: Number, letters in the parenthesis indicate the LoE (quality of evidence), and the GoR (quality of evidence and clinical significance/magnitude of benefit or harm), ESMO- Magnitude of Clinical Benefit Scale (MCBS) score; ^aIn absence of contraindications and conditioned by the registration and accessibility of anti-PD-(L)1 combinations with platinum-based ChT, this strategy will be preferred to platinum- based ChT in patients with PS 0-1 and PD-L1 < 50%. Alternatively, if TMB can accurately be evaluated, and conditioned by the registration and accessibility, nivolumab plus ipilimumab should be preferred to platinum-based standard ChT in patients with NSCLC with a high TMB; ^bNot EMA-approved

Squamous NSCLC

In patients with advanced/metastatic squamous NSCLC, the molecular test investigation is not recommended except in the rare circumstances the patient is found to be a never, long-time ex, or light smoker (**Figure 1.10**). Also in this cancer the clinical condition as also other information are considered in the treatment algorithm such as smoking status, PD-L1 expression, TMB, age and performance status.



Figure 1.10. Treatment algorithm for stage IV squamous NSCLC, molecular tests negative

Source: <u>https://www.esmo.org/guidelines/lung-and-chest-tumours/clinical-practice-living-guidelines-</u> metastatic-non-small-cell-lung-cancer

BSC: Best supportive care; **ChT**: Chemotherapy; **PS**: Performance Status; **SCC**: Squamous Cell Carcinoma; **TMB**: Tumor mutational burden; []: Number, letters in the parenthesis indicate the LoE (quality of evidence), and the GoR (quality of evidence and clinical significance/magnitude of benefit or harm), ESMO- Magnitude of Clinical Benefit Scale (MCBS) score; ^a Molecular testing is not recommended in SCC, except in those rare circumstances when SCC is found in a never-, long-time ex- or light-smoker (< 15 pack-years); ^b In absence of contraindications and conditioned by the registration and accessibility of anti-PD-(L)1 combinations with platinum-based ChT, this strategy will be preferred to platinum-based ChT in patients with PS 0-1 and PD-L1 < 50%. Alternatively, if TMB can accurately be evaluated, and conditioned by the registration and accessibility, nivolumab plus ipilimumab should be preferred to platinum-based standard ChT in patients with NSCLC with a high TMB; ^c Not EMA-approved

1.5.3 Extensive/metastatic stages (III/IV): SCLC

In the extensive (advanced) stage, target therapy is not approved for SCLC yet: while these drugs brought a real revolution in the treatment of NSCLC, this revolution for SCLC has yet to be observed. In fact, preclinical findings of the various aberrant processes in this type of cancer have not yet been translated into better outcomes with the addition of the targeted therapies to standard chemotherapy.⁴⁸ The low prevalence of SCLC represents the main

hurdle to the development of innovative pharmacotherapies. However, the orphan designation of different drugs intended for the treatment of SCLC has fostered the study of a number of promising treatments,⁴⁹ mainly immunotherapies, that were recently marketed, or will be possibly approved in the near future:⁵⁰ Nivolumab was approved for this indication by FDA in 2018. Durvalumab and atezolizumab are reported in ESMO guidelines for the treatment of patients with performance status 0-1 and with no contraindication to immunotherapy (**Figure 1.11**).



(i.e. stage IV or stage III SCLC not eligible for curative treatment)

Source: https://www.esmo.org/guidelines/lung-and-chest-tumours/small-cell-lung-cancer

BSC, best supportive care; **ChT**, chemotherapy; **G-CSF**, granulocyte colony-stimulating factor; **IO**, immunotherapy; **MCBS**, ESMO-Magnitude of Clinical Benefit Scale; **MRI**, magnetic resonance imaging; **PCI**, prophylactic cranial irradiation; **PS**, performance status; **RT**, radiotherapy; **SCLC**, small-cell lung cancer. ^aESMO-MCBS v1.1 score for new therapy/indication approved by the EMA or FDA. The score has been calculated by the ESMO-MCBS Working Group and validated by the ESMO Guidelines Committee. ^b Carboplatin may be replaced by cisplatin in patients

1.6 Value of conducting pharmacoepidemiology studies in lung cancers

Pharmacoepidemiology is the study of interactions between drugs and human populations, assessing the benefits, risks and the use of drugs in real word setting.⁵¹ This discipline focus on clinical patient outcomes by using methods of clinical epidemiology and applying them to understanding the determinants of beneficial and adverse drug effects, duration-response relationships, clinical effects of drug-drug interactions, and the effects of non-adherence to drug therapy.⁵¹

The pharmacological treatment (r)evolution of advanced / non-resectable lung cancer promoted by the approval of target- and immuno-therapies over the last ten year had a deep impact on the treatment of patients, on the healthcare system and on clinical research.

Figure 1.12 shows the number of new trials on immunotherapies (PD-1/PD-L1 monoclonal antibodies) started per year by different advanced cancer from 2006 to 2018.⁵²



specify cancer types to be tested

As shown in the figure, after 2013, the prevalence of trial evaluating immunotherapy has grown in the last 4 years. This trend was confirmed in the last four years (**Figure 1.13**).



Source: <u>https://www.cancerresearch.org/en-us/scientists/immuno-oncology-landscape/pd-1-pd-11-landscape</u> **PDx**: (PD-L1/PD-1); 4,400 clinical trials are in the current landscape as of September 2020, nearly tripling since in September 2017. FDA-approved monoclonal antibodies include pembrolizumab, nivolumab, durvalumab, atezolizumab, avelumab, and cemiplimab. Other PDx (PD-L1/PD-1) include monoclonal antibodies approved by regulatory agencies other than the FDA such as the EMA as well as those in clinical development and not yet approved

In this scenario where the number of trials (especially on immunotherapy) exploited in the last years, the marketing authorizations of anti-cancer therapies are usually based on the results of randomized controlled trials with a low number of patients, with a short duration of exposure and stringent selection criteria.^{53–56}

Effectiveness of anti-cancer drugs may be different than expected as daily practice differs from the experimental setting: as a matter of fact, poor evidence of survival benefits of anticancer drugs is present in literature.^{57,58} EMA approved 48 anticancer drugs for a total of 68 indications from 2009 to 2013 and only the 51% of them demonstrated to significant improve survival or quality of life.⁵⁷ Moreover, regulatory agencies (FDA and EMA) are increasingly required to take market approval decisions for new drugs on the basis of limited clinical evidence which is a situation commonly encountered in cancer.⁵⁹

Clinical trials also present limitations themself: 1) important safety issues may not be detectable in clinical trials because the frequency of many adverse events is too low and adverse effects of cancer therapies may occur many years after drug administration;⁵⁸ 2) reviews showed that women and racial/ethnic minorities are deeply underrepresented in clinical trials and in particular in clinical trials concerning cancer.^{55,60}

Taking into account the above considerations, monitoring the real-world pattern of utilization of antineoplastic drugs is fundamental for the generation of evidence from large and unselected populations, identify and address difference in the utilization between specific populations which are not or under- represented in clinical setting and finally complement information provided from randomized clinical trials using pharmacoepidemiology studies.⁶¹

1.7 Datasources to conduct pharmacoepidemiology studies in oncology

The conduction of pharmacoepidemiology studies in the oncology setting remains often a challenge since information to reliably describe utilization of cancer drugs, patients' characteristics and outcomes are often scattered in distinct and heterogenous data sources.^{62,63} The main three types of data sources available providing routinely collected electronical healthcare data (rcEHD) to conduct pharmacoepidemiology studies in oncology are: cancer registries, administrative/claims databases and electronical health/medical records.⁶²

- <u>Cancer registries:</u> Cancer registries are usually used for public health surveillance by monitoring trends in cancer incidence and mortality by age, gender, race and ethnicity, and type/stage. Researchers make use of registry data for describing patterns in cancer occurrence and for generating hypotheses for clinical and epidemiologic studies to inform cancer prevention and treatment improvements. These registries often include information on cancer diagnosis and mortality.
- <u>Administrative/claims database:</u> Administrative data are used increasingly to understand the delivery of health care services to a population of patients, and to understand care effectiveness or impact of regulatory actions. Administrative data can allow the assessment of large and generalizable populations of patients, thus limiting concerns about limited generalizability. The large sample sizes often available with administrative data may be particularly valuable for studying relatively smaller subgroups of

individuals, such as older individuals or racial/ethnic minorities. These data allow researchers, providers, and policy makers to track care delivered and can provide evidence about health care utilization and costs of care.

Typically, these data include information about services delivered, such as emergency visits, hospitalizations, diagnostic tests, procedures, and drug dispensations. Notably, such data are nearly always collected for purposes other than research, i.e. assessing policy changes for billing purposes and, thus, lack rich clinical information that may be important to accurately interpret findings.

• <u>Electronical healthcare records/ electronical medical records</u>: electronic healthcare records are defined as an electronic collection of medical information about the health histories of patients (i.e., diagnosis, drugs, laboratory tests). The electronic health record is a longitudinal collection of the electronic health information of individual patients or populations, while and electronical medical record is the patient record created by providers for specific encounters in hospitals or ambulatory environments. The latter could serve as a data source for the broader electronical healthcare records. Anyway, these terminologies have often been used interchangeably despite there is a difference between them.⁶⁴

Electronical healthcare records are being used increasingly for observational research, post-marketing safety evaluation, and to inform decision making. In February 2011, the FDA issued draft guidance regarding best practices for use such data in conducting pharmacoepidemiologic studies. The primary advantages of these data are their potentially comprehensive and relatively timely clinical information, with the possibility of including physicians' notes, patient symptoms and history, diagnostic information, and planned and actual treatments. Unfortunately, sometimes information in such data

26

sources are unable to communicate effectively between each other,⁶⁵ and sometimes information can only be found in free-text field.⁶⁶

1.8 Thesis objectives

In the thesis project we aimed to answer the following research questions:

- What is the available evidence in literature of studies using routinely collected electronical healthcare data about lung cancer?
- Is it possible to conduct a drug utilization study on NSCLC patients using the available Italian data?
- How is changed the pharmacotherapy of NSCLC in the last decade?
- Is the survival of these patients improved in a real-word setting?

Based on the research questions three works were conducted. First, a scoping review of the literature searching for drug utilization studies on lung cancer; the second study was conducted to validate a NSCLC case-finding algorithm in the regional pathology registry of the Tuscany region which is included into the regional administrative database; the last study concerned drug utilization and survival of patients with advanced NSCLC using real world data coming from the Tuscany region medico-administrative database.

The final goals of this thesis were to validate Italian medico-administrative data, to describe the drug utilization of NSCLC therapies over the last decade and to evaluate the impact of the introduction of these drugs on survival of NSCLC patients.

CHAPTER 2.

Scoping review of the literature: using routinely collected electronical healthcare data to study drug utilization in patients with lung cancer

2.1 Introduction

Observational studies based on large databases of rcEHD have the potential to complement information from clinical trials by allowing the observation of the "real world" clinical practice, thus leveraging data from wider and less strictly selected populations, during long-term follow-up periods.

The study of utilization of antineoplastic drugs is of paramount importance both for patients, to achieve the highest benefit from the pharmacotherapy received, and for the healthcare service, to maximize the results obtained from the resources invested. Unfortunately, information, such as histology, molecular characteristics, and stage of the tumor, that should be considered to plan the most appropriate pharmacotherapy for each patient, is frequently scattered in different databases. One emerging research direction is to leverage large, rcEHD to facilitate clinical and translational research.⁶² As for lung cancer, an overview of available data to conduct pharmacoepidemiologic studies on this topic was needed.

2.2 Study objectives

We decided to conduct a scoping review of the literature to provide a structured overview of the available studies using rcEHD to facilitate the design and benchmark of future works on this topic.

The objectives of this study were to retrieve available information from published studies using electronical healthcare data to facilitate the conduction of drug utilization study on lung cancer patients.

2.3 Publication

Real-World Utilization of Target- and Immunotherapies for Lung Cancer: A Scoping Review of Studies Based on Routinely Collected Electronic Healthcare Data



International Journal of Environmental Research and Public Health

Received: 27 April 2021 Accepted: 15 July 2021

Link to publication: https://www.mdpi.com/1660-4601/18/14/7679

Link to supplementary materials: https://www.mdpi.com/1660-4601/18/14/7679/s1







Real-World Utilization of Target- and Immunotherapies for Lung Cancer: A Scoping Review of Studies Based on Routinely Collected Electronic Healthcare Data

Andrea Spini ^{1,2,*}, Giulia Hyeraci ³, Claudia Bartolini ³, Sandra Donnini ^{4,*}, Pietro Rosellini ^{5,6}, Rosa Gini ³, Marina Ziche ², Francesco Salvo ^{1,6} and Giuseppe Roberto ³

- ¹ INSERM, BPH, U1219, Team Pharmacoepidemiology, University of Bordeaux, Bordeaux 33000, France; francesco.salvo@u-bordeaux.fr (F.S.)
- ² Department of Medical Science, Surgery and Neuroscience, University of Siena, Siena 53100, Italy; marina.ziche@unisi.it (M.Z.)
- ³ Osservatorio di Epidemiologia, Agenzia regionale di sanità della Toscana, Florence 50141, Italy; giulia.hyeraci@ars.toscana.it (G.H.); claudia.bartolini@ars.toscana.it (C.B.); rosa.gini@ars.toscana.it (R.G.); giuseppe.roberto@ars.toscana.it (G.R.)
- ⁴ Department of Life sciences, University of Siena, Siena 53100, Italy
- ⁵ CIC1401, CIC Bordeaux, Bordeaux 33000, France; rosellini.pietro@u-bordeaux.fr (P.R.)
- 6 Pole de Santé Publique, Service de Pharmacologie Médicale, Centre de Pharmacovigilance de Bordeaux, CHU de Bordueax, Bordeaux 33000, France
- * Correspondence: andrea.spini@u-bordeaux.fr (A.S.); donnini4@unisi.it (S.D.)

Abstract: Routinely collected electronic healthcare data (rcEHD) have a tremendous potential for enriching pre-marketing evidence on target- and immunotherapies used to treat lung cancer (LC). A scoping review was performed to provide a structured overview of available rcEHD-based studies on this topic and to support the execution of future research by facilitating access to pertinent literature both for study design and benchmarking. Eligible studies published between 2016 and 2020 in PubMed and ISI Web of Science were searched. Data source and study characteristics, as well as evidence on drug utilization and survival were extracted. Thirty-two studies were included. Twenty-six studies used North American data, while three used European data only. Thirteen studies linked ≥1 data source types among administrative/claims data, cancer registries and medical/health records. Twenty-nine studies retrieved cancer-related information from medical records/cancer registries and 31 studies retrieved information on drug utilization or survival from medical records or administrative/claim data. Most part of studies concerned non-small-cell-LC patients (29 out of 32) while none focused on small-cell-LC. Study cohorts ranged between 85 to 81,983 patients. Only two studies described first-line utilization of immunotherapies. Results from this review will serve as a starting point for the execution of future rcEHD-based studies on innovative LC pharmacotherapies.

Keywords: electronic healthcare data; big data; real-word data; real-word evidence; drug utilization; lung cancer; immunotherapy; target-therapy; scoping review

1. Introduction

Lung cancer is the most commonly diagnosed cancer worldwide (2.09 million cases in 2018). It accounts for 14.5% of the total cases of cancer in men and 8.4% in women, being the leading cause of cancer death in men (22.0%) [1,2].

Based on histological characteristics, the World Health Organization classifies lung cancers in small-cell lung cancers (SCLC) and non-small-cell lung cancers (NSCLC). [3] In 2018, SCLC accounted for about 300,000 cases while NSCLC for 1.8 million [1,2]. The

Citation: Spini, A.; Hyeraci, G.; Bartolini, C.; Donnini, S.; Rosellini, P.; Gini, R.; Ziche, M.; Salvo, F.; Roberto, G. Real-World Utilization of Target- and Immunotherapies for Lung Cancer: A Scoping Review of Studies Based on Routinely Collected Electronic Healthcare Data. Int. J. Environ. Res. Public Health 2021, 18, 7679. https://doi.org/10.3390/ijerph18147679

Academic Editor: Sara Malo and Antonio Gimeno-Miguel

Received: 27 April 2021 Accepted: 15 July 2021 Published: 19 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses /by/4.0/). latter can be distinguished in two main histotypes: squamous and non-squamous carcinoma [2,4–6].

Treatment of lung cancer relies on one or more therapeutic approaches among surgery, radiation therapy and pharmacotherapy [7]. Currently, a wide range of medications is available for the treatment of advanced NSCLC. The choice of a specific pharmacological regimen is mainly based on the stage of the cancer, although other factors such as the overall patient's health and lung function, as well as some specific molecular traits of the cancer itself, are also important. Early-stage NSCLC shows no overt clinical symptoms, and surgery represents the treatment of choice. In such cases, pharmacotherapy can be used both before, as neoadjuvant treatment aimed to reduce the size of the tumor, and after surgery, as adjuvant treatment intended to decrease the risk of cancer recurrence [5,8]. In advanced stages, where cancer has already spread, treatment choice depends on the specific site and number of metastases, other than age and overall health status of the patient. In particular, while pharmacotherapy of SCLC is based mainly on standard chemotherapy (the FDA approved nivolumab in August 2018), during the last 15 years, the pharmacological treatment for advanced stage NSCLC was revolutionized by the authorization of innovative anticancer therapies, such as target therapy and immunotherapy [9].

Target therapy includes drugs that can counteract specific mechanisms underlying the development of tumors [10]. These include the neutralizing antibody bevacizumab, which acts by binding the pro-angiogenic vascular endothelial growth factor (VEGF), the tyrosine kinase inhibitors (TKIs) directed to the epithelial growth factor receptor (EGFR), and TKIs of anaplastic lymphoma (ALK). Immunotherapy includes nivolumab (approved in 2015) and pembrolizumab (approved in 2016): these drugs inhibit the binding between lymphocyte protein death 1 (PD-1) and tumor ligand of PD-1 (PD-L1) by maintaining the immune system's response to the tumor [11,12]. In advanced NSCLC and in non-operable patients, some of these drugs are the first-line treatment (e.g., anti-TKIs) in patients with activating mutations in EGFR or ALK genes, while others are licensed as second-line treatment (e.g., Nivolumab) [8,13–15]. No target therapies are approved for the treatment of advanced SCLC.

Knowledge on efficacy and safety of authorized anticancer drugs mostly relies on evidence from clinical trials [16]. Such studies are usually based on relatively small samples of strictly selected, well monitored, patient populations, which are generally followed for short time periods [17].

In this context, observational studies based on large databases of routinely collected electronic healthcare data (rcEHD) has the potential to complement information from clinical trials by allowing the observation of the "real world" clinical practice, thus leveraging data from wider and less strictly selected populations, during long-term follow-up periods [18,19]. Given also the hot topic of using big data, as well as artificial intelligence, for longitudinal data mining in healthcare [20], an overview of available data to conduct pharmacoepidemiologic studies is needed. Unfortunately, the conduction of such studies in the oncology setting remains often a challenge since information to reliably describe utilization of cancer drugs, patients' characteristics and outcomes are often scattered in distinct data sources.

We performed a scoping review [21] of the published rcEHD-based studies concerning the utilization of target- and immuno-therapies in LC patients with the aim of providing a structured overview of the available studies to facilitate the design and benchmark of future works on this topic.

2. Material and Methods

2.1. Literature Search

We searched PubMed and ISI Web of Science databases for retrieving the articles of interest that were published from January 2016 to August 2020. Due to the approval of immunotherapy in 2015, January 2016 was chosen as starting date to give a more up-to-date picture of the issue [11]. The search string used was composed by three sets of keywords respectively related to the concepts "lung cancer", "drug-utilization measures", and "type of study/Data", respectively (see supplementary material—Table S1 for more details). Snowballing search was also conducted to retrieve additional papers of interest by examining the references cited in the included articles.

2.2. Eligibility Criteria

Retrospective observational studies based on information retrieved from rcEHD that reported evidence on target therapies and immunotherapies in patients with lung cancer were selected. Eligible studies had to be published between January 2016 and August 2020 and written in English. Studies based on ad hoc data collection or with no abstract or full-text available were excluded.

2.3. Study Selection

Two authors (AS and GH) screened all titles and abstracts of the references retrieved. Potentially relevant studies were further assessed through examination of full texts. The reviewers worked independently, in parallel, and blinded to each other. Disagreements between the two reviewers were solved through discussion with a third author (GR).

2.4. Data Extraction

The following information was extracted from the included studies:

- (i) Data source characteristics: type of source, name, catchment area. Notably, data source types were classified into three main categories: (a) administrative/claims data (i.e., data for health system planning and management and health assistance claims), (b) "medical/health records" (i.e., documentation of clinical care) and (c) "cancer registries" [22,23];
- (ii) Study characteristics: study population, population size, cohort type (population-based, hospital-based), study period, follow-up duration and drugs or drugs regimens. Additionally, relevant information items such, as cancer-related characteristics, patients-related characteristics, drug utilizations, vital status, were also classified by sources of rcEHD used, whenever possible;
- (iii) Information on the utilization of target- and immunotherapies based on treatment line and LC histology (e.g., pattern of use, frequency molecular testing, survival).

As for study selection, two authors extracted independently the data (AS and GH), and a third author (GR) was consulted in case of disagreement.

In particular, median overall survival (OS) values were extracted, whenever reported. Median OS values were grouped by treatment line and presented as the range between the maximum and minimum reported value.

3. Results

3.1. Literature Search Results

A total of 594 study references were retrieved from PubMed and ISI Web of Science (Figure 1).

Screening of titles and abstracts allowed the selection of 131 potentially eligible studies. Among them, a total of 32 studies fulfilled the eligibility criteria and were finally included into the review [24–55]. No further studies were retrieved through a snowballing search.



Figure 1. Flow chart.

3.2. Source of Routinely Collected Electronic Healthcare Data (rcEHD) Used by Study

Twenty-six out of 32 included studies used rcEHD from North America [27–31,33–41,43–54], two studies used data from Asia [24,32], three from Europe [26,42,55], and one from Australia [25]. Thirteen studies used record linkage of \geq 1 type of data source [24–26,28,30,32,35,36,42,44,45,50,55], while 19 studies were based on one data source type only. Among the latter, 14 studies used medical/health records only [29,33,34,37,41,43,46–49,51–54], four were based on administrative/claim data, [31,38–40] and one study used data from cancer registries (Table 1) [27]. Four studies [30,44,45,50] linked administrative/claim data with cancer registries.

				Datasource Type	
Study Reference Study	Data Source Name	Catchment Area	Administrative/	Medical/Health	Cancer
50			Claims Data	Records	Registry
	Institute for Clinical Evaluative Sciences				
	databases:Ontario Cancer Registry, OHIP billing				
Derver et al. 2016 [26]	claims data, Ontario Drug Benefit, Hospital	Canada			
Dawe et al., 2016 [56]	Discharge Abstracts, National Ambulatory Care	Canada	¥.	¥	¥.
	Reporting System, Home Care Database, Ontario				
	vital statistics				
Spence et al., 2017 [28]	Kaiser Permanente California (KPSC)	USA			
	-Care for Outcome registry (built on the				
Cramer-van der Welle et	Dutch cancer registry)	Notherlands			
al., 2018 [42]	 Clinical data from Netherland hospitals 	Nemeriands		¥.	×
	– Santeon Farmadatabase				
	-Care for Outcome registry (built on the				
Potors at al 2017 [26]	Dutch cancer registry)	Notherlands			
1 eters et al., 2017 [20]	 Clinical data from Netherland hospitals 	Weiterlands		Ł	Ł
	– Santeon Farmadatabase				
Brodor et al. 2018 [25]	– Truven Health Analytics MarketScan Database	LICA			
biodel et al., 2010 [55]	 IMS PharMetrics Database 	UJA	¥.	Ł	
Arupachalam at al 2018	–Surveillance, Epidemiology and End				
	Results-Medicare database (SEER)	USA			
[11]	– Medicare files				
	-Surveillance, Epidemiology and End				
Bittoni et al., 2018 [45]	Results-Medicare database (SEER)	USA			
	– Medicare files				
Bobbili et al., 2019 [50]	Surveillance, Epidemiology and	USA	$\mathbf{\nabla}$		

Table 1. Data source characteristics.

	End Results-Medicare database (SEER)				
	– Medicare files				
	Surveillance, Epidemiology and End				
Gilden et al., 2017 [30]	Results-Medicare database (SEER)	USA			
	– Medicare files				
Liang et al., 2016 [32]	Taiwan Cancer Ragistry, National Health	T-:			
	Insurance and National Death Registry	Taiwan			
Abernethy et al., 2017 [29]	Flatiron Health database	USA			
Aguilar et al., 2018 [41]	US Oncology Network's iKnowMed database	USA			
Chiang et al., 2020 [49]	Flatiron Health database	USA			
Davies et al., 2019 [46]	Flatiron Health database	USA			
	Electronic medical records from South Western	A		7	
Ding et al., 2017 [25]	Sydney Local Health District (SWSLHD)	Australia			
Jahanzeb et al., 2020 [48]	Flatiron Health database	USA			
Khozin et al., 2019 [52]	Flatiron Health database	USA			
	-International Oncology Network (ION)				
Lunacsek et al.,2016 [33]	electronic medical record (EMR) database	USA		\checkmark	
	– Social Security Death Master File				
McKay et al., 2016 [37]	Flatiron Health database	USA			
Molife et al., 2019 [54]	Flatiron Health database	USA			
N. J. 1. 1. 2010 [24]	US oncology iKnowMedTM (iKM) database				
Nadier et al., 2018 [34]	Social Security Death Index	USA			
Schwartzberg et al., 2019	The Group The label and the second				
[53]	Flatiron Health database	USA		\checkmark	
Simeone et al., 2019 [51]	Flatiron Health database	USA			
Waterhouse et al., 2020	US On colorer Naturarly's Wasser Mad database	TIC A			
[47]	US Oncology Network's INnowined database	UJA	V		
Winfree et al., 2018 [43]	Flatiron Health database	USA			

Kasymjanova et al., 2017	Jewish General Hospital's Pulmonary Division	Canada		
[27]	Lung Cancer Registry	Callada		¥.
	-Medical and pharmacy claims of insured			
Dalal et al. 2018 [28]	employees and their dependents		Z	
Dalal et al., 2016 [56]	-Medicare-eligible retirees with	USA		
	employer-provided Medicare Supplemental plans			
Hopson et al., 2018 [31]	Humana Research Database	USA		
Lowra et al. 2020 [55]	Programme de Médicalisation des Systèmes	France	R	
Levia et al., 2020 [55]	d'Information (PMSI)	Trance	V	
Shen et al., 2017 [39]	Truven Health MarketScan database	USA	N	
Shinde et al., 2016 [40]	Truven Health MarketScan database	USA	\square	
Wang et al., 2017 [24]	Medical Data Vision Database	Japan	N	

3.3. Characteristics of the Included Studies

The majority of the included studies concerned NSCLC patients (28 out of 32) [24–30,32–34,36–38,40–54], three studies included patients with unspecified lung tumor [31,39,55], and one study concerned neuroendocrine lung tumor [35]. Patients with SCLC were identified and included in one study, although the latter study primarily concerned NCSLC treatment (Table 2) [27].

The size of the study populations ranged from 85 to 81,983 patients. Sixteen studies included only patients \geq 18 years of age [24,28,29,31,33–35,38,41–43,46,47,51,53,54], five concerned elderly patients only (\geq 65 years old) [30,32,44,45,50], while 11 studies did not apply any age restriction to the study population [25–27,36,37,39,40,48,49,52,55]. Most part of the studies were population-based (28 out of 32) [28–55], while four were hospital-based [24–27]. All the included studies were longitudinal. Studies' observation period ranged between 2000 and 2018 [36,48]. Among studies that reported follow-up duration (13 out of 32), the mean follow-up time ranged from 6.9 to 20 months [44,53].

Study		Study Population,	Observation	Follow-Up		Drugs or Drug Regimens under Study				
Reference	•	Sample Size Cohort Type	Period	Duration	Tai	get Therapy	Immunotherapy	Other		
Dawe et al. [36]	• •	NSCLC 81,983 patients Population-based	January 2000– December 2010	-	•	Target therapy NOS	7	Standard chemotherapy, Complex single or multi agents, Special single agents or multi agents		
Spence et al. [28]	•	Patients ≥18 years old, NSCLC (III-IV) 2081 patients Population-based	January 2008– September 2014	-	•	Bevacizumab Erlotinib	-	Carboplatin, Cisplatin, Docetaxel, Etoposide, Gemcitabine, Paclitaxel, Pemetrexed, Vinorelbine		
Cramer-van der Welle et al., [42]	•	Patients ≥18 years old, NSCLC (IV) 1214 patients Population-based	January 2008– December 2014	-	•	Bevacizumab Erlotinib Gefinitib	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Paclitaxel		
Peters et al. [26]	•	NSCLC (III-IV) 2158 patients Hospital-based	January 2008– December 2012	-	•	TKI	~	Carboplatin, Cisplatin, Gemcitabine, Pemetrexed		
Broder et al. [35]	•	Patients ≥18 years old; lung neuroendocrine tumor 785 patients Population-based	July 2009–June 2014	14.3 months (Median: 11 months)	•	Target therapy NOS	/	Cytotoxic chemotherapy, Somatostatin analogues (+/-interferon)		
Arunachalam et al. [44]	•	Patients ≥65 years old, NSCLC (III-IV) 4033 patients Population-based	January 2007– December 2011	20 months; (Median 15. months)	7•	Bevacizumab	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Pemetrexed, Vinorelbine		
Bittoni et al. [45]	• •	Patients ≥65 years old, NSCLC (III-IV) 5931 patients Population-based	January 2007– December 2011	13.6 months (Median 8.9 months)	•	Bevacizumab	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Pemetrexed, Vinorelbine		
Gilden et al. [30]	•	Patients ≥65 years old, NSCLC (IIIB-IV) 77756 patients	January 2008– December 2010	177	•	Bevacizumab	-	Carboplatin, Cisplatin, Paclitaxel, Pemetrexed		

Table 2. Study characteristics.

•	Population-based					
• Liang et al. [32] •	Patients ≥65 years old, NSCLC (advanced) 25,008 patients Population-based	January 2005– December 2009	Median 14 months	Bevacizumab Erlotinib Gefitinib	-	Carboplatin, Cisplatin, Docetaxel, Etoposide, Epirubicin, Gemcitabine, Paclitaxel Pemetrexed, Tegafur, Vinorelbine
Abernethy et al. [29]	Patients ≥18 years old, NSCLC (IV) 4441 patients Population-based	November 2012- January 2015	- - •	Bevacizumab Erlotinib	PD-1 inhibitors	Carboplatin, Cisplatin, Docetaxel, Etoposide, Gemcitabine, Paclitaxel, Pemetrexed, Vinorelbine
• Aguilar et al. [41] •	Patients ≥18 years old, NSCLC (metastatic) 3108 patients Population-based	January 2011– June 2015	10.3 months (Median 7.6 months)	Bevacizumab Erlotinib	Nivolumab	Carboplatin, Docetaxel, Pemetrexed
• Davies et al. [46] •	Patients ≥18 years old NSCLC (IIIB-IV, ALK mutated) 300 patients Population-based	January 2011– December 2014	Median 16.6 • months •	Ceritinib Crizotinib	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Pemetrexed, Vinorelbine
• Ding et al. [25] •	NSCLC (advanced, EGFR- mu- tated) 85 patients Hospital-based	January 2010– June 2016	Median 10.7 • months •	Afatinib Erlotinib Gefinitib Rociletinib	-	-
• Lunacsek et al. [33] •	Patients ≥18 years old, NSCLC (advanced/ metastatic, non-squamous) 431 patients Population-based	April 2006–July 2013	- • •	Bevacizumab Cetuximab Erlotinib	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Metotrexate, Paclitaxel, Pemetrexed, Temozolomide, Vinflunine, Vinorelbine
McKay et al. [37]	NSCLC (advanced) 6867 patients Population-based	January 2011– April 2015	- •	Bevacizumab Erlotinib	-	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Paclitaxel, Pemetrexed Vinorelbine
Nadler et al. [34]	Patients ≥18 years old, NSCLC (IV) 10,689 patients	January 2012– April 2016	Median 6.9 months	Bevacizumab Crizotinib Erlotinib	Nivolumab	Carboplatin, Docetaxel, Gemcitabine, Paclitaxel, Pemetrexed

•	Population-based					
Winfree et al. [43]	Patients ≥18 years old, NSCLC (advanced, non-squamous) 715 patients Population-based	January 2011– October 2015	Median 13.8 months	•	Bevacizumab Erlotinib Nivolumab	Carboplatin, Cisplatin, Gemcitabine, Pemetrexed
Kasymjanova et al. [27]	NSCLC and SCLC 751 patients Hospital-based	January 2010– December 2014	-	•	Target therapy _ NOS	Chemotherapy: Combined agents, Double agents, Single agent
• Dalal et al. [38] •	Patients ≥18 years old NSCLC (with at least one prescription for ceritinib) 164 patients Population-based	January 2006– December 2015	-	•	Ceritinib Crizotinib - Other (NOS)	Standard chemotherapy
• Hopson et al. [31] •	Patients 18–89 years old, lung cancer (metastatic) + other tu- mors 3199 lung cancer patients Population-based	January 2007– December 2013	-	•	Target therapy _ NOS	Carboplatin, Cisplatin, Docetaxel, Oxaliplatin, Paclitaxel
Shen et al. [39]	lung cancer (advanced) stage 5842 patients Population-based	January 2013– June 2014	-	•	Bevacizumab Erlotinib	Pemetrexed
Shinde et al. [40]	NSCLC (metastatic) 4926 patients Population-based	January 2009– September 2012	-	:	Crizotinib Erlotinib	-
• Wang et al. [24] •	Patients ≥18 years old, NSCLC (IIIB-IV) 16,413 patients Hospital-based	April 2008– September 2015	10.3 months	•	Bevacizumab Erlotinib - Gefinitib	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Paclitaxel, Tegafur
• Waterhouse et al. [47] •	Patients ≥18 years old NSCLC (ALK-mutated, with at least one prescription of anti-ALK) 410 patients	September 2011- December 2017	-	•	Alectinib Brigatinib Ceritinib Crizotinib	-

•	Population-based					
Jahanzeb et al. [48] •	NSCLC (IIIB-IV, ALK-mutated, with at least one prescription of anti-ALK) 581 patients Population-based	January 2011– June 2018	- •	Alectinib Brigatinib Ceritinib Crizotinib	-	-
• Chiang et al. [49] •	NSCLC (metastatic, with at least one prescription of an- ti-EGFR) 782 patients Population-based	January 2011– September 2017	Median: 10.3 months	Afatinib Erlotinib Gefitinib Osimertinib	Immunotherapy not specified	Chemotherapy not specified
Bobbili et al. [50] •	Patients ≥65 years old, NSCLC (III) 4564 patients Population-based	January 2009– December 2014	- •	Bevacizumab Erlotinib	-	Carboplatin, Cisplatin, Docetaxel, Etoposide, Gemcitabine, Paclitaxel, Pemetrexed
• Simeone et al. [51] •	Patients ≥18 years old, NSCLC (IV) 9656 patients Population based	January 2013– January 2017	Median: 8.4 months	Bevacizumab Erlotinib Ramucirumab	Nivolumab	Carboplatin, Cisplatin, Docetaxel, Etoposide, Gemcitabine, Paclitaxel, Pemetrexed, Vinorelbine
• Khozin et al. [52] •	NSCLC (advanced or pro- gressed, with at least a pre- scription of immunotherapy) 5257 patients Population based	January 2011– December 2017	- •	Alk inhibitors Anti-EGFR Anti-VEGF	Atezolizumab Nivolumab Pembrolizumab	Platinum based chemotherapy, Non platinum based chemotherapy, Single agent Chemotherapy
Schwartzberg• et al. [53] •	Patients ≥18 NSCLC (advanced or progressed) 6597 patients (2 cohorts): Historical: 2357 patients Current: 4240 patients Population based	Historical: January 2011– December 2013 Current: January 2015– May 2017	Median: 5.8 • months •	Anti-EGFR Ramucirumab	Atezolizumab Nivolumab Pembrolizumab	Carboplatin, Docetaxel, Gemcitabine, Paclitaxel, Pemetrexed, Vinorelbine
Molife et al. • [54]	Patients ≥18 NSCLC (advanced or metastatic, with at least a	December 2014– May 2017		Afatinib Alectinib	Atezolizumab Ipilimumab	Carboplatin, Cisplatin, Docetaxel, Gemcitabine, Etoposide, Paclitaxel,

	•	prescription of immunotherapy or ramucirumab) 4054 Population-based	7		Bevacizumab Ceritinib Crizotinib Erlotinib Gefitinib Osimertinib Ramucirumah	Nivolumab Pembrolizumab	Pemetrexed, Vinorelbine
Levra et al. [55]	•	Patients with NSCLC (with at least a prescription of nivolumab) 10,452 patients Population based	January 2015– December 2016	-		Nivolumab Pembrolizumab	Chemotherapy not specified

3.4. Sources of rcEHD Used for Information Retrieval

Medical/health records and cancer registries were most frequently used to retrieve cancer-related information (i.e., histology, stage, molecular/genetic characterization, tumor response and disease progression—see Table S2 in supplementary file): on a total of 23 studies where the source used to retrieve the reported cancer-related information could be assessed, 13 used medical/health records [25,29,33,37,41,43,47-49,51-54] and seven used cancer registry data [26-28,32,36,42,50]. Notably, tumor response was reported in four studies only: the information was always retrieved from medical/health records [25,33,43,48]. In three studies based on administrative/claim data, instead, proxies of cancer-related information were used to identify tumor histology and/or stage [31,39,55]. A study based on French administrative healthcare data used bevacizumab or pemetrexed dispensing as a proxy for non-squamous NSCLC histology [55]. Two other studies based on administrative/claims data from US, identified patients with metastatic cancer by using algorithms based on a combination of ICD-9CM codes (e.g., excluding patients with a claim for lung surgery, and then selected only those patients with ICD-9CM codes referring to a metastatic disease—see Table S3 in the supplementary material further details on algorithms used to derive missing variables from administrative/claim data) [31,39].

Out of the 32 studies reporting information on drug utilization, 18 studies used medical/health records [25,26,28,29,33,36,37,41–43,47–49,51–54] and seven administrative/claim data [30–32,38–40,55]. Notably, four studies derived information on treatment line from administrative/claims data [31,32,35,55] and 16 from medical/health records [25,26,29,33,37,41,42,46–49,51–54].

As for information on vital status, it was retrieved from administrative data in seven studies [28,32–34,36,39,55] and from medical/health records in 11 studies, [25,37,41,43,47–49,51–54] in a total of 20 studies in which the source used to retrieve vital status could be assessed.

3.5. Utilization of Target- or Immuno- Therapies for Non-Small-Cell Lung Cancers (NSCLC)

Twenty-nine out of 32 studies described the use of innovative treatments in patients with NSCLC. [24–26,28–30,32–34,37,40–55].

3.5.1. First-Line Treatments for Advanced NSCLC Patients (III-IV Stage)

Twenty-one studies reporting information on the use of innovative pharmacotherapy as first-line treatment for advanced NSCLC were found [24–26,28–30,32–34,41– 43,45,46,48–54]: two studies concerned immunotherapy [52,54], 13 studies concerned the anti-angiogenic drug bevacizumab [24,28–30,33,34,41–43,45,50,51,54], and fourteen concerned TKIs.

The use of first-line immunotherapy was described by two studies [52,54]. Information about pembrolizumab and nivolumab (e.g., changes in treatment line during study period and trend for utilization) was reported in both studies. The study of Molife et al., reported also that no patients received atezolizumab as a first-line treatment in a population extracted from the US Flatiron healthcare database from 2014 to 2017 [54].

Eight studies described the use of bevacizumab in relation with histology (i.e., squamous/non-squamous) [24,28–30,33,42,43,45], while five did not specify the histology of NSCLC [34,41,50,51,54]. The prevalence of use of first-line bevacizumab among patients with advanced NSCLC was reported in 10 studies [24,28–30,34,41–43,45,50]. Among the latter studies, the use of bevacizumab in non-squamous NSCLC patients, varied between 6.0% and 50.9%, while it was negligible in patients with squamous NSCLC (from 0.0% to 1.5%; see Figure 2a). The study of Molife et al., included also a cohort of patients treated with ramucirumab [54].



Panel (a) First-line utilization of bevacizumab by tumor histology

Panel (b) First-line utilization of anti-EGFR drugs according to EGFR mutation status



Figure 2. Reported incidence of use of innovative drugs as first-line pharmacotherapy for advanced non-small-cell lung cancers (NSCLC). (a) First-line utilization of bevacizumab by tumor histology; (b) First-line utilization of anti-EGFR drugs according to EGFR mutation status; NOS: not otherwise specified.

Among the 11 studies concerning first-line anti-EGFR TKIs, eight reported the use of erlotinib [24,25,34,41,42,49,50]. Among these, three studies also provided information on the use of gefitinib [24,42,49], and two on the use of afatinib [24,49]. Five studies reported that the incidence of use of anti-EGFR among advanced NSCLC patients varied between 3.7% and 32.9% [24,34,41,42,50]. Two studies described the use of anti-EGFR TKIs in a population of NSCLC patients with an activating mutation of the related gene, and found an incidence of use between 77.8% and 85.0% (Figure 2b) [25,41]. The anti-EGFR TKIs median duration of first-line treatment in patients with activating mutation ranged between 6.5 months and 9 months [25,41]. Four studies concerned the use of the anti-ALK TKI crizotinib [34,46,48,54], three ceritinib [46,48,54], two alectinib [48,54] and one brigatinib [48]. Notably, two out of four studies [46,48] investigated the use of first-line anti-ALK TKIs in a population with ALK-mutated NSCLC. 3.5.2. Second-Line Treatments for Advanced NSCLC (III-IV Stage)

Thirteen studies described the use of innovative anticancer drugs as second-line pharmacotherapy for advanced NSCLC (Figure 3) [29,33,34,37,41,44,46,48,50–54].

Seven studies described the use of immunotherapy as second-line treatment in patients with NSCLC [29,34,41,51–54]. PD-L1 cancer expression ranged between 1.3% and 57.7% and was reported in five studies [34,51–54]. Studies that reported information about the use of nivolumab [34,41,51–54], pembrolizumab [52–54] and atezolizumab [52– 54] were six, three and three, respectively. One study described immunotherapies utilization without distinction on the active substance concerned [29].



Figure 3. Reported incidence of use of innovative drugs as second-line pharmacotherapy for NSCLC. * Utilization was assessed in patient with anaplastic lymphoma (ALK)-mutated NSCLC.

Five studies described the use of bevacizumab as second-line treatment for advanced NSCLC in the US [37,44,50,51,54]. The reported incidence of use varied between 6.2% and 15%. Two studies shows that second-line bevacizumab was used to treat non-squamous NSCLC patients only [37,44].

Eight studies from the US concerned anti-EGFR therapies as second-line in patients with advanced NSCLC [34,37,49–54]. All the eight studies described the use of erlotinib, of which three described also the use of gefinitib, afatinib and osimertinib [49,53,54]. Five studies showed that the incidence of use of second-line anti-EGFR utilization among advanced NSCLC patients ranged between 3.6% and 18.6% [34,37,50,51,53]. Four US studies reported the use of anti-ALK therapies as second-line therapy [46,48,52,54]. Two out of four studies described the use of anti-ALK medications in an ALK mutated NSCLC cohort [46,48], while the remaining two studies concerned a cohort of patients included regardless of molecular characteristics of the tumor.

3.6. Utilization of Target- or Immuno-Therapies for Neuroendocrine Lung Cancer

One out of the 32 included studies referred to patients with neuroendocrine lung cancer [35]. Using the MarketScan Database and PharMetrics Database between July 2009 and June 2014, the authors reported that in a total of 785 patients, 78.2% started first-line therapy with cytotoxic chemotherapy, 18.1% with somatostatin analogues, and 1.1% with other drugs such as sunitinib or everolimus.

3.7. Utilization of Target- or Immuno-Therapies for Unspecified Lung Cancer Histology

Using administrative data only, two studies (two from the US) included advanced stage lung cancer patients regardless whether they were diagnosed with NSCLC or SCLC [31,39,55]. Both studies used data sources from the USA [31,39]: the first described the first-line use of biologic therapy (bevacizumab, crizotinib, erlotinib and cetuximab) in patients with metastatic lung cancer by site-of-care [31] and the second one the use of erlotinib in patients with EGFR mutated metastatic lung cancer [39].

3.8. Survival of Patients Treated with Target- or Immuno- Therapies for Advanced NSCLC

Twenty-two out of 32 studies reported the median overall survival (OS) of patients with advanced NSCLC [25–30,32–34,36,42–48,50–54]. The shortest median OS was reported for patients with stage IV NSCLC in the period 2011–2013 without known EGFR or ALK mutations (7.3 months), while the longest median OS was reported in patients with ALK activating mutations in a US population during the period 2011–2017 (27.6 months) [47,53]. In patients without EGFR or ALK mutations, the reported median OS ranged from 7.8 to 10 months for non-squamous NSCLC and from 6.5 to 8.5 months for patients with squamous NSCLC [29,53]. Three studies reported evidence on median OS in elderly patients (\geq 65 years old) [44,45,50] which ranged between 6.4 and 6.7 months for squamous advanced NSCLC, and between 7.5 and 7.8 for non-squamous advanced NSCLC.

Eight studies reported evidence on OS by drug treatment (see Table S4 in the supplementary material for the range of reported median OS found by treatment among advanced/metastatic NSCLC patients) [28,33,34,42,43,48,51,54]. Among the latter, five were referred to an advanced stage (III–IV) NSCLC [28,33,43,48,54], and three concerned metastatic stage (IV) NSCLC only [34,42,51]. Seven studies reported median OS for first-line drug treatments [28,33,34,42,43,48,54], and three studies also reported median OS for second- or third-line pharmacotherapies [34,48,51,52].

Median progression-free survival (PFS) was reported by five studies [25,33,43,48,52]. Among them, three studies reported PFS in relation to first-line treatment [33,43,48] and two studies in relation to second-line treatments [48,52] (see Table S4 in the supplementary material).

4. Discussion

With this scoping review we provided a structured overview of the available literature concerning recently published rcEHD-based studies concerning the utilization of target- or immunotherapies for LC. Our results highlighted a paucity of studies performed in Europe concerning immunotherapies, particularly as first-line pharmacotherapy, and the absence of papers reporting on the utilization of innovative drugs in SCLC patients. Focusing on the different types of rcEHD and methodologies used to retrieve information, results from this review represent a starting point for future studies on this topic, also highlighting current gaps of knowledge and facilitating access to pertinent literature both for study design and for benchmarking of results.

As for countries of data provenance, most of the studies included in this scoping review were conducted using data from the USA or Canada. This is probably because in regions other than North America, healthcare data are often scattered in different and heterogeneous databases, so that the performing studies on rare events that requires information from different healthcare settings, as in the case of lung cancer, remains a challenge [56]. Moreover, the approval of new anticancer medication in Europe is often delayed compared to the US [57], possibly contributing to the higher number of studies from the US included in this review compared to those using European Union (EU) data.

Indeed, results from this review demonstrated that a unique source of electronic healthcare data among administrative/claims data, medical/health records, cancer registry is often insufficient for performing an observational study on the real-world utilization of drugs for LC, as well as for other types of tumor [58].

Concerning the specific sources of rcEHD used for information retrieval, medical/health records were the most frequently used source of information for assessing drug exposure. Among the studies included in this review, administrative/claims data were less frequently used for retrieving such information. This was probably because in-hospital drug utilization might not always be tracked at patient-level in this type of rcEHD [30–34,36,39,55]. Also information on treatment-line is not usually available in administrative/claims data, although ad hoc algorithms can be adopted to derive this information (Table S3 supplementary material) [31,32,35,55]. Moreover, administrative/claims data usually do not record clinical information, such as tumor stage, histology, or gene mutations, which are crucial for studying drug utilization patterns and health outcomes in cancer patients. In this respect, the use of medical/health records or cancer registries appeared to be in most cases necessary [25,28,29,32]. Our results showed that information on disease progression and tumor response was only retrieved from medical/health records [25], while vital status was assessed using administrative/claims data or medical health records, although the former are usually considered as the gold standard for such information [59]. Indeed, each type of data source has its strengths and limitations with respect to the specific research question that needs to be addressed. Even within each of the three general categories of data sources adopted in this review [19], a significant heterogeneity in terms of content and validity has to be expected (see Table S2 in the supplementary file). Therefore, as has already happened in other contexts [56,60,61], fostering the development of methodologies for leveraging data diversity will be of paramount importance for the generation of solid evidence on the real-world utilization of drugs in LC.

As for evidence on the real-world utilization of innovative anticancer drugs, most of the included studies concerned patients with advanced stage NSCLC while no studies focusing on SCLC were found. The absence of licensed target therapies and the recent approval of immunotherapies for SCLC (Nivolumab was the first approved in August 2018 in the US [62]) apparently explains the absence of any published study focusing on SCLC in our literature review. Given the very low prevalence of SCLC [8], rcEHD has the potential to play an important role in capturing and studying far larger populations of SCLC patients than those recruited in clinical trials. The orphan designation of different drugs intended for the treatment of SCLC has promoted the study of a number of promising treatments [63], mainly immunotherapies, that were recently marketed, or will be possibly approved in the near future [62,64]. However, further initiatives are desirable to foster SCLC genotyping for the discovery of new molecular targets useful to develop innovative medications. Findings from this review showed that available evidence on immunotherapies from rcEHD-based studies concerning immunotherapy used for advanced stage NSCLC is still scarce, particularly with respect to their use as first-line pharmacotherapy. Notably, only two studies reported evidence on the real-world utilization of immunotherapies administered as first-line pharmacotherapy in patients with advanced NSCLC [52,54]. Such paucity of literature is mostly due to the recent approval of this class of anticancer medications for such indications. In fact, pembrolizumab was the first immunotherapy approved for first-line treatment of advanced NSCLC in 2017 [11,12]. The reported estimates of the incidence of use of immunotherapies as second-line pharmacotherapy for NSCLC, instead, appeared extremely variable from one study to the other mostly due to the different study period, cohort characteristics, and active principles concerned (from 9.8% to 48.8%) [29,34,41,53]. Such heterogeneity of study characteristics and results, however, represents an important resource for benchmarking results of future studies. Conversely, from immunotherapy a markedly higher number of studies on target therapies as first or second-line treatment for NSCLC were found. These studies provided information on the real-world utilization of such a class of medications, such as estimates of the frequency of the use in the relevant study populations, by histology as well as by molecular test execution.

The main strength of this review is the systematic approach adopted for reviewing the available body of recently published literature on the topic, with an in-depth screening of the records retrieved from two comprehensive databases like PubMed and ISI web of science. In particular, the choice of including studies published starting from 2016 was mainly due to the recent approval of some of the drugs and indications of interest (e.g., the first included studies concerning immunotherapies was published in 2017). Moreover, this approach, other than increasing the efficiency of the literature search efforts (i.e., the number of observational studies published increased in the last few years [65]), allowed us to provide an overview of studies concerning the most up-to-date evidence and methodologies on the topic. Indeed, given the scoping nature of this review, quality assessment of the included studies was not performed.

5. Conclusions

In conclusion, this scoping review provided a structured overview of the published rcEHD-based studies that investigated the real-world utilization of target and immunotherapies in lung cancer patients. The characteristics of studies included in this review showed that record-linkage of different sources of rcEHD often appears to be necessary. Cancer-related information were mainly retrieved from medical/health records or cancer registries while information on drug utilization or vital status were extracted in most of cases from medical/health records or administrative/claim data. As for evidence collected on the utilization of innovative medications for lung cancer, our results highlighted a paucity of studies performed in Europe as well as concerning immunotherapies, particularly as first-line pharmacotherapy. Notably, no study reporting drug utilization evidence concerning SCLC patients was found due to the absence of licensed target therapies and the very recent approval of immunotherapies for this indication.

Finally, this work will serve as a starting point for the execution of future real-world studies based on rcEHD facilitating access to pertinent literature both for study design and for benchmarking of results.

Supplementary Materials: The following are available online at www.mdpi.com/1660-4601/18/14/7679/s1, Table S1; Research Strategy, Table S2; Sources of the data used by the included studies per study variable, Table S3; Algorithms used to extract information from administrative/claims data as reported in the papers included in the review, Table S4; Minimum and maximum median OS reported per treatment-line among stage III-IV NSCLC patients.

Author Contributions: M.Z., R.G., G.R. and S.D. conceived the study. A.S., F.S. and G.R. analyzed data and wrote the manuscript; A.S., G.H., C.B. and G.R. selected eligible articles and extracted information from retrieved studies. P.R. contributed to the interpretation and discussion of study results. All authors revised and approved the final version of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are all available within this manuscript and its supplementary material.

Acknowledgments: The work was partially sponsored by Progetto CORELAB-Bando Salute 2018-Regione Toscana.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

- World Health Organization. Cancer. 2020. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 5 October 2019).
- World Health Organization PRGlobocanFinal Pdf. Available online: https://www.who.int/cancer/PRGlobocanFinal.pdf (accessed on 22 May 2019).
- Travis, W.D.; Brambilla, E.; Burke, A.P.; Marx, A.; Nicholson, A.G. Introduction to the 2015 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart. J. Thorac. Oncol. 2015, 10, 1240–1242, doi:10.1097/jto.00000000000663.
- 4. Aiom Linee Guida Neoplasie Del Polmone–2019. Available online: https://www.aiom.it/linee-guida-aiom-neoplasie-del-polmone-2019/ (accessed on 7 July 2020).
- Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.R.; Bharat, A.; Bruno, D.S.; Chang, J.Y.; Chirieac, L.R.; D'Amico, T.A. et al. NCCN Clinical Practice Guidelines in Oncology: Non-Small Cell Lung Cancer, Version 2.2021. Available online: https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf (accessed on 12 January 2021).
- Ho, C.; Tong, K.; Ramsden, K.; Ionescu, D.; Laskin, J. Histologic Classification of Non-Small-Cell Lung Cancer over Time: Reducing the Rates of Not-Otherwise-Specified. *Curr. Oncol.* 2015, 22, 164–170, doi:10.3747/co.22.2339.
- Zappa, C.; Mousa, S.A. Non-Small Cell Lung Cancer: Current Treatment and Future Advances. Transl. Lung Cancer Res. 2016, 5, 288–300, doi:10.21037/tlcr.2016.06.07.
- 8. AIOM Linee Guida Associazione Italiana Di Oncologia Medica 2018: Tumore al Polmone. Available online: https://www.aiom.it/wp-content/uploads/2018/11/2018_LG_AIOM_Polmone.pdf (accessed on 12 January 2021).
- Toschi, L.; Rossi, S.; Finocchiaro, G.; Santoro, A. Non-Small Cell Lung Cancer Treatment (r)Evolution: Ten Years of Advances and more to Come. *Ecancermedicalscience* 2017, 11, 787, doi:10.3332/ecancer.2017.787.
- Mayekar, M.K.; Bivona, T.G. Current Landscape of Targeted Therapy in Lung Cancer. Clin. Pharmacol. Ther. 2017, 102, 757– 764, doi:10.1002/cpt.810.
- 11. EMA Opdivo. Available online: https://www.ema.europa.eu/en/medicines/human/EPAR/opdivo (accessed on 4 October 2019).
- 12. EMA Keytruda. Available online: https://www.ema.europa.eu/en/medicines/human/EPAR/keytruda (accessed on 4 October 2019).
- 13. Absenger, G.; Terzic, J.; Bezan, A. ASCO Update: Lung Cancer. *Memo Mag. Eur. Med. Oncol.* 2017, 10, 224–227, doi:10.1007/s12254-017-0373-x.
- 14. Horn, L. Advances in the Treatment of Non-Small Cell Lung Cancer. J. Natl. Compr. Cancer Netw. 2014, 12, 764–767.
- Melosky, B. Current Treatment Algorithms for Patients with Metastatic Non-Small Cell, Non-Squamous Lung Cancer. Front. Oncol. 2017, 7, 38, doi:10.3389/fonc.2017.00038.
- Unger, J.M.; Cook, E.; Tai, E.; Bleyer, A. The Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies. Am. Soc. Clin. Oncol. Educ. Book 2016, 36, 185–198, doi:10.14694/edbk_156686.
- 17. Fahey, T. Applying the Results of Clinical Trials to Patients to General Practice: Perceived Problems, Strengths, Assumptions, and Challenges for the Future. Br. J. Gen. Pr. 1998, 48, 1173–1178.
- Snyder, C.F.; Wu, A.W.; Miller, R.S.; Jensen, R.E.; Bantug, E.T.; Wolff, A. The Role of Informatics in Promoting Patient-Centered Care. Cancer J. 2011, 17, 211–218, doi:10.1097/ppo.0b013e318225ff89.
- Houser, S.H.; Colquitt, S.; Clements, K.; Hart-Hester, S. The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama. *Perspect. Health Inf. Manag.* 2012, 9, 9. Available online: http://europepmc.org/article/MED/22737098 (accessed on 19 July 2021)
- Benke, K.; Benke, G. Artificial Intelligence and Big Data in Public Health. Int. J. Environ. Res. Public Health 2018, 15, 2796, doi:10.3390/ijerph15122796.
- Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic Review or Scoping Review? Guidance for Authors When Choosing between A Systematic or Scoping Review Approach. BMC Med. Res. Methodol. 2018, 18, 1–7, doi:10.1186/s12874-018-0611-x.
- Benchimol, E.I.; Smeeth, L.; Guttmann, A.; Harron, K.; Moher, D.; Petersen, I.; Sørensen, H.T.; von Elm, E.; Langan, S.M. RECORD Working Committee The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement. *PLoS Med.* 2015, *12*, e1001885, doi:10.1371/journal.pmed.1001885.
- Hiatt, R.A.; Tai, C.G.; Blayney, D.W.; Deapen, D.; Hogarth, M.; Kizer, K.W.; Lipscomb, J.; Malin, J.; Phillips, S.K.; Santa, J.; et al. Leveraging State Cancer Registries to Measure and Improve the Quality of Cancer Care: A Potential Strategy for California and Beyond. J. Natl. Cancer Inst. 2015, 107, djv047, doi:10.1093/jnci/djv047.
- Wang, F.; Mishina, S.; Takai, S.; Le, T.K.; Ochi, K.; Funato, K.; Matsuoka, S.; Ohe, Y. Systemic Treatment Patterns with Advanced or Recurrent Non–Small Cell Lung Cancer in Japan: A Retrospective Hospital Administrative Database Study. *Clin. Ther.* 2017, 39, 1146–1160, doi:10.1016/j.clinthera.2017.04.010.

- Ding, P.N.; Roberts, T.L.; Chua, W.; Becker, T.M.; Descallar, J.; Yip, P.Y.; Bray, V. Clinical Outcomes in Patients with Advanced Epidermal Growth Factor Receptor-Mutated Non-Small-Cell Lung Cancer in South Western Sydney Local Health District. *Intern. Med. J.* 2017, 47, 1405–1411, doi:10.1111/imj.13555.
- Peters, B.J.; Welle, C.M.C.-V.; Smit, A.A.; Schramel, F.M.; Van De Garde, E.M. Trends in Prescribing Systemic Treatment and Overall Survival for Non-Small Cell Lung Cancer Stage IIIB/IV in the Netherlands: 2008–2012. *Cancer Epidemiol.* 2017, 51, 1–6, doi:10.1016/j.canep.2017.08.001.
- Kasymjanova, G.; Small, D.; Cohen, V.; Jagoe, R.T.; Batist, G.; Sateren, W.; Ernst, P.; Pepe, C.; Sakr, L.; Agulnik, J. Lung Cancer Care Trajectory at A Canadian Centre: An Evaluation of How Wait Times Affect Clinical Outcomes. *Curr. Oncol.* 2017, 24, 302–309, doi:10.3747/co.24.3611.
- Spence, M.M.; Hui, R.L.; Chang, J.T.; Schottinger, J.E.; Millares, M.; Rashid, N. Treatment Patterns and Overall Survival Associated with First-Line Systemic Therapy for Patients with Advanced Non-Small Cell Lung Cancer. J. Manag. Care Spéc. Pharm. 2017, 23, 195–205, doi:10.18553/jmcp.2017.23.2.195.
- Abernethy, A.P.; Arunachalam, A.; Burke, T.; McKay, C.; Cao, X.; Sorg, R.; Carbone, D.P. Real-World First-Line Treatment and Overall Survival in Non-Small Cell Lung Cancer without Known EGFR Mutations or ALK Rearrangements in US Community Oncology Setting. *PLoS ONE* 2017, *12*, e0178420, doi:10.1371/journal.pone.0178420.
- Gilden, D.M.; Kubisiak, J.M.; Pohl, G.M.; Ball, D.E.; Gilden, D.E.; John, W.J.; Wetmore, S.; Winfree, K.B. Treatment Patterns and Cost-Effectiveness of First Line Treatment of Advanced Non-Squamous Non-Small Cell Lung Cancer in Medicare Patients. J. Med. Econ. 2016, 20, 151–161, doi:10.1080/13696998.2016.1230550.
- Hopson, S.; Casebeer, A.; Stemkowski, S.; Antol, D.D.; Tao, Z.; Howe, A.; Patton, J.; Small, A.; Masaquel, A. Does Site-Of-Care for Oncology Infusion Therapy Influence Treatment Patterns, Cost, and Quality in the United States? *J. Med. Econ.* 2017, *21*, 152–162, doi:10.1080/13696998.2017.1384736.
- Liang, Y.-H.; Shao, Y.-Y.; Liao, B.-C.; Lee, H.-S.; Yang, J.C.-H.; Chen, H.-M.; Chiang, C.-J.; Cheng, A.-L.; Lai, M.-S. Cytotoxic Chemotherapy as First-Line Therapy for Advanced Non-Small-Cell Lung Cancer in Taiwan: Daily Practice. J. Cancer 2016, 7, 1515–1523, doi:10.7150/jca.15180.
- Lunacsek, O.E.; Ravelo, A.; Coutinho, A.D.; Hazard, S.J.; Green, M.R.; Willey, J.; Eaddy, M.; Goertz, H.-P. First-Line Treatment with Bevacizumab and Platinum Doublet Combination in Non-Squamous Non-Small Cell Lung Cancer: A Retrospective Cohort Study in US Oncology Community Practices. *Drugs-Real World Outcomes* 2016, 3, 333–343, doi:10.1007/s40801-016-0090-5.
- Nadler, E.; Espirito, J.L.; Pavilack, M.; Boyd, M.; Vergara-Silva, A.; Fernandes, A. Treatment Patterns and Clinical Outcomes Among Metastatic Non-Small-Cell Lung Cancer Patients Treated in the Community Practice Setting. *Clin. Lung Cancer* 2018, 19, 360–370, doi:10.1016/j.cllc.2018.02.002.
- Broder, M.S.; Cai, B.; Chang, E.; Neary, M.P.; Papoyan, E.; Iii, A.B.B. Real-World Treatment Patterns for Lung Neuroendocrine Tumors: A Claims Database Analysis. Oncology 2018, 94, 281–288, doi:10.1159/000486282.
- Dawe, D.E.; Pond, G.R.; Ellis, P.M. Assessment of Referral and Chemotherapy Treatment Patterns for Elderly Patients with Non-Small-Cell Lung Cancer. Clin. Lung Cancer 2016, 17, 563–572, doi:10.1016/j.cllc.2016.05.012.
- McKay, C.; Burke, T.; Cao, X.; Abernethy, A.P.; Carbone, D.P. Treatment Patterns for Advanced Non-Small-Cell Lung Cancer after Platinum-Containing Therapy in U.S. Community Oncology Clinical Practice. *Clin. Lung Cancer* 2016, 17, 449–460.e7, doi:10.1016/j.cllc.2016.03.008.
- Dalal, A.A.; Guérin, A.; Mutebi, A.; Culver, K.W. Treatment Patterns, Clinical and Economic Outcomes of Patients with Anaplastic Lymphoma Kinase-Positive Non-Small Cell Lung Cancer Receiving Ceritinib: A Retrospective Observational Claims Analysis. J. Drug Assess. 2018, 7, 21–27, doi:10.1080/21556660.2018.1445092.
- Shen, C.; Kehl, K.L.; Zhao, B.; Simon, G.R.; Zhou, S.; Giordano, S.H. Utilization Patterns and Trends in Epidermal Growth Factor Receptor (EGFR) Mutation Testing Among Patients with Newly Diagnosed Metastatic Lung Cancer. *Clin. Lung Cancer* 2017, 18, e233–e241, doi:10.1016/j.cllc.2016.11.002.
- Shinde, R.; Cao, X.; Kothari, S. Biopsy Procedures and Molecular Testing Utilization and Related Costs in Patients with Metastatic Lung Cancer. J. Manag. Care Spéc. Pharm. 2016, 22, 1194–1203, doi:10.18553/jmcp.2016.15404.
- Aguilar, K.M.; Winfree, K.B.; Muehlenbein, C.E.; Zhu, Y.E.; Wilson, T.; Wetmore, S.; Nadler, E.S. Treatment Patterns by EGFR Mutation Status in Non-Small Cell Lung Cancer Patients in the USA: A Retrospective Database Analysis. *Adv. Ther.* 2018, 35, 1905–1919, doi:10.1007/s12325-018-0811-0.
- Cramer-van der Welle, C.M.; Peters, B.; Schramel, F.M.; Klungel, O.H.; Groen, H.J.; Van De Garde, E.M. Systematic Evaluation of the Efficacy—Effectiveness Gap of Systemic Treatments in Metastatic Nonsmall Cell Lung Cancer. *Eur. Respir. J.* 2018, *52*, 1801100, doi:10.1183/13993003.01100-2018.
- Winfree, K.B.; Torres, A.Z.; Zhu, Y.E.; Muehlenbein, C.; Aggarwal, H.; Woods, S.; Abernethy, A. Treatment Patterns, Duration and Outcomes of Pemetrexed Maintenance Therapy in Patients with Advanced NSCLC in A Real-World Setting. *Curr. Med. Res. Opin.* 2018, 35, 817–827, doi:10.1080/03007995.2018.1547273.
- Arunachalam, A.; Li, H.; Bittoni, M.A.; Camacho, R.; Cao, X.; Zhong, Y.; Lubiniecki, G.M.; Carbone, D.P. Real-World Treatment Patterns, Overall Survival, and Occurrence and Costs of Adverse Events Associated with Second-Line Therapies for Medicare Patients with Advanced Non-Small-Cell Lung Cancer. *Clin. Lung Cancer* 2018, 19, e783–e799, doi:10.1016/j.cllc.2018.05.016.
- 45. Bittoni, M.A.; Arunachalam, A.; Li, H.; Camacho, R.; He, J.; Zhong, Y.; Lubiniecki, G.M.; Carbone, D.P. Real-World Treatment Patterns, Overall Survival, and Occurrence and Costs of Adverse Events Associated with First-Line Therapies for Medicare

Patients 65 Years and Older with Advanced Non-Small-Cell Lung Cancer: A Retrospective Study. *Clin. Lung Cancer* 2018, 19, e629–e645, doi:10.1016/j.cllc.2018.04.017.

- Davies, J.; Martinec, M.; Coudert, M.; Delmar, P.; Crane, G. Real-World Anaplastic Lymphoma Kinase (ALK) Rearrangement Testing Patterns, Treatment Sequences, and Survival of ALK Inhibitor-Treated Patients. *Curr. Med. Res. Opin.* 2018, 35, 535– 542, doi:10.1080/03007995.2018.1533458.
- Waterhouse, D.M.; Espirito, J.L.; Chioda, M.D.; Baidoo, B.; Mardekian, J.; Robert, N.J.; Masters, E.T. Retrospective Observational Study of ALK-Inhibitor Therapy Sequencing and Outcomes in Patients with ALK-Positive Non-Small Cell Lung Cancer. *Drugs-Real World Outcomes* 2020, *7*, 261–269, doi:10.1007/s40801-020-00207-6.
- Jahanzeb, M.; Lin, H.M.; Pan, X.; Yin, Y.; Wu, Y.; Nordstrom, B.; Socinski, M.A. Real-World Treatment Patterns and Progression-Free Survival Associated with Anaplastic Lymphoma Kinase (ALK) Tyrosine Kinase Inhibitor Therapies for ALK + Non-Small Cell Lung Cancer. Oncologist 2020, 25, 867–877, doi:10.1634/theoncologist.2020-0011.
- Chiang, A.C.; Fernandes, A.W.; Pavilack, M.; Wu, J.W.; Laliberté, F.; Duh, M.S.; Chehab, N.; Subramanian, J. EGFR Mutation Testing and Treatment Decisions in Patients Progressing on First- or Second-Generation Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors. *BMC Cancer* 2020, 20, 1–13, doi:10.1186/s12885-020-06826-0.
- Bobbili, P.; Ryan, K.; Duh, M.S.; Dua, A.; Fernandes, A.W.; Pavilack, M.; Gomez, J. Treatment Patterns and Overall Survival among Patients with Unresectable, Stage III Non-Small-Cell Lung Cancer. *Future Oncol.* 2019, 15, 3381–3393, doi:10.2217/fon-2019-0282.
- Simeone, J.C.; Nordstrom, B.L.; Patel, K.; Klein, A.B. Treatment Patterns and Overall Survival in Metastatic Non-Small-Cell Lung Cancer in A Real-World, US Setting. *Future Oncol.* 2019, *15*, 3491–3502, doi:10.2217/fon-2019-0348.
- Khozin, S.; Miksad, R.A.; Adami, J.; Boyd, M.; Brown, N.R.; Gossai, A.; Kaganman, I.; Kuk, D.; Rockland, J.M.; Pazdur, R.; et al. Real-World Progression, Treatment, and Survival Outcomes during Rapid Adoption of Immunotherapy for Advanced Non-Small Cell Lung Cancer. *Cancer* 2019, 125, 4019–4032, doi:10.1002/cncr.32383.
- Schwartzberg, L.; Korytowsky, B.; Penrod, J.R.; Zhang, Y.; Le, T.K.; Batenchuk, C.; Krug, L. Real-World Clinical Impact of Immune Checkpoint Inhibitors in Patients with Advanced/Metastatic Non-Small Cell Lung Cancer After Platinum Chemotherapy. *Clin. Lung Cancer* 2019, 20, 287–296.e4, doi:10.1016/j.cllc.2019.04.004.
- Molife, C.; Hess, L.M.; Cui, Z.L.; Li, X.I.; Beyrer, J.; Mahoui, M.; Oton, A.B. Sequential Therapy with Ramucirumab and/or Checkpoint Inhibitors for Non-Small-Cell Lung Cancer in Routine Practice. *Futur. Oncol.* 2019, 15, 2915–2931, doi:10.2217/fon-2018-0876.
- 55. Levra, M.G.; Cotté, F.-E.; Corre, R.; Calvet, C.; Gaudin, A.-F.; Penrod, J.R.; Grumberg, V.; Jouaneton, B.; Jolivel, R.; Assié, J.-B.; et al. Immunotherapy Rechallenge after Nivolumab Treatment in Advanced Non-Small Cell Lung Cancer in the Real-World Setting: A National Data Base Analysis. *Lung Cancer* 2020, *140*, 99–106, doi:10.1016/j.lungcan.2019.12.017.
- Roberto, G.; Leal, I.; Sattar, N.; Loomis, A.K.; Avillach, P.; Egger, P.; Van Wijngaarden, R.; Ansell, D.; Reisberg, S.; Tammesoo, M.-L.; et al. Identifying Cases of Type 2 Diabetes in Heterogeneous Data Sources: Strategy from the EMIF Project. *PLoS ONE* 2016, 11, e0160648, doi:10.1371/journal.pone.0160648.
- Lythgoe, M.; Krell, J.; Warner, J.L.; Desai, A.; Khaki, A.R. Time Intervals between U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA) New Cancer Therapy Approvals. J. Clin. Oncol. 2021, 39, 1575, doi:10.1200/jco.2021.39.15_suppl.1575.
- Major, J.M.; Penberthy, L.; McGlynn, K.A. Data Systems and Record Linkage: Considerations for Pharmacoepidemiologic Studies Examining Cancer Risk. Ann. Epidemiol. 2016, 26, 746–748, doi:10.1016/j.annepidem.2016.08.015.
- Carrigan, G.; Whipple, S.; Taylor, M.D.; Torres, A.Z.; Gossai, A.; Arnieri, B.; Tucker, M.; Hofmeister, P.P.; Lambert, P.; Griffith, S.D.; et al. An Evaluation of the Impact of Missing Deaths on Overall Survival Analyses of Advanced Non-Small Cell Lung Cancer Patients Conducted in An Electronic Health Records Database. *Pharmacoepidemiol. Drug Saf.* 2019, 28, 572–581, doi:10.1002/pds.4758.
- 60. Sturkenboom, M.; Braeye, T.; van der Aa, L.; Danieli, G.; Dodd, C.; Duarte-Salles, T.; Emborg, H.-D.; Gheorghe, M.; Kahlert, J.; Gini, R.; et al. ADVANCE Database Characterisation and Fit for Purpose Assessment for Multi-Country Studies on the Coverage, Benefits and Risks of Pertussis Vaccinations. *Vaccine* 2020, *38*, B8–B21, doi:10.1016/j.vaccine.2020.01.100.
- Willame, C.; Dodd, C.; van der Aa, L.; Picelli, G.; Emborg, H.-D.; Kahlert, J.; Gini, R.; Huerta, C.; Martín-Merino, E.; McGee, C.; et al. Incidence Rates of Autoimmune Diseases in European Healthcare Databases: A Contribution of the ADVANCE Project. Drug Saf. 2021, 44, 383–395, doi:10.1007/s40264-020-01031-1.
- 62. FDA OPDIVO (Nivolumab) Injection, for Intravenous Use. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/125554s070lbl.pdf (accessed on 13 July 2021).
- 63. Orphanet. The Portal of Rare Diseases and Orphan Drugs. Small Cell Lung Cancer Schedule. Available online: https://www.orpha.net/consor/cgi-bin/Drugs_Search_Simple.php?lng=EN&LnkId=10953&Typ=Pat&fdp=y&from=rightMenu. (accessed on 13 July 2021).
- European Medicines Agency. OPDIVO (Nivolumab). Injection, for Intravenous Use. Available online: https://www.ema.europa.eu/en/documents/product-information/opdivo-epar-product-information_en.pdf (accessed on 13 July 2021).
- 65. Bajwa, S.S.; Theerth, K.; Gupta, A. The Increasing Trend of Observational Studies in Clinical Research: Have We Forgotten and Started Defying the Hierarchy? *Indian J. Anaesth.* **2021**, *65*, 186–190, doi:10.4103/ija.ija_176_21.

CHAPTER 3.

Identification and validation of a casefinding algorithm to extract NSCLC cases from the regional pathology registry of Tuscany
3.1 Introduction

In the scoping review of the literature, we found that the information necessary to appropriately characterized lung cancer patients and describe drug utilization is huge. Given that, the record linkage of data sources was confirmed to be the most appropriate methodology to study drug utilization and survival of patients with lung cancer.

The scoping review highlighted that cancer registries when linked to other rcEHD are a valuable data source for the conduction of pharmacoepidemiology studies in lung cancer. Even if the Tuscany region has a regional cancer registry, unfortunately, these data could not be linked to other datasources as an anonymized ID patients different from that utilized in Tuscany region administrative database was used. Nevertheless, the Tuscany region database includes multiple registries that can be linked with each other through a pseudo-anonymized regional person identifier code. Linked with administrative registries (i.e., hospital discharge records, outpatients encounter data) the Tuscany region has also a regional pathology registry. The latter covers the whole Tuscan population and contains free text field and codes useful for the identification of patients with NSCLC.

3.1.1 Pathology registries to conduct observational studies in oncology

Diagnostic histological and cytological specimens are routinely stored in pathology department archives. These biobanks are a valuable research resource for many diseases, particularly if they can be linked to high quality electronical healthcare data, allowing large retrospective epidemiological/pharmaco-epidemiological studies to be carried out.⁶⁷ In other settings, such as Denmark, it was demonstrated that in the era of personalized medicine, access to high quality pathology registries provides a valuable clinical and research resource.⁶⁷ In fact, pathology

registries can provide a valid complement to other data sources that are widely used for research purposes, such as hospital discharge records, where information on histology is lacking.⁶⁸ However, validity of variables retrieved from healthcare data is imperfect, and its measurement is necessary to assess reliability of such evidence.⁶⁹ Validated algorithms can be used to identify patients with specific diseases and to extract other relevant information (i.e. histology) which are not presented in administrative datasources.

Anyway, no evidence from the literature using such registries for the conduction of pharmacoepidemiologic studies for lung cancer patients was found in literature.⁷⁰

3.1.2 SNOMED coding in pathology registries

The Systematized Nomenclature of Medicine (SNOMED) is a systematic collection of medical terms, in human and veterinary medicine, providing codes, terms, synonyms and definitions which cover anatomy, diseases and findings. It allows to index, retrieve, and aggregate medical data across specialties and sites of care. SNOMED was started in the U.S. by the College of American Pathologists in 1973 and revised into the 1990s.⁷¹ In 2002 College of American Pathologists's SNOMED Reference Terminology was merged with the National Health Service's Clinical Terms Version 3 to produce SNOMED Clinical terms (SNOMED CT).

Versions of SNOMED released prior to 2001 were based on a multiaxial, hierarchical classification system. In such versions, a disease may be located in a body organ, which results in a code in a topography axis, and may lead to morphological alterations represented by a morphology code.

As for SNOMED CT, the primary purpose is to support the effective clinical recording of data with the aim of improving patient care. It is a structured collection of medical terms that are used internationally for recording clinical information and are coded in order to be computer processable.⁷²

Despite the advent in years 2000 of SNOMED CT,⁷³ in the report of 2016 "Assessing SNOMED CT for Large Scale eHealth Deployments in the EU" most part of European nations reported that

their use of SNOMED CT was very limited (**Figure 3.1**).⁷⁴ The pathology registry of Tuscany (as well as other pathology registries, at international level across all Europe, i.e. Denmark)⁶⁷ still use previous SNOMED version of nomenclature.



3.2 Study objectives

Given that the regional pathology registry of the Tuscany region, was not yet used for the

conduction of pharmacoepidemiology study on cancer, the development and the validation of a

NSCLC case-finding algorithm patients was needed.

The objective of this study was to develop and to validate a case-finding algorithm for the

identification of NSCLC patients in the regional pathology registry of Tuscany.

3.3 Publication

Development and validation of a case-finding algorithm for the identification of non-small cell lung cancers in a region-wide Italian pathology registry

Submitted on 17/07/2021

Abstract

Purpose: To develop and validate a case-finding algorithm for the identification of NSCLC cases in a region-wide Italian pathology registry. **Methods:** Data collected between 2009 and 2017 in the pathology registry and the Pharmacy Database of the University Hospital of Siena and the pathology registry of Tuscany region were used. A NSCLC-identification algorithm based on free-text keywords and SNOMED morphology and topography codes was designed and tested on data from Siena: indication for drug use (i.e., NSCLC) was the reference standard for sensitivity; positive predictive value was estimated through manual review. Algorithm modifications were then tested to improve algorithm performance: positive predictive value was estimated in pathology registry of Tuscany using analytical formulae that assumed NSCLC incidence equal either to 80% or 90% of overall lung cancer incidence recorded in Tuscany. The algorithm modification with the best performance was chosen as the final version of the algorithm. A random sample of 200 cases was extracted from the pathology registry of Tuscany for manual review. **Results:** The first version of the algorithm showed a positive predictive value of 74.7% and sensitivity of 79% in pathology registry of Siena. The final version of the algorithm had a

sensitivity in pathology registry of Tuscany that grew with calendar time (2009=[24.7%-28%]; 2017=[57.9%-65.1%]) and a positive predictive value of 93%. **Conclusion:** The final NSCLC-finding algorithm showed with very high positive predictive value. SE was in line with the expected contribution of pathology registry to overall cases captured in the regional Cancer Registry, with a trend of increase over calendar time. Given the promising algorithm validity and the wide use of SNOMED terminology in electronic pathology records, the proposed algorithm is expected to be easily adapted to other electronic databases for (pharmaco)epidemiology purposes.

Introduction

Lung cancer is the most commonly diagnosed cancer worldwide (2.09 million cases in 2018) with approximatively 42.500 new diagnosis in Italy in 2019.⁷⁵ NSCLC represents about 85% of all cases of lung cancer and can be distinguished in three main histotypes: squamous, big cells, and adenocarcinoma.⁷⁶ In the past fifteen years, a number of novel anticancer medications, such as target- and immunotherapies, have been authorized for the pharmacological treatment of advanced stage NSCLC. ⁷⁷ Nevertheless, knowledge on the efficacy and safety of such anticancer drugs mostly relies on evidence from clinical trials which are usually based on relatively small samples of selected patient populations.^{78,79}

Large databases of routinely collected electronic healthcare data can be used to generate evidence on the epidemiology of diseases as well as on the real-world utilization of medicines in population that were not included in clinical trials.^{80,81} However, validity of variables retrieved from healthcare data is imperfect, and its measurement is necessary to assess reliability of such evidence.⁶⁹

The aim of the study was to develop and to validate a case-finding algorithm for the identification of patients with NSCLC in a region-wide pathology registry in Italy.

Material and methods

Data sources

For the purposes of this study, three data sources were used: 1) the pathology registry of the University Hospital of Siena, which is a city in Tuscany that hosts the University Hospital "Le Scotte", one of the major hospitals of the region, 2) the Hospital Pharmacy database of Siena and 3) the region-wide administrative healthcare database of Tuscany.

- Pathology Registry of the University Hospital of Siena

The pathology registry of Siena collects records of histological and cytological examinations of patients attending the University Hospital of Siena. It contains patient demographics and pathological diagnoses, which are both recorded using the SNOMED morphology and topography codes, and free-text description of macroscopical and microscopical findings, as well as other clinical information. Biomolecular characterization (i.e., EGFR positive mutations) can be also recorded with SNOMED codes.

- The Hospital Pharmacy database of Siena

The Hospital Pharmacy database of Siena (HPS) collects information concerning intravenous drugs prepared by the hospital pharmacy and administered to both inpatients and outpatients treated at the Siena University Hospital. The database contains demographic data of treated patients as well as information on the administered drug, including indication of use.

- Regional administrative healthcare database of Tuscany

The administrative healthcare database of Tuscany collects information on all healthcare services dispensed to Tuscan inhabitants (around 3.7 million subjects) and reimbursed by the National Healthcare Service. This database includes multiple registries that can be linked with each other through a pseudo-anonymized regional person identifier code. In this study, two registries were used: the inhabitant registry, which records demographic information, including vital status of all inhabitants entitled to public healthcare assistance, and the regional pathology registry. The latter

covers the whole Tuscan population and contains all variables available in the pathology registry of Siena, except for molecular characterization.

Design of a case-finding algorithm for NSCLC identification

The first version of the case-finding algorithm for the identification of NSCLC cases was designed by a multi-expertise work group including clinicians from the Pathology Department of the Siena University Hospital, epidemiologists, and information technology experts. The population of cases the algorithm was intended to identify corresponded to the intersection of cases identified by a morphological (Morpho_1) and a topographical component (Topog_1). Morpho_1 was developed to identify histologies that fall within the pathological definition of NSCLC, while Topog_1 aimed at identifying the lung site. For both components, topography, and morphology SNOMED codes together with keywords from the free text diagnosis field were used.

Application and validation of the first version of the NSCLC case-finding algorithm against the training dataset

The first version of the case-finding algorithm was applied to all the records collected in the pathology registry of Siena, i.e. the training dataset, between 01/01/2009 and 31/08/2017. The competent authority assigned the pseudo-anonymized ID to each extracted case in order to allow the linkage to the regional inhabitant registry of Tuscany.

Sensitivity (SE) of the algorithm was estimated using the indication for drug use recorded in HPS as the reference standard. Cases labeled as NSCLC by the first version of the algorithm were linked to all patients with a record of an administration of any drug used to treat NSCLC in the HPS database: assuming that SE of the algorithm in the population of patients treated for NSCLC was the same as in the NSCLC general population, the proportion of cases retrieved among patients that received a drug treatment for NSCLC was used as the estimate of SE. Positive predictive value (PPV) was calculated based on the results from a manual review of records of all cases of NSCLC identified in the pathology registry of Siena (*reference standard*). Information used for the manual review was extracted from the free-text fields (pathological diagnosis, macroscopy, microscopy, clinical information and sent material) of the first record identified for each patient during the study period. The information was extracted manually and independently by two researchers (AS and PR) which worked blinded to each other. A senior pathologist (CB) from Siena University Hospital resolved disagreements between the two assessors. Data were extracted in a standard extraction form, encompassing four items: 1) Malignancy: Malignant tumor, Yes/No/Maybe; 2) Topography: Lung, Yes/No/Maybe; 3) Morphology: NSCLC, Yes/No/Maybe; and 4) Origin of the tumor: primary, Yes/No/Maybe. Those cases that received "*yes*" in all the four validation items, or those who received "*yes*" in the first three validation items and "*maybe*" with respect to the primary origin of the tumor were considered as confirmed NSCLC cases. Cohen's Kappa for concordance between assessors was also calculated.

Refinement of the first version of the NSCLC case-finding algorithm: validation of different algorithm modifications

Based on the experience from the manual review, we tested the impact of several algorithm modifications. Namely, a set of additional algorithm *components* were used as exclusion criteria with respect to the population of cases identified by the first version of the algorithm. The validity of the tested algorithms modifications was then estimated as follow:

- PPV: the validated dataset extracted from pathology registry of Siena (see above) was used as the reference standard.

- SE: analytical formulae based on algebraic interrelation among PPV, SE, observed incidence, P, and true incidence, \prod , were used.^{82,83} P was obtained from direct application of the relevant algorithm variant in the pathology registry of Tuscany. \prod was obtained from the most recent estimate of overall lung cancer incidence in the Region, i.e. 2014, reported from the Cancer

Registry of Tuscany.⁸⁴ namely, as NSCLC is known to represent about 85% of all lung cancer cases,⁷⁵ we considered two extreme values of \prod , corresponding to 80% and 90%, respectively, of the overall lung cancers in order to obtain a range [min-max] of SE values. We also assumed that \prod did not change during study period.

The combination of algorithm components with the best performance in terms of SE and PPV was then selected as final version of the algorithm.

Application of final version of the NSCLC case-finding algorithm to the pathology registry of Tuscany

The PPV of the final version of the algorithm in the general Tuscan population was estimated in the pathology registry of Tuscany. A random sample of 200 cases retrieved from 2009 to 2019 was extracted and manually validated by one researcher (AS) who consulted two other experts (PR and CB) for resolving any doubt on the assessment.

Results

Application of the first version of the algorithm and validation against the training dataset

SNOMED codes and free-text keywords selected for the components Morpho_1 and Topog_1 of the first version of the algorithm are reported in **Table 3.1**.

The algorithm allowed to retrieve 2309 cases from the pathology registry of Siena during the period 01/01/2009-31/08/2017, of which 2003 were linkable to the inhabitant registry of the regional administrative healthcare database of Tuscany and, thus, were considered for manual chart validation.

As for PPV, out of 2003 cases labeled as NSCLC cases, 1496 were classified as confirmed cases, leading to an estimate of 74.7% for PPV (see **Table 3.2**).

Table 3.1. Components of the first version of the algorithm used for the identification of

Acronym of algorithm's	Description	Tumor	Terminology	Codes/free text words	Criteria
component		characteristic			
Morpho_1	NSCLC codes	Morphology	SNOMED	M-81403 OR M-80703 OR	Inclusion
	and free text		and free text	M-82463 OR M-85603 OR	
				M-83233 OR-85503 OR	
				M-84303 OR M-82503 OR	
				M-80033 OR M-80123 OR	
				M-80463 OR (M-8* AND	
				(Text in diagnosis	
				field:"*non microcit*" OR	
				"*adenocarcin*"	
				"*squamocell*" OR	
				"*grandi cell*"))	
Topog_1	Chest	Topography	SNOMED	T-2* OR Text in diagnosis	Inclusion
	topography		and free text	filed: "POLM*" OR	
				"BRONCH*"	

NSCLC cases in the pathology registry of Siena*

* First version of the algorithm corresponded to the intersection of the two components reported in the table, i.e. Morpho_1 AND Topog_1.

Table 3.2. Manual validation results of first version of the algorithm

Population	Malignancy:	Topography:	Morphology:	Origin:	
	Malignant tumor, n (%)	Lung, n (%)	NSCLC, n (%)	Primary, n (%)	
n=2003 patients from the	e Yes: 1992 (99.46%)	Yes: 1754	Yes: 1836 (91.66%)	Yes: 995 (49.67%)	
pathology registry of	No: 2 (0.09%)	(87.56%)	No: 93 (4.64%)	No: 292 (14.57%)	
Siena	Maybe: 9 (0.45%)	No: 180 (8.98%)	Maybe: 74 (3.69%)	Maybe: 716	
		Maybe: 69		(35.74%)	
		(3.44%)			

Cohen's k measure of concordance between reviewers was as follows: 1) k=0.1350 for the malignancy item (slight agreement); 2) k=0.7617 for the lung item; (substantial agreement); 3) k=0.4307 for the morphology item (moderate agreement); 4) k= 0.3128 for the origin item (fair agreement).

Patients receiving a drug treatment with a recorded NSCLC indication in HPS were 469: among those, 373 were identified by the first version of the algorithm, leading to an estimate of 79,5% for SE.

Refinement of the first version of the NSCLC case-finding algorithm: validation of different algorithm modifications

Manual validation identified SNOMED codes and free-text keywords that were used to design additional algorithm components as exclusion criteria for the improvement of the first version of the algorithm (**Table 3.3**).

Acronym of algorithm's	Description	Tumoral	Terminology	Codes/free text words	Criteria
component		characteristic			
Morpho_K_neuroend	Neuroendocrine carcinoma	Morphology	SNOMED	M-82463	Exclusion
Topog_Upper_airways	Upper airways	Topography	SNOMED	T-21* OR T-22* OR T-23* OR T-24* OR T-25*	Exclusion
Topog_Resp_Not_specified	Respiratory system not specified	Topography	SNOMED	T-20*	Exclusion

 Table 3.3. Additional algorithm components used for algorithm modifications

Two algorithm modifications were obtained based on different logical combination of additional algorithm components (**Table 3.4**).

Table 3.4. First version of the algorithm and algorithm modifications

First version of the algorithm

Morfo_1 AND Topog_1

First modification

(Morfo_1 AND NOT Morfo_K_neuroend) AND ((Topog_1 AND NOT (Topog_Upper_airways OR

$Topog_Resp_Not_specified))$

Second modification*

(Morfo_1 AND NOT Morfo_K_neuroend) AND (Topog_1 AND NOT Topog_Upper_airways)

*Chosen as final version of the algorithm

The two algorithm modifications showed similar results in terms of PPV and SE. The algorithm modification with the best performance showed a PPV of 87.9% and an increasing sensitivity over the years (**Figure 3.2**): in 2009 SE range was [min=24.9%-max=28.0%] while in 2017 SE range was [min=57.9%-max=65.1%].



Figure 3.2. Positive predictive value (PPV) and sensitivity (SE) of first version and final version of the algorithm

The figure shows the validity measures of the first and the final version of the algorithm. PPV was represented in red and SE in green (dark green – higher value of SE; light green – lower value of SE)

This algorithm modification was then chosen as final version of the NSCLC case-finding algorithm for the regional pathology registry of Tuscany.

Performance of final version of the algorithm in regional Pathology Registry of Tuscany

The algorithm was applied to the records collected in the pathology registry of Tuscany between 2009 and 2019. Overall, 15,169 cases were identified. A random sample of 200 cases was validated manually (see **Table 3.5**). The number of confirmed NSCLC cases was 186/200, corresponding to a PPV of 93%.

Population Malignancy: **Topography:** Morphology: **Origin:** Malignant tumor, n (%) NSCLC, n (%) Primary, n (%) Lung, n (%) Yes: 200 (100%) Yes: 200 (100%) Yes: 34 (17%) n=200 patients from Yes: 190 (95%) PR of Tuscany No: 1 (0.5%) No: 4 (2%)

Maybe: 9 (4.5%)

Maybe: 162 (81%)

 Table 3.5. Manual validation results of final version of the algorithm

Discussion and conclusion

In this paper we reported our experience concerning the development and validation of a casefinding algorithm for the identification of NSCLC cases in a region-wide Italian pathology registry. The algorithm was found to have an high PPV and an acceptable SE which estimates appeared to increase over the study period and could be considered in line with the share of regional NSCLC cases that are likely to be captured by the pathology registry.⁸⁵ Notably, two possibly concurrent reasons can explain the observed trend of increase of the estimated SE. First, the regional pathology registry of Tuscany was established in 2006 so that completeness of data recording was expected to increase over the years. Second, the establishment of clinically and scientifically accredited regional centers during the last decade might have contributed to attract an increasing number of patients towards Tuscan hospitals, rather than to other out-of-the-region facilities. Thus, over the study period, the regional pathology registry of Tuscany have reasonably accounted for an increasing share of the total NSCLC cases captured by the regional Cancer Registry, which is also feed by hospital discharge records and death certificates.⁸⁵

The process adopted in this study for developing the final version of the NSCLC-finding algorithm was stepwise. Manual ascertainment of the 2003 cases extracted through the application of the first version of the algorithm to the training dataset allowed to identify two additional exclusion criteria, concerning neuroendocrine morphology and upper airways topography, respectively. Adding the latter two components to the first version of the algorithm allowed to increase PPV without compromising SE.

Given the promising evidence on validity reported in this work, the final version of the NSCLCfinding algorithm designed and tested in this study can be considered suitable for epidemiology and pharmacoepidemiology studies. Pathology registries can therefore provide a valid complement to other data sources that are widely used for research purposes, such as hospital discharge records, where information on histology is often lacking.⁶⁸ Notably, information on pathological staging (TNM – see section 1.3) and histology can be also found in the pathology registry, however assessing the completeness and validity of such variables in the pathology registry of Tuscany was out of the scope of this work.

This study has several strengths. First, to the best of our knowledge this is the first validation study concerning the application of a NSCLC case-finding algorithm in an Italian pathology registry. Since the algorithm developed in this study was based on SNOMED codes and free-text keywords, this work will serve as the starting point for local NSCLC case-finding algorithm design and application. In fact, SNOMED terminology is widely used in electronic pathology reports⁸⁶ so that the proposed algorithm is expected to be easily adapted to other electronic databases. Second, two manual case validations were performed in this study. In particular, the double-blinded manual validation of the 2003 cases identified in the training dataset (i.e., the pathology registry of Siena)

allowed gaining insight into the specific codes and keywords used in the pathology reports which were finally adopted for algorithm refinement. Third, an innovative methodology for estimating missing validity indices was used.^{82,83} This allowed obtaining also estimates of SE, a validity measure that is usually difficult to obtain and, thus, rarely reported in validation studies.⁸⁷ However, SE estimations reported in this work have limitations that should be taken in consideration. For the first version of the algorithm, for which the indication for drug use recorded in the Hospital Pharmacy database of Siena was used as the reference standard, we assumed that SE was the same in both pharmacologically treated and untreated patients. For the final version of the algorithm, the adopted reference standard was the most recent estimate of the overall lung cancer incidence from the regional Cancer Registry of Tuscany (i.e., 2014), which we assumed that remained stable during the whole study period. Considering the evidence of a possible trend of slight decrease of the overall lung cancer incidence during the same period,⁸⁸ the latter assumption might have led to an under-estimation of SE of the final version of the proposed NSCLC-finding algorithm. Notably, incidence of overall lung cancer cases was only available so we used two extreme values of the expected share of NSCLC cases (80%-90%) which allowed to obtain a range of SE estimates,⁷⁶ within which the true SE was likely to actually fall. Another study limitation concerns the classification of confirmed cases based on the results of the manual review. PPV value was calculated considering as confirmed NSCLC cases also those for which the primary origin of the tumor could not be excluded based on all the available information recorded. This choice was made because pathology registries do not always contain information on the tumor's origin (see Table 3.2 and Table 3.5). However, to reduce possible misclassification of secondary NSCLC, the first available record in the pathology registry was considered for each patient. Nevertheless, for future applications of the present algorithm, other sources of data should be used whenever linkable to pathology registry in order to refine the selection of primary NSCLC cases (e.g., hospital discharge registries).

In conclusion, with this study we designed and validated a case-finding algorithm for the identification of NSCLC cases in a region-wide pathology registry. The algorithm was found to have a high PPV (i.e., 93%) and an acceptable SE which estimates appeared to increase over the study period and could be considered in line with the share of regional NSCLC cases that are expected to be captured by the pathology registry. Given its promising validity the proposed NSCLC-finding algorithm can be considered suitable for future epidemiology and pharmacoepidemiology studies. Given the wide use of SNOMED terminology in electronic pathology reports, the proposed algorithm is expected to be easily adapted to other electronic databases.

CHAPTER 4. Drug utilization and survival of patients with non-resectable NSCLC: a real-world study with gender perspective

4.1 Introduction

While in the scoping review most part of the included studies investigated the drug utilization of innovative drugs in relation to histology, only few studies reported some of the presented result stratified by sex (not taking into consideration the patients demographical characteristics table).^{89,90} Real-world evidence can be useful to describe treatment efficacy in a "clinical practice" population characterized by lower selection bias compared with randomized clinical trials, adding details on the outcome of special patients' populations (i.e. men and women), under-represented or excluded from clinical trials. Of course, the absence of a randomized group does not allow to estimate the efficacy compared with other treatments, and quality of data could be lower compared with clinical trials.⁹¹ Anyway, given the treatment revolution of lung cancer in the last decade, a descriptive study about use of these drugs and the survival of patients diagnosed in an advanced stage NSCLC in relation to gender could provide a picture in real-world setting.

4.1.1 Gender difference in cancer

The epidemiology, the natural history of the diseases and the outcomes are gender-specific, but in the daily clinical practice there are still many disadvantages for women.^{92,93} The differences, or at least some of them, are known, but they are unfortunately often not translated into clinical practice. For example, the differences in body size and composition (women are usually shorter and thinner than men and have more adipose tissue and less muscle mass than males) have important consequences on the parameters of pharmacokinetic and toxicokinetic of drugs and therefore should be taken into account when determining the dosage of drugs, but even now unfortunately the average dose is set considering a Caucasian man of 70 kg.^{94,95}

If we analyze what has happened in the recent past, it emerges that preclinical and clinical research has been mainly, if not exclusively, conducted on male subjects.⁹⁶ The application of a sex-gender

approach, as well as contributing to the appropriateness and practice of evidence-based medicine, can lead to a reduction in the cost of care. In fact, healthcare costs are higher in women, even after subtracting those related to the birth pathway,^{97,98} perhaps because they receive less appropriate care.⁹⁹

As for anti-cancer therapy, given the growing interest and use of immunotherapy, it should be remembered that there is a distinct immune profile between males and females and this plays an important role in the response to these drugs.¹⁰⁰ Hormonal and genetic differences contribute to the sexual dimorphism of the immune system. The 17β -estradiol - estrogen receptor α axis is the main regulator of innate immune populations (17β -estradiol is able to reduce the mobility and inflammatory activity of neutrophils),¹⁰¹ and as it well known male and female have different hormone profile. The hormonal difference affects both innate and adaptive immune responses, (**Figure 4.1**), leading to a significant functional diversity between females and males.



Source: Capone I, Marchetti P, Ascierto PA, Malorni W, Gabriele L. Sexual Dimorphism of Immune Responses: A New Perspective in Cancer Immunotherapy. *Front Immunol*. 2018 Mar 21;9:552

4.1.2 Gender difference in NSCLC

Epidemiology features

Among NSCLC, adenocarcinoma is currently the most common subtype both in men and women, but women present with proportionally more adenocarcinomas and fewer squamous cell carcinomas than men.^{102,103} As already reported in the introduction section of the thesis (see section 1.1), smoking habits explain the difference in histological distribution, but other genetic and hormonal factors are likely to contribute.

Women and men also present different frequencies about molecular mutations which can be target by drugs (see section 1.4.2). About 15% of all NSCLC presents EGFR mutations: anyway, these are found with higher frequency in women and Asians people.^{104,105} ALK translocations are more common among non-smoker patients, non-squamous cancer histologies and at a younger age. In this case, gender conflicting results have been reported.^{106,107}

Treatment and outcomes

Many studies have shown better overall survival rates in women with early stage lung cancer after surgical resection, both in NSCLC and SCLC.^{108–112} A meta-analysis of 39 publications including only studies that compared overalls survival data for men and women for a total of about 75,000 patients, reported that survival of women was significantly better than the survival of men.¹¹³ Moreover women have also a significantly longer survival than men independently from stage and treatment or age at diagnosis.¹¹⁴

As for SCLC patients an improved benefit from chemotherapy has been observed for women compared to men.^{115,116} However, in advanced NSCLC available literature is controversial.^{117,118} Women reported a greater toxicity from certain standard chemotherapy drugs,¹¹⁸ but the choice of chemotherapy is currently not influenced by the patient's sex. As for immunotherapy, the analysis of overall survival carried out across patient subgroups in the pivotal studies of nivolumab and

pembrolizumab did not report a significant difference in the hazard ratio between women and men.^{119,120}

Since the beginning target therapy use in NSCLC patients, the female gender was associated to better drug activity to EGFR inhibitors, and then this variable was lost in the large amount of data involving genetic aberrations used in the current context of precision medicine in NSCLC and the *gender* was missed as variable. Precision medicine has revealed that the genomics could represent the axis of drugs prescription; however, there are classic prognostic variables that can be taken in consideration for precision medicine. Gender is one of the oldest prognostic variables in oncology that is being more and more left behind in the current research.¹²¹

4.2 Study objectives

The scoping review highlighted a paucity of studies conducted in Europe, and without taking gender as stratification variable. Moreover, the NSCLC-case finding algorithm demonstrated good validity measures and was considered suitable for pharmacoepidemiology studies. We decided to describe the drug utilization and the survival of patients with a non-resectable, primary NSCLC with gender perspective. We take the decision to study only non-resectable NSCLC patients because are those for which pharmacotherapies in the last decade was revolutionized.

Thus, the study objective was to describe the drug utilization and the survival of patients with a non-resectable, primary NSCLC also including a gender perspective. A protocol was written and submitted to the EU PAS register of the *European Network of Centers for Pharmacoepidemiology and Pharmacovigilance* (EUP PAS register number EUPAS36692).

4.3 Publication

First-line pharmacotherapies and survival among patients diagnosed with non-resectable NSCLC: a real-life setting study with gender prospective

Submitted on 15/10/2021

Abstract

Purpose: This study aims to describe first-line pharmacotherapy and overall survival in nonresectable NSCLC patients by gender. Methods: Incident cases of non-resectable NSCLC recorded between 2009 and 2019 (cohort entry) in the pathology registry of the regional administrative healthcare database of Tuscany were identified. Records of antineoplastic therapies delivered up to 4 months following cohort entry were classified as chemotherapy, target therapies, immunotherapies, undefined monoclonal antibodies. First-line treatment and survival of patients receiving drug treatment was described. Analyses were stratified according to histology, gender and cohort entry year. Results: 4,393 incident cases of non-resectable NSCLC were included. Women with non-squamous-NSCLC received a target-therapy more frequently than men (14.9% vs. 6.5%). Immunotherapy incidence of use varied between 3.8% (2017) and 9.1% (2019). The two-year survival rate increased over time: for non-squamous-NSCLC was 22.3% (2009-2011) and 30.6% (2018-2019), while for squamous-NSCLC was 13.5% and 22.5%, respectively. Among nonsquamous-NSCLC, median survival was longer in women than in men (389 vs. 276 days). Conclusion: Considering sex-related biomolecular differences, among non-squamous-NSCLC women received target-therapy more frequently than men. Survival improved over the study period, for both histologies.

Introduction

Lung cancer is the most diagnosed cancer in the world. It represents 14.5% of total cancer cases in men and 8.4% in women, being the leading cause of cancer death in men (22.0%). As a result of the reduction in the smoking habit in men and the increase in women, in recent years there has been a marked reduction in the incidence of this cancer in men, amounting to -1.6% per year, while in women there has been an average increase of 1.7% per year.^{14,122}

Based on histological characteristics, the World Health Organization divides lung cancer into two classes: SCLC and NSCLC. The latter represents 85% of all cases of lung cancer and its diagnosis is usually based on diagnostic imaging (chest X-ray and computed tomography), and the anatomopathological evaluation of a biological sample.¹⁴

The two main histological subtypes of NSCLC are: squamous and non-squamous carcinoma (mainly adenocarcinoma and large cell carcinoma). According to epidemiological data, the latter subtype is proportionally much more frequent in women than in men.¹²³ Regardless of histological type, in the early stages of the disease the NSCLC shows no obvious clinical symptoms, and the first-choice treatment is the surgical resection of the neoplastic tissue. However, it often happens that the diagnosis of NSCLC is done when the tumor is already advanced or metastatic and therefore no longer eligible for surgical resection and so drug treatment becomes the first therapeutic choice.

During the last decade, the pharmacological treatment of non-resectable NSCLC, has undergone a real revolution thanks to the licensing of different innovative antineoplastic medications, namely target therapy and immunotherapy drugs.¹⁴ The first target therapy, gefitinib, an anti-EGFR drug, was approved in 2009 by EMA, followed by Crizotinib, an ALK inhibitor, in 2012. More recently also immunotherapies like nivolumab (2015) were approved by EMA as second/third line treatment for advanced stage NSCLC, while in 2016 the use of pembrolizumab as first-line treatment was also approved.

Up today, real-world evidence on the utilization of innovative anticancer treatments for NSCLC is still poor, particularly in Europe.⁷⁰ With this respect, existing electronic healthcare databases that routinely collect information from everyday clinical practice have a tremendous potential for filling such gaps of knowledge. Moreover, the importance of sex and gender as modulator of treatment use and outcomes in cancer patients is still an under-evaluated issue, and the reporting of results by sex in publications need to be reemphasized in oncology.¹²⁴ These sources of observational data can be re-used to execute large pharmaco-epidemiological studies to generate evidence on the utilization of anti-cancer medications in the real-word setting by gender.^{80,81}

Since around two decades, in Tuscany region, in Italy, a region-wide population-based administrative healthcare database collects information on healthcare services reimbursed by the National Health Service and delivered to Tuscan inhabitants (around 3.7 million subjects). Although the administrative healthcare database of Tuscany has been extensively used for pharmaco-epidemiological studies, ^{125–127} no experience currently exists in linking patient-level data from the different available registries to the information recorded in the regional pathology registry to describe the utilization of anticancer medications.

The aim of this study was to describe survival and drug utilization of first-line anticancer therapies approved for the treatment of advanced, non-resectable, NSCLC patients in relation to gender by analysing data from the regional administrative database of Tuscany.

Methods

Data sources

Data from the regional administrative database of Tuscany region were used. The database collects longitudinal pseudonymized patient-level information on the utilization of healthcare services reimbursed by the National Healthcare Service and dispensed to all Tuscan inhabitants who are registered with a general practitioner. For each subject registered in the inhabitant registry of the database, demographic data and vital status can be linked to different registries. In particular, the

regional pathology registry was used for the identification of patients with NSCLC. This registry contains free-text fields (i.e., pathological diagnosis) and *Systematized NOmenclature of human and veterinary MEDicine* (SNOMED) codes concerning pathological evaluation reports coming from regional hospitals. Moreover, data from three different registries were used to retrieve information on drug treatments: 1) the *Outpatient Drug Dispensings* (ODD) registry, which records the specific medicinal product used during outpatient encounters or dispensed for out-of-hospital use (high-cost drugs used during hospitalization are in some cases recorded, although it rarely occurs), 2) the registries of *Hospital Discharge Records* (HDR) and 3) the *Outpatient Encounter and diagnostic procedures Data* (OED) in which infusive antineoplastic administration procedures executed during hospitalization and outpatients encounters, respectively, are recorded as either unspecified standard chemotherapy or monoclonal antibody therapy .

Cohort selection

NSCLC cases recorded between January 2009 and June 2019 were extracted from the regional pathology registry through the application of an identification algorithm developed *ad hoc* for the objectives of the study. As described elsewhere, the algorithm was based on the selection of records containing SNOMED codes and keywords used within the free-text description of the pathological diagnosis.¹²⁸ The date of the first record in pathology registry was considered the cohort entry date. To identify patients who with a primary cancer diagnosis, those with a record of non-pulmonary cancer diagnosis identified in AHD during the five years prior to the cohort entry date were excluded. Underage patients and those with less than two years of look-back in AHD were also excluded. Finally, to select incident cases of primary, non-resectable NSCLC cases, only patients without a lung cancer diagnosis recorded in ADH from three months to five years prior cohort entry date were retained in the study cohort. The selected cases were subsequently stratified in three sub-cohorts based on NSCLC histology: squamous, non-squamous, or unknown histology.

First-line identification

The pharmacological treatment of cancer is rarely represented by the administration of a single anticancer drug but more often it is based on the administration of more than one antineoplastic drug administered during a time frame of theoretically maximum 21 days (i.e., one treatment cycle). For each patient in the study cohort, the first record of dispensing/administration of any antineoplastic medication observed in ODD, HDR and OED during the four months following the cohort entry date was considered as the start date of the first treatment cycle. All the subsequent record of drug treatment occurred during 21 days following the date of initiation of the treatment cycle were identified. The observed treatment sequences were then grouped in four main treatment categories: standard chemotherapy, target therapy, immunotherapy and therapy based on unspecified monoclonal antibodies (obtained from records of drug administration procedures identified in HDR and OED specific codes).

Patients who had no drug treatment registered within the 4 months following cohort entry were classified as untreated unless a record of radiotherapy was found in HDR and OED.

Survival

Overall survival (OS) of the selected cohort of patients with NSCLC was described. In particular, the survival of patients receiving any first-line pharmacotherapy was observed based on the year of cohort entry and the study cohort was categorized as follows: 2009-2011, 2012-2014, 2015-2017, and 2018-2019.

Survival time of patients receiving a first-line pharmacotherapy was defined as the time between patient's date of first pharmacotherapy and death. A Kaplan-Meier curve was plotted, and the analysis was stratified on patients' histology characteristics (squamous and non-squamous), and gender (male and female). Also, the median OS (CI: 95%), the two-years survival and five-years survival were calculated.

Statistical analysis

Descriptive analyses were conducted to assess demographic and clinical characteristics of selected NSCLC patients in relation to the histology of NSCLC (squamous, non-squamous, unknown). Continuous variables were described by means and standard deviation or by median and range while categorical variables were described by patient counts and percentages. Temporal trends of the proportion of patients by first-line treatment received were evaluated using a time series model. Mann-Kendall test was used to assess the downward or upward statistical significance of the observed trends. R software was used to perform the analyses.

Results

Cohort characteristics and first-line antineoplastic medications

Overall, the case-finding algorithm identified 11,335 patients with NSCLC in the pathology registry of Tuscany region. Of these, a total of 4,393 incident patients with unresectable primary NSCLC were included in the analysis and followed until June 2020. Cohort selection is reported in detail in

Figure 4.2.



Within the selected cohort 2,793 patients (63.6%) had a non-squamous, 1,559 (35.5%) a squamous, and 41 (0.9%) unknown histology (**Table 4.1**).

	Non squamous		Squamous		Unknown H	Overall	
	n=2,793		n=1,559		n=41		n=4,393 (%)
	Women	Men	Women	Men	Women	Men	-
	n=1,057 (%)	n=1,736 (%)	n=289 (%)	n=1,270 (%)	n=13 (%)	n=28 (%)	
Age at cohort entry date,	67.5 (11.5)	69.6 (9.7)	70.7 (9.0)	71.8 (9.2)	68.9 (11.1)	70.6 (10.7)	69.8 (10.1)
mean (SD)			()			/0.0 (10.7)	
Age bands at cohort entry							
date							
18-54	144 (13.6)	136 (7.8)	17 (5.9)	67 (5.3)	2 (15.4)	4 (14.3)	370 (8.4)
55-69	424 (40.1)	671 (38.7)	100 (34.6)	395 (31.1)	3 (23.1)	6 (21.4)	1,599 (36.4)
70-84	434 (41.1)	856 (49.3)	163 (56.4)	732 (57.6)	8 (61.5)	15 (53.6)	2,208 (50.3)
85+	55 (5.2)	73 (4.2)	9 (3.1)	76 (6.0)	0	3 (10.7)	216 (4.9)
Time in the study cohort,	338.0	211.0	201.0	228.0	400.0	134.0	239.0
median [IQR]	[113 - 685]	[72 -533]	[70 - 424]	[78 - 444]	[140 - 750]	[50 - 370]	[80 - 540]
First-line pharmacotherapies							
Standard chemotherapy	526 (49.8)	921 (53.1)	151 (52.2)	694 (54.6)	6 (46.2)	13 (46.4)	2,311 (52.6)
Target therapies	158 (14.9)	112 (6.5)	3 (1.0)	2 (0.2)	4 (30.8)	0 (0.0)	279 (6.4)
Immunotherapy	33 (3.1)	52 (3.0)	7 (2.4)	19 (1.5)	0 (0.0)	0 (0.0)	111 (2.5)
Unspecified anti-body	10 (0.0)	22 (1 2)	1 (0.2)	12 (0.0)	0 (0 0)	0 (0 0)	46 (1.0)
therapy	10 (0.9)	23 (1.3)	1 (0.3)	12 (0.9)	0 (0.0)	0 (0.0)	46 (1.0)
Radiotherapy only	44 (4.2)	105 (6.0)	20 (6.9)	84 (6.6)	1 (7.7)	2 (7.1)	256 (5.8)
Untreated	286 (27.1)	523 (30.1)	107 (37.0)	459 (36.1)	2 (15.4)	13 (46.4)	1,390 (31.6)
Concomitant radiotherapy	103 (9.7)	151 (8.7)	18 (6.2)	59 (4.6)	2 (15.4)	0	333 (7.6)
Diagnostic imaging, yes (%)							
Tomography	745 (70.5)	1,280 (73.7)	215 (74.4)	973 (76.6)	10 (76.9)	15 (53.6)	3,238 (73.7)
Radiography	509 (48.2)	875 (50.4)	154 (53.3)	726 (57.2)	7 (53.8)	21 (75.0)	2,292 (52.2)

Table 4.1. Cohort characteristics

SD: Standard deviation; **IQR**: interquartile range

Concerning gender, female patients were more represented in non-squamous than squamous histology (37.8% *vs.* 18.5%). The mean age of the overall study cohort was 68,9 years and the age band 70-84 years was the most represented for all histology and genders.

Overall, most part of the patients with NSCLC had received a first-line standard chemotherapy (52.6%). Approximately 10% of patients with non-squamous histology received target therapies, while these drugs were used only in few patients with squamous histology (n=5). Overall, 2.5% of patients in the cohort received first-line immunotherapy. A higher proportion of patients who did not receive any treatment within four months of cohort entry was observed in patients with squamous NSCLC than patients with non-squamous NSCLC: 36.3% (female 37.0%; male 36.1%) *vs.* 29.0% (female 27.1%; male 30.1%). Characteristics of patients in relation to first-line treatment are available in **Table 4.2**. Concomitant radiotherapy was observed in a total of 7.6% of cases (9.1% in non-squamous and 4.9% in squamous).

	Non-squamous					Squamous						
	Standard Chemotherapy	Target therapy	Immunotherapy	uMAB	Radiotherapy	Untreated	Standard Chemotherapy	Target therapy	Immunotherapy	uMAB	Radiotherapy	Untreated
n	1447	270	85	33	149	809	845	5	26	13	104	566
Female (%)	526 (36.4)	158 (58.5)	33 (38.8)	10 (30.3)	44 (29.5)	286 (35.4)	151 (17.9)	3 (60.0)	7 (26.9)	1 (7.7)	20 (19.2)	107 (18.9)
Year of Cohort Entry Date (%)												
2009	74 (5.1)	1 (0.4)	0 (0.0)	0 (0.0)	7 (4.7)	26 (3.2)	91 (10.8)	0 (0.0)	0 (0.0)	0 (0.0)	10 (9.6)	57 (10.1)
2010	89 (6.2)	1 (0.4)	0 (0.0)	3 (9.1)	12 (8.1)	28 (3.5)	95 (11.2)	2 (40.0)	0 (0.0)	1 (7.7)	13 (12.5)	55 (9.7)
2011	120 (8.3)	8 (3.0)	0 (0.0)	1 (3.0)	7 (4.7)	55 (6.8)	94 (11.1)	0 (0.0)	0 (0.0)	1 (7.7)	7 (6.7)	51 (9.0)
2012	95 (6.6)	13 (4.8)	0 (0.0)	0 (0.0)	4 (2.7)	54 (6.7)	98 (11.6)	0 (0.0)	0 (0.0)	1 (7.7)	12 (11.5)	72 (12.7)
2013	128 (8.8)	18 (6.7)	0 (0.0)	3 (9.1)	8 (5.4)	65 (8.0)	68 (8.0)	1 (20.0)	0 (0.0)	0 (0.0)	12 (11.5)	45 (8.0)
2014	137 (9.5)	28 (10.4)	0 (0.0)	2 (6.1)	11 (7.4)	79 (9.8)	63 (7.5)	0 (0.0)	0 (0.0)	0 (0.0)	7 (6.7)	48 (8.5)
2015	150 (10.4)	19 (7.0)	0 (0.0)	3 (9.1)	17 (11.4)	76 (9.4)	58 (6.9)	1 (20.0)	0 (0.0)	1 (7.7)	11 (10.6)	31 (5.5)
2016	181 (12.5)	32 (11.9)	0 (0.0)	5 (15.2)	20 (13.4)	103 (12.7)	70 (8.3)	1 (20.0)	0 (0.0)	3 (23.1)	12 (11.5)	63 (11.1)
2017	189 (13.1)	61 (22.6)	17 (20.0)	8 (24.2)	24 (16.1)	121 (15.0)	86 (10.2)	0 (0.0)	5 (19.2)	2 (15.4)	10 (9.6)	58 (10.2)
2018	168 (11.6)	66 (24.4)	45 (52.9)	6 (18.2)	33 (22.1)	123 (15.2)	92 (10.9)	0 (0.0)	12 (46.2)	2 (15.4)	8 (7.7)	64 (11.3)
2019	116 (8.0)	23 (8.5)	23 (27.1)	2 (6.1)	6 (4.0)	79 (9.8)	30 (3.6)	0 (0.0)	9 (34.6)	2 (15.4)	2 (1.9)	22 (3.9)
Age, mean (SD)	66.38 (9.83)	68.51 (10.94)	66.89 (10.31)	67.94 (8.94)	72.93 (9.14)	72.82 (10.42)	69.12 (8.76)	70.00 (5.52)	71.04 (8.77)	68.77 (8.86)	75.72 (7.68)	74.64 (8.93)
Age bands (%)												
18-54	183 (12.6)	32 (11.9)	12 (14.1)	2 (6.1)	8 (5.4)	43 (5.3)	57 (6.7)	0 (0.0)	0 (0.0)	1 (7.7)	1 (1.0)	25 (4.4)
55-69	676 (46.7)	101 (37.4)	35 (41.2)	15 (45.5)	38 (25.5)	230 (28.4)	344 (40.7)	3 (60.0)	11 (42.3)	5 (38.5)	20 (19.2)	112 (19.8)
70-84	573 (39.6)	127 (47.0)	36 (42.4)	16 (48.5)	92 (61.7)	446 (55.1)	434 (51.4)	2 (40.0)	12 (46.2)	7 (53.8)	75 (72.1)	365 (64.5)
85+	15 (1.0)	10 (3.7)	2 (2.4)	0 (0.0)	11 (7.4)	90 (11.1)	10 (1.2)	0 (0.0)	3 (11.5)	0 (0.0)	8 (7.7)	64 (11.3)
Time in the study cohort, median [IQR]	322.00 [154.00 - 615.50]	538.00 [255.25 - 793.50]	438.00 [157.00 - 575.00]	374.00 [147.00 - 635.00]	154.00 [58.00 - 609.00]	62.00 [26.00 - 294.00]	301.00 [157.00 - 496.00]	199.00 [105.00 - 645.00]	344.50 [115.25 - 527.50]	232.00 [171.00 - 453.00]	178.50 [69.00 - 694.25]	77.50 [32.00 - 275.50]
Tomography, yes (%)	1095 (75.7)	191 (70.7)	62 (72.9)	16 (48.5)	114 (76.5)	547 (67.6)	657 (77.8)	4 (80.0)	18 (69.2)	8 (61.5)	83 (79.8)	418 (73.9)
Rx chest, yes (%)	783 (54.1)	119 (44.1)	32 (37.6)	16 (48.5)	66 (44.3)	368 (45.5)	501 (59.3)	3 (60.0)	15 (57.7)	6 (46.2)	43 (41.3)	312 (55.1)

Table 4.2. Characteristics of patients on the basis of first-line received

uMAB: Unspecified monoclonal antibodies



The observed incidence of use of immunotherapy as first-line treatment in the study cohort varied between 3.8% in 2017 and 9.1% in 2019, without any substantial difference for histology or gender (**Figure 4.3**: data not shown in the figure). In patients with non-squamous NSCLC, first-line standard chemotherapy decreased from 2009 to 2019, both for women (from 61.7% to 41.1%) and for men (from 73.8% to 49.7%) (**Figure 4.3** - Panel A). Target therapy was used as first-line for non-squamous NSCLC more frequently in women than in men (14.9% *vs.* 6.5%). First-line immunotherapy for non-squamous NSCLC was observed in both sexes: from 5.4% in 2017 to 9.4% in 2019 for men and from 1.9% in 2017 to 8.9% in 2019 for women. In patients with squamous NSCLC, first-line standard chemotherapy also decreased from 2009 to 2019, both for women (60.9% to 50.0%) and for men (57.0% to 45.3%) (**Figure 4.3** - Panel B). The use of target therapy was negligible in both sexes, while an increase of the use of first-line immunotherapy for squamous NSCLC was observed in 2017 to 11.3% in 2019 for men and from 0.0% in 2017 to 25.0% in 2019 for women).

The time series analysis showed a significant downward trend of the proportion of male with nonsquamous and squamous NSCLC receiving standard chemotherapy (tau = -0.546, *p value* < 0.001; tau = -0.224, *p value* =0.040, respectively). Also female with non-squamous NSCLC receiving standard chemotherapy had a significant downward trend (tau = -0.495, *p value* <0.001) (**Figure 4.4**). The time series analysis for target and immuno-therapies was not performed due to the low incidence patients treated with those two categories.

A higher utilization of anti-EGFR therapies in women than men out of the total number of patients treated with target therapies (85.4% *vs.* 69.6%) was observed (**Figure 4.5**). Conversely, men receiving treatment with anti-VEGF and anti-ALK was higher than the percentage of women (19.6% *vs.* 7.0%, and 9.8% *vs.* 6.3%, respectively). Finally, only a small part of patients had received anti-BRAF drugs (approximately 1% for both sexes).



Percentage of untreated patients (grey) and of those receiving standard chemotherapy (yellow) during the study period per calendar quarters; Panel A1) Non-squamous NSCLC, Male; Panel A2) Non-squamous NSCLC, Female; Panel B1) Squamous NSCLC, Male; Panel B2) Squamous NSCLC, Female *Statistically significant



Survival

Concerning non-squamous NSCLC, among patients that received a pharmacological treatment, the median survival was 389 days in women, and 276 days in men.

The two-year survival showed a trend to an increase over time: in 2009-2011 was 30.4% for women and 17.7% for men, while it was 35.5% for women and 26.9% for men in 2018-2019. Also, the five-year survival showed a trend to an increase over time: in 2009-2011 it was 8.5% for women and 3.2% for men, while in 2015-2017 it was 14.7% for women and 8.3% for men (**Figure 4.6** - Panel A).

Concerning squamous NSCLC, among patients that received a pharmacological treatment, the median survival was 210 days in women and 274 days in men, without any meaningful trend over years.

The two-year survival showed a trend to an increase over time for men: in 2009-2011 was 12.3%, while it was 26.9% 2018-2019. Also, the five-year survival showed a trend to an increase over time in men: in 2009-2011 was 4.7% while it was 6.4% in 2015-2017 (**Figure 4.6** - Panel B). Among women with squamous NSCLC, the estimate of five-year survival in 2015-2017 (n=0) and of two-years survival in 2018-2019 (n=1) was not possible due to sample size.

Characteristics of treated patients by histology and gender are presented in **Table 4.3**. While for squamous patients' characteristics of men and women are comparable, for non-squamous ones men were slightly older than female (68 *vs.* 66 years).


	Non squamous		Squamous	
	Male	Female	Male	Female
n	1213	771	811	182
Year Cohort Entry Date (%)				
2009	48 (4.0)	34 (4.4)	86 (10.6)	15 (8.2)
2010	65 (5.4)	40 (5.2)	101 (12.5)	10 (5.5)
2011	96 (7.9)	40 (5.2)	85 (10.5)	17 (9.3)
2012	72 (5.9)	40 (5.2)	90 (11.1)	21 (11.5)
2013	91 (7.5)	66 (8.6)	66 (8.1)	15 (8.2)
2014	112 (9.2)	66 (8.6)	59 (7.3)	11 (6.0)
2015	124 (10.2)	65 (8.4)	58 (7.2)	13 (7.1)
2016	140 (11.5)	98 (12.7)	67 (8.3)	19 (10.4)
2017	184 (15.2)	115 (14.9)	80 (9.9)	23 (12.6)
2018	174 (14.3)	144 (18.7)	86 (10.6)	28 (15.4)
2019	107 (8.8)	63 (8.2)	33 (4.1)	10 (5.5)
Age, mean (SD)	68.07 (9.36)	65.85 (11.03)	70.00 (8.80)	69.26 (9.17)
Age bands (%)				
18-54	114 (9.4)	123 (16.0)	47 (5.8)	12 (6.6)
55-69	532 (43.9)	333 (43.2)	305 (37.6)	78 (42.9)
70-84	542 (44.7)	302 (39.2)	440 (54.3)	90 (49.5)
85+	25 (2.1)	13 (1.7)	19 (2.3)	2 (1.1)
Time in the study cohort, median [IQR]	298.00 [138.00 - 606.00]	399.00 [189.50 - 729.50]	307.00 [147.00 - 533.00]	244.50 [131.50 - 433.50]

Table 4.3. Characteristics of patients treated with a first-line by sex

SD: Standard deviation; **IQR**: interquartile range

Discussion and conclusions

With this study we described the first-line treatment utilization and survival in NSCLC patients from a whole Italian region between 2009 and 2019. During the study period we observed a trend of decrease of the percentage of patients treated with chemotherapy. The highest percentage of cases treated with first-line target therapy was observed among women with non-squamous NSCLC, in which anti-EGFR medications were the most frequently used target drugs. At the same time the survival of patients in the cohort seemed to improve over the study period, especially among women with non-squamous NSCLC. This study also add knowledge on the re-use of data from the pathology registry of the regional AHD of Tuscany that, to the best of our knowledge, was never used before for pharmacoepidemiology purposes. The use of pathology registry allowed to identify patients with NSCLC, also providing information on the specific cancer histology. The characteristics of the included patients are in line with epidemiological data:¹⁴ as expected, non-squamous histology was more frequent than squamous histology on the total of NSCLC cases (70% *vs.* 30%) and the proportion of women with non-squamous histology was higher than the proportion of those with squamous NSCLC.¹²⁹

Concerning drug utilization, this study found that among patients with non-squamous NSLCL women received a first-line target therapy more than twice as frequently as men (14.9% *vs.* 6.5%). These data are in line with the frequency of lung cancer mutations in the Caucasian population, where EGFR and BRAF mutations are known to be more frequent in women.^{130–132} During the study period, the proportion of patients who did not receive any drug treatment in the four months following cohort entry was stable and concerned about 1/3 of patients with NSCLC whatever the histology (36% squamous *vs.* 29% non-squamous). A previous study using the Iowa Cancer Registry reported that more than 20% of patients with advanced NSCLC were untreated and that such percentage was higher than that observed for any other

cancer, with exception of non-Hodgkin lymphoma.¹³³ Also, a National Cancer Data Base study showed comparable rates from 1998 to 2012: among patients with stage IIIB and IV NSCLC, 22.3% and 25.9% of patients did not received any anticancer treatment, respectively.¹³⁴ The high rate of untreated patients is mainly due to the advanced stage of the disease and the often poor performance status of patients for which the small benefits expected from the treatment and their potential harms can lead physicians and patients to avoid any anticancer treatment.¹³⁵ According to the guidelines for the treatment of NSCLC, identification of activating mutations of EGFR and ALK is recommended.¹⁴ According to literature, 10-13% of Caucasian patients with adenocarcinoma (which represents the most part of non-squamous NSCLC) presents an EGFR activating mutation, 2-7% rearrangements of ALK and 1-4% BRAF mutations.^{14,136} In line with these data, about 10% of patients with nonsquamous NSCLC in our study cohort were treated with target therapies. Moreover, as expected from the existing differences between sexes in the expression of EGFR activating mutation,^{130,131} women with non-squamous NSCLC in the present study cohort received a first-line target therapy more than twice as frequently as men with non-squamous NSLCL. During the study period, immunotherapy (e.g. pembrolizumab) was approved as first-line treatment for NSCLC by FDA and EMA. In non-squamous NSCLC, pembrolizumab is reserved to patients not eligible to target therapy, in combination with chemotherapy or as monotherapy, but only in patients with >50% of PD-L1 mutation. It could be conversely used in combination to standard chemotherapy in the squamous ones. Other observational studies that focused on the use of first-line pembrolizumab in NSCLC patients ^{137–139} did not analyze the incidence of use of pembrolizumab in patients with advanced NSCLC. Thus, to our knowledge, the present study estimated for the first time the incidence of use of first-line immunotherapy in non-resectable NSCLC cancer after approval in Italy and showed an increase in its use in both squamous and non-squamous histologies, and in both sexes.

Concerning survival, the present study seemed to show an overall trend of improvement of patient's prognosis over time, which may be in part related to the marketing of novel anticancer drugs. Overall, independently from NSCLC histology and with no specific difference in the characteristics of treated patients between male and female (**Table 4.3**), OS increased over time both in terms of two- or five-year survival. Howlader *et al.* reported that in a US cohort with NSCLC (irrespective of histology and stage) the estimated two-year survival increased from 2001 to 2016: from 26 to 35% in men and from 35 to 44% in women, respectively.¹⁴⁰ In the present study, among patients with non-squamous NSCLC that received any anticancer drugs, a better OS was observed in women compared to men; on the other hand, for squamous one, OS was conversely better in men than in women and these results are in line with other studies.^{141,142}

This study has several strengths. The first is represented by the linkage of different registries from the administrative healthcare database of Tuscany which provided a detailed information at patient-level, treatment information, and cancer histology. The second strength is represented by the use of an *ad hoc* case-finding algorithm developed and validated for the purpose of this study. The algorithm was shown to have an high PPV (93%).¹²⁸ Finally, to the best of our knowledge, this is the first population-based observational study that estimated the incidence of use of immunotherapy as fist-line treatment in patients with non-resectable NSCLC in the real-world clinical practice in Italy. The study has also some limitations. The first concerns the sensitivity of the algorithm used for the algorithm to the regional pathology registry was found to be able to identify about the half of the cases of NSCLC occurred in the Tuscan population.¹²⁸ However, the distribution of patient characteristics (i.e., histological types) in the present study cohort was line with the epidemiology of the disease, so that the sample of cases identified can be considered representative of the underlying

source case population. Secondly, the accuracy of the identification of the first-line pharmacotherapy was not validated. Given the universal healthcare assistance granted to all Italian inhabitants and the extremely high costs associated to NSCLC treatment, the possible impact of private healthcare assistance (which is not tracked in the administrative data used for this study) on exposure misclassification is expected to be negligible. Nevertheless, in some cases we cannot exclude that the date of actual administration of drugs might in some cases differ from the date registered in the administrative databases used for this study. Finally, two of the databanks used for the identification of drug exposure (HDR and OED), only allowed to identify "unspecified anti-body therapy" which could correspond either to an immunotherapy or a bevacizumab-based target therapy. Nevertheless, patients treated with drugs classified as "unspecified anti-body therapy" represents only 1.0% of the entire cohort.

In conclusion, this study demonstrated that Italian administrative healthcare data can be a valuable tool for the study of drug utilization and survival in oncology field. This study described, over last decade, the introduction of target and immunotherapies for the treatment of patients with non-resectable NSCLC in the clinical practice. Our findings highlighting a different use of target therapies between sexes: in women with non-squamous NSCLC, anti-EGFR medications were the most frequently used target drugs according to the known biomolecular profile of both sexes. Survival of patients in the cohort seemed to slightly improve over the study period, and male and female patients also presented a different survival profile if diagnosed with a squamous or a non-squamous NSCLC. Future studies on this topic will be able to leverage evidence and methodologies from this work to provide a more detailed picture of NSCLC pharmacotherapy utilization (i.e. second-line treatment), safety and effectiveness in real-life, with a special focus on innovative medications and gender.

CHAPTER 5. Thesis conclusion and perspectives

All the primary objectives of this thesis were investigated, and we provided results that could be the basis for further works. Here we discuss the results of the presented studies, taking into consideration future prospective of this thesis.

5.1 Review of the existing literature

The aim of this study was to provide a structured overview of the available studies using routinely collected electronical healthcare data to facilitate the design and benchmark of future works on this topic.

From this study we understood that - unless big electronical healthcare records are available a record linkage between different data types usually gives complementary information to correctly characterized oncologic patients. Administrative/claims data usually do not record clinical information, such as tumor stage, histology, or gene mutations, which are crucial for studying drug utilization patterns and outcomes in cancer patients. Despite medical/health records were the most frequently used source of information for assessing drug exposure, administrative/claims data can be used to track this information.

This study also highlighted a paucity of studies performed in Europe. This is probably because in regions other than North America, healthcare data are often scattered in different and heterogeneous databases, so that the performing studies on rare events that requires information from different healthcare settings, as in the case of lung cancer, remains a challenge. Few evidence also concerned immunotherapies, and in particular when they are used as first-line pharmacotherapy. Such paucity of literature is mostly due to the recent approval of these class of anticancer medications for such indication: pembrolizumab was the first immunotherapy approved for first-line treatment of advanced NSCLC in 2017.

Notably, no study reporting drug utilization evidence concerning SCLC patients was found. Given the very low prevalence of SCLC, rcEHD has the potential of playing an important role in capturing and studying far larger populations of SCLC patients than those recruited in clinical trials.

Results from this review represent a starting point for future studies on this topic, also highlighting current gaps of knowledge and providing summary of pertinent literature both for study design and for results benchmarking. We used the acquired knowledge from this work for the conduction of the validation and the drug utilization/survival study.

5.2 Validation study

The scoping review highlighted that cancer registries when linked to other rcEHD are a valuable data source for the conduction of pharmacoepidemiology studies in lung cancer. Even if the Tuscany region has a regional cancer registry, unfortunately data from coming this registry could not be linked to other datasources such as administrative data, making it impossible to use it for the purposes of the thesis. Anyway, the administrative database of the Tuscany region has different registries linked among each other's. Among them, a regional pathology registry that collects information about all the inhabitants of Tuscany was not already used for conducting observational studies. The aim of the study was thus to develop and to validate a case-finding algorithm for the identification of patients with NSCLC in a region-wide pathology registry in Italy.

The algorithm was found to have a high PPV (93%) and an acceptable SE (up to 65%) which estimates appeared to increase over the study period and could be considered in line with the share of regional NSCLC cases that are likely to be captured by the pathology registry.

We thus demonstrated the Tuscany pathology registry could be effectively used with other data to provide information on histology that can be used for research purposes.

Given the wide use of SNOMED terminology in electronic pathology reports, the proposed algorithm is expected to be easily adapted to other electronic databases. The promising evidence on validity demonstrated the final version of the NSCLC-finding algorithm designed and tested in this study can be considered suitable for epidemiology and pharmacoepidemiology studies. Other available pathology registries in other Italian regions should be investigated to understand the reproducibility of this algorithm, and its validity measures. This would allow the retrieval of a far larger number of patients with NSCLC to be included in a unique cohort, to perform further studies on drug utilization and survival in patients with NSCLC.

5.3 Study using the administrative database of the Tuscany region

5.3.1 Drug utilization study

While in the scoping review most part of the included studies investigated the drug utilization of innovative drugs in relation to histology, only few studies reported some of the presented result stratified by sex.

One of the aims of last study of this thesis was to describe drug utilization of first-line anticancer therapies approved for the treatment of advanced, non-resectable, NSCLC patients in relation to gender by analysing data from the regional administrative database of Tuscany.

We found that during the study period, the proportion of patients who did not receive any drug treatment in the four months following cohort entry was stable and concerned about 1/3 of patients with NSCLC whatever the histology. The high rate of untreated patients could mainly be due to the advanced stage of the disease and to the often too poor performance status of patients for which the small benefits expected from the treatment and their potential harms can lead physicians and patients to avoid any anticancer treatment.

In this study we found that, in line with epidemiological data, about 10% of patients with nonsquamous NSCLC in our study cohort were treated with target therapies. Moreover, as expected from the existing differences between sexes in the expression of EGFR activating mutation, women with non-squamous NSCLC in the present study cohort received a first-line target therapy more than twice as frequently as men with non-squamous NSCLC.

This study is to the best of our knowledge, the first study assessing the incidence of first-line pharmacotherapies in patients with non-resectable, primary NSCLC in Italy. This could provide important information for the scientific community for the benchmarking of the results, and for the health authorities to understand the pattern of use of these drugs in a real word setting which is further changing and evolving in NSCLC patients. Only in 2021 (at October), six novel target/immunotherapies are approved by FDA for the treatment of advanced NSCLC.¹⁴³ This study could be replicated over the years for the characterization of the users in treatment with pharmacological treatments and to discover the different use of these drugs in relation to important prognostic variables, such as gender. Of course, an implementation of the present study could be performed in order to understand also the switch to other pharmacological treatments and the pattern of the second- and further treatment line.

5.3.2 Survival

After drug utilization, the other aim of last study of this thesis was to describe the survival of first-line anticancer therapies approved for the treatment of advanced, non-resectable, NSCLC patients in relation to gender by analysing data from the regional administrative database of Tuscany.

Despite the five-years survival of these patients remained low, this work suggests that survival of patients with non-resectable NSCLC receiving a first-line of treatment was slightly improved in the last decade. This could be at least in part due to an added benefit of first-line treatments available for eligible patients with non-resectable NSCLC. This work also highlighted difference in the survival between men and women on the basis of NSCLC histology (squamous and non-squamous). The two histology of NSCLC, squamous and nonsquamous, differed on the basis of druggable mutations by target therapies. This in turn could explain the different survival between men and women on the basis of histologies, given that women had a different NSCLC biomolecular profile when compared to men.

This work provided an up-to-date knowledge about the possible impact of novel pharmacotherapies in the last decade in patients with NSCLC receiving at least a first-line pharmacological treatment. Anyway, given that the benefits of immunotherapy are more evident in the long-term period and that pembrolizumab was approved only in 2017 as a first-line, an update of this study in the next years could be useful to understand the evolution of survival of these patients. Another future prospective of this thesis could regard the evaluation of the incidence of the adverse effects of these drugs, which has not been considered yet. The understanding of the safety profile of these drugs along with their effectiveness in a real word setting could represent a great chance to be taken.

6. BIBLIOGRAPHY

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209-249. doi:10.3322/caac.21660

2. Vavalà T, Mariniello A, Reale ML, Novello S. Gender differences in lung cancer. *Italian Journal of Gender-Specific Medicine*. 2016;2(3):99-109.

3. Rossi S, Baili P, Capocaccia R, et al. The EUROCARE-5 study on cancer survival in Europe 1999-2007: Database, quality checks and statistical analysis methods. *Eur J Cancer*. 2015;51(15):2104-2119. doi:10.1016/j.ejca.2015.08.001

4. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. *Cancer Manag Res.* 2019;11:943-953. doi:10.2147/CMAR.S187317

5. Postmus PE, Kerr KM, Oudkerk M, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2017;28:iv1-iv21. doi:10.1093/annonc/mdx222

6. Musk AW, de Klerk N, Reid A, Hui J, Franklin P, Brims F. Asbestos-related diseases. *Int J Tuberc Lung Dis.* 2020;24(6):562-567. doi:10.5588/ijtld.19.0645

7. Hubbard R, Venn A, Lewis S, Britton J. Lung cancer and cryptogenic fibrosing alveolitis. A population-based cohort study. *Am J Respir Crit Care Med.* 2000;161(1):5-8. doi:10.1164/ajrccm.161.1.9906062

8. Lorigan P, Radford J, Howell A, Thatcher N. Lung cancer after treatment for Hodgkin's lymphoma: a systematic review. *Lancet Oncol.* 2005;6(10):773-779. doi:10.1016/S1470-2045(05)70387-9

9. Kirk GD, Merlo C, O' Driscoll P, et al. HIV infection is associated with an increased risk for lung cancer, independent of smoking. *Clin Infect Dis.* 2007;45(1):103-110.

doi:10.1086/518606

10. Yano T, Haro A, Shikada Y, Maruyama R, Maehara Y. Non-small cell lung cancer in never smokers as a representative "non-smoking-associated lung cancer": epidemiology and clinical features. *Int J Clin Oncol.* 2011;16(4):287-293. doi:10.1007/s10147-010-0160-8

11. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy. *Semin Cancer Biol.* 2018;52(Pt 1):103-109. doi:10.1016/j.semcancer.2017.11.019

Myers DJ, Wallen JM. Lung Adenocarcinoma. In: *StatPearls*. StatPearls Publishing;
 2021. Accessed August 4, 2021. http://www.ncbi.nlm.nih.gov/books/NBK519578/

13. Schwendenwein A, Megyesfalvi Z, Barany N, et al. Molecular profiles of small cell lung cancer subtypes: Therapeutic implications. *Molecular Therapy - Oncolytics*.
2021;20:470-483. doi:10.1016/j.omto.2021.02.004

14. AIOM - LINEE GUIDA NEOPLASIE DEL POLMONE - 2019. AIOM. Published October 25, 2019. Accessed July 7, 2020. https://www.aiom.it/linee-guida-aiom-neoplasie-del-polmone-2019/

15.CancerStaging.AccessedOctober21,2021.https://www.cancer.org/treatment/understanding-your-diagnosis/staging.html

16. Detterbeck FC. The eighth edition TNM stage classification for lung cancer: What does it mean on main street? *The Journal of Thoracic and Cardiovascular Surgery*. 2018;155(1):356-359. doi:10.1016/j.jtcvs.2017.08.138

17. Pirker R. Adjuvant Chemotherapy of Non-Small Cell Lung Cancer. *Tanaffos*.2012;11(1):12-17.

Drugs Approved for Lung Cancer - National Cancer Institute. Published April 5, 2011.
 Accessed October 21, 2021. https://www.cancer.gov/about-cancer/treatment/drugs/lung

19. Rossi A, Di Maio M. Platinum-based chemotherapy in advanced non-small-cell lung cancer: optimal number of treatment cycles. *Expert Rev Anticancer Ther*. 2016;16(6):653-660. doi:10.1586/14737140.2016.1170596

20. Treat J, Scagliotti GV, Peng G, Monberg MJ, Obasaju CK, Socinski MA. Comparison of pemetrexed plus cisplatin with other first-line doublets in advanced non-small cell lung cancer (NSCLC): a combined analysis of three phase 3 trials. *Lung Cancer*. 2012;76(2):222-227. doi:10.1016/j.lungcan.2011.10.021

21. Falzone L, Salomone S, Libra M. Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. *Front Pharmacol.* 2018;9:1300. doi:10.3389/fphar.2018.01300

22. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular Cancer*. 2018;17(1):58. doi:10.1186/s12943-018-0782-4

23. Guo Y, Cao R, Zhang X, et al. Recent Progress in Rare Oncogenic Drivers and Targeted Therapy For Non-Small Cell Lung Cancer. *Onco Targets Ther*. 2019;12:10343-10360. doi:10.2147/OTT.S230309

24. Hsu P-C, Jablons DM, Yang C-T, You L. Epidermal Growth Factor Receptor (EGFR) Pathway, Yes-Associated Protein (YAP) and the Regulation of Programmed Death-Ligand 1 (PD-L1) in Non-Small Cell Lung Cancer (NSCLC). *International Journal of Molecular Sciences*. 2019;20(15):3821. doi:10.3390/ijms20153821

25. Du X, Shao Y, Qin H-F, Tai Y-H, Gao H-J. ALK-rearrangement in non-small-cell lung cancer (NSCLC). *Thorac Cancer*. 2018;9(4):423-430. doi:10.1111/1759-7714.12613

26. Mustachio LM, Roszik J. Current Targeted Therapies for the Fight against Non-Small Cell Lung Cancer. *Pharmaceuticals (Basel)*. 2020;13(11):374. doi:10.3390/ph13110374

27. Remon J, Pignataro D, Novello S, Passiglia F. Current treatment and future challenges in ROS1- and ALK-rearranged advanced non-small cell lung cancer. *Cancer Treat Rev.*

2021;95:102178. doi:10.1016/j.ctrv.2021.102178

28. O'Leary CG, Andelkovic V, Ladwa R, et al. Targeting BRAF mutations in non-small
cell lung cancer. *Transl Lung Cancer Res.* 2019;8(6):1119-1124.
doi:10.21037/tlcr.2019.10.22

29. Khunger A, Khunger M, Velcheti V. Dabrafenib in combination with trametinib in the treatment of patients with BRAF V600-positive advanced or metastatic non-small cell lung cancer: clinical evidence and experience. *Ther Adv Respir Dis.* 2018;12:1753466618767611. doi:10.1177/1753466618767611

30. Bačić I, Karlo R, Zadro AŠ, Zadro Z, Skitarelić N, Antabak A. Tumor angiogenesis as an important prognostic factor in advanced non-small cell lung cancer (Stage IIIA). *Oncology Letters*. 2018;15(2):2335-2339. doi:10.3892/ol.2017.7576

31. Sitohy B, Nagy JA, Dvorak HF. Anti-VEGF/VEGFR Therapy for Cancer: Reassessing the Target. *Cancer Res.* 2012;72(8):1909-1914. doi:10.1158/0008-5472.CAN-11-3406

32. Manzo A, Montanino A, Carillio G, et al. Angiogenesis Inhibitors in NSCLC. *Int J Mol Sci.* 2017;18(10):E2021. doi:10.3390/ijms18102021

33. Janning M, Loges S. Anti-Angiogenics: Their Value in Lung Cancer Therapy. *Oncol Res Treat*. 2018;41(4):172-180. doi:10.1159/000488119

34. Esfahani K, Roudaia L, Buhlaiga N, Del Rincon SV, Papneja N, Miller WH. A review of cancer immunotherapy: from the past, to the present, to the future. *Curr Oncol*. 2020;27(Suppl 2):S87-S97. doi:10.3747/co.27.5223

35. Koury J, Lucero M, Cato C, et al. Immunotherapies: Exploiting the Immune System for Cancer Treatment. *J Immunol Res.* 2018;2018:9585614. doi:10.1155/2018/9585614

36. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol.* 2008;26:677-704.
 doi:10.1146/annurev.immunol.26.021607.090331

37. Good-Jacobson KL, Szumilas CG, Chen L, Sharpe AH, Tomayko MM, Shlomchik MJ. PD-1 regulates germinal center B cell survival and the formation and affinity of long-lived plasma cells. *Nat Immunol.* 2010;11(6):535-542. doi:10.1038/ni.1877

38. Lamberti G, Sisi M, Andrini E, et al. The Mechanisms of PD-L1 Regulation in Non-Small-Cell Lung Cancer (NSCLC): Which Are the Involved Players? *Cancers (Basel)*. 2020;12(11):3129. doi:10.3390/cancers12113129

39. Schmidt LH, Kümmel A, Görlich D, et al. PD-1 and PD-L1 Expression in NSCLC Indicate a Favorable Prognosis in Defined Subgroups. *PLoS One*. 2015;10(8):e0136023. doi:10.1371/journal.pone.0136023

40. Juneja VR, McGuire KA, Manguso RT, et al. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med*. 2017;214(4):895-904. doi:10.1084/jem.20160801

41. Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol.* 2017;7(9):170070. doi:10.1098/rsob.170070

42. Liang Y, Wakelee HA. Adjuvant chemotherapy of completely resected early stage non-small cell lung cancer (NSCLC). *Transl Lung Cancer Res.* 2013;2(5):403-410. doi:10.3978/j.issn.2218-6751.2013.07.01

43. Chaft JE, Rimner A, Weder W, Azzoli CG, Kris MG, Cascone T. Evolution of systemic therapy for stages I–III non-metastatic non-small-cell lung cancer. *Nat Rev Clin Oncol.* Published online April 28, 2021:1-11. doi:10.1038/s41571-021-00501-4

44. Clinical Practice Living Guidelines – Metastatic Non-Small-Cell Lung Cancer | ESMO. Accessed October 19, 2021. https://www.esmo.org/guidelines/lung-and-chest-tumours/clinical-practice-living-guidelines-metastatic-non-small-cell-lung-cancer

45. Rose-James A, TT S. Molecular Markers with Predictive and Prognostic Relevance in

Lung Cancer. Lung Cancer Int. 2012;2012:729532. doi:10.1155/2012/729532

46. Lindeman NI, Cagle PT, Aisner DL, et al. Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. *Arch Pathol Lab Med*. 2018;142(3):321-346. doi:10.5858/arpa.2017-0388-CP

47. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med*. 2017;377(25):2500-2501. doi:10.1056/NEJMc1713444

48. Taniguchi H, Sen T, Rudin CM. Targeted Therapies and Biomarkers in Small Cell Lung Cancer. *Front Oncol.* 2020;10:741. doi:10.3389/fonc.2020.00741

49. RESERVED IU-AR. Orphanet: Search for an orphan drug. Accessed August 6, 2021. https://www.orpha.net/consor/cgi-

bin/Drugs_Search_Simple.php?lng=EN&LnkId=10953&Typ=Pat&fdp=y&from=rightMenu

50. FDA Approves Opdivo (nivolumab) for Certain Patients with Previously Treated Small Cell Lung Cancer. Drugs.com. Accessed November 12, 2019. https://www.drugs.com/newdrugs/fda-approves-opdivo-nivolumab-certain-patientspreviously-treated-small-cell-lung-cancer-4803.html

51. Montastruc J-L, Benevent J, Montastruc F, et al. What is pharmacoepidemiology? Definition, methods, interest and clinical applications. *Therapie*. 2019;74(2):169-174. doi:10.1016/j.therap.2018.08.001

52. 2019-09 PD-1/L1 trial landscape dashboard. Tableau Software. Accessed October 21, 2021.https://public.tableau.com/shared/F7WBQTSSC?:display_count=y&:origin=viz_share_l ink&:embed=y&:showVizHome=n&:tabs=n&:toolbar=n&:apiID=host0#navType=0&navSrc =Parse

53. Al-Refaie WB, Vickers SM, Zhong W, Parsons H, Rothenberger D, Habermann EB.

Cancer trials versus the real world in the United States. *Ann Surg.* 2011;254(3):438-442; discussion 442-443. doi:10.1097/SLA.0b013e31822a7047

54. Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. *J Oncol Pract*. 2012;8(6):365-370. doi:10.1200/JOP.2012.000646

55. Kwiatkowski K, Coe K, Bailar JC, Swanson GM. Inclusion of minorities and women in cancer clinical trials, a decade later: Have we improved? *Cancer*. 2013;119(16):2956-2963. doi:10.1002/cncr.28168

56. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*. 2013;347:f6409. doi:10.1136/bmj.f6409

57. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. *BMJ*. 2017;359:j4530. doi:10.1136/bmj.j4530

58. Herk-Sukel MPP van, Lemmens VEPP, Poll-Franse LV van de, Herings RMC, Coebergh JWW. Record linkage for pharmacoepidemiological studies in cancer patients. *Pharmacoepidemiology and Drug Safety*. 2012;21(1):94-103. doi:10.1002/pds.2205

59. Salcher-Konrad M, Naci H, Davis C. Approval of Cancer Drugs With Uncertain Therapeutic Value: A Comparison of Regulatory Decisions in Europe and the United States. *The Milbank Quarterly*. 2020;98(4):1219-1256. doi:10.1111/1468-0009.12476

60. Yakerson A. Women in clinical trials: a review of policy development and health equity in the Canadian context. *Int J Equity Health*. 2019;18:56. doi:10.1186/s12939-019-0954-x

61. Conte C, Vaysse C, Bosco P, et al. The value of a health insurance database to conduct

pharmacoepidemiological studies in oncology. *Therapies*. 2019;74(2):279-288. doi:10.1016/j.therap.2018.09.076

62. Hiatt RA, Tai CG, Blayney DW, et al. Leveraging state cancer registries to measure and improve the quality of cancer care: a potential strategy for California and beyond. *J Natl Cancer Inst.* 2015;107(5). doi:10.1093/jnci/djv047

63. Rivera DR, McGlynn KA, Freedman AN. Connections between pharmacoepidemiology and cancer biology: designing biologically relevant studies of cancer outcomes. *Ann Epidemiol*. 2016;26(11):741-745. doi:10.1016/j.annepidem.2016.10.001

64. EHRs, Meaningful Use, and a Model EMR. Patient Care Online. Accessed October 21, 2021. https://www.patientcareonline.com/view/ehrs-meaningful-use-and-model-emr

65. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity*. 2020;124(4):525-534. doi:10.1038/s41437-020-0303-2

66. Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Sci Rep*. 2017;7(1):46226. doi:10.1038/srep46226

67. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol*. 2010;2:51-56.

68. 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

69. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. :35.

70. Spini A, Hyeraci G, Bartolini C, et al. Real-World Utilization of Target- and

Immunotherapies for Lung Cancer: A Scoping Review of Studies Based on Routinely Collected Electronic Healthcare Data. *Int J Environ Res Public Health*. 2021;18(14):7679. doi:10.3390/ijerph18147679

71. Cote RA. Architecture of SNOMED: Its Contribution to Medical Language
Processing. *Proc Annu Symp Comput Appl Med Care*. Published online October 26, 1986:7480.

72. Doe J. SNOMED. EUROPEAN INNOVATION PARTNERSHIP - European Commission. Published August 26, 2016. Accessed August 5, 2021. https://ec.europa.eu/eip/ageing/standards/healthcare/e-health/snomed_en

73. Semantic Strategy in EU and SNOMED CT. Presented at the: https://confluence.ihtsdotools.org/download/attachments/73368385/SnomedCtShowcase2014 Present 14100.pdf?version=1&modificationDate=1535032086000&api=v2

74. Assessing SNOMED CT for Large Scale eHealth Deployments in the EU | ASSESS CT Project | H2020 | CORDIS | European Commission. Accessed August 5, 2021. https://cordis.europa.eu/project/id/643818/it

75. AIRC - tumore al polmone. Accessed June 15, 2020. https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-al-polmone

76. AIOM. Linee guida Associazione Italiana di Oncologia Medica - 2018: Tumore al polmone. https://www.aiom.it/wp-content/uploads/2018/11/2018_LG_AIOM_Polmone.pdf

77. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res.* 2016;5(3):288-300. doi:10.21037/tlcr.2016.06.07

78. Fahey T. Applying the results of clinical trials to patients to general practice: perceived problems, strengths, assumptions, and challenges for the future. *Br J Gen Pract*. 1998;48(429):1173-1178.

79. Unger JM, Cook E, Tai E, Bleyer A. Role of Clinical Trial Participation in Cancer

Research: Barriers, Evidence, and Strategies. *Am Soc Clin Oncol Educ Book*. 2016;35:185-198. doi:10.14694/EDBK_156686

80. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. THE ROLE OF INFORMATICS IN PROMOTING PATIENT-CENTERED CARE. *Cancer J*. 2011;17(4):211-218. doi:10.1097/PPO.0b013e318225ff89

81. Houser SH, Colquitt S, Clements K, Hart-Hester S. The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama. *Perspect Health Inf Manag*.
2012;9(Spring). Accessed November 29, 2018. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3329210/

82. Gini R, Dodd CN, Bollaerts K, et al. Quantifying outcome misclassification in multidatabase studies: The case study of pertussis in the ADVANCE project. *Vaccine*. Published online October 31, 2019. doi:10.1016/j.vaccine.2019.07.045

83. Bollaerts K, Rekkas A, Smedt TD, Dodd C, Andrews N, Gini R. Disease misclassification in electronic healthcare database studies: Deriving validity indices—A contribution from the ADVANCE project. *PLOS ONE*. 2020;15(4):e0231333. doi:10.1371/journal.pone.0231333

84. Tumori in Toscana 2014. :26.

85. Dati disponibili | Istituto per lo Studio e la Prevenzione Oncologia ISPO. Accessed October 26, 2020. http://www.ispro.toscana.it/rtt/stato_rilevazione

86. The value of SNOMED CT. SNOMED. Accessed June 10, 2021. https://www.snomed.org/snomed-ct/why-snomed-ct

87. Hall GC, Lanes S, Bollaerts K, Zhou X, Ferreira G, Gini R. Outcome misclassification: Impact, usual practice in pharmacoepidemiology database studies and an online aid to correct biased estimates of risk ratio or cumulative incidence. *Pharmacoepidemiology and Drug Safety*. n/a(n/a). doi:10.1002/pds.5109

88. I numeri del cancro in Italia. AIOM. Accessed October 26, 2020. https://www.aiom.it/i-numeri-del-cancro-in-italia/

89. Levra MG, Cotté F-E, Corre R, et al. Immunotherapy rechallenge after nivolumab treatment in advanced non-small cell lung cancer in the real-world setting: A national data base analysis. *Lung Cancer*. 2020;140:99-106. doi:10.1016/j.lungcan.2019.12.017

90. Schwartzberg L, Korytowsky B, Penrod JR, et al. Real-World Clinical Impact of Immune Checkpoint Inhibitors in Patients With Advanced/Metastatic Non–Small Cell Lung Cancer After Platinum Chemotherapy. *Clinical Lung Cancer*. 2019;20(4):287-296.e4. doi:10.1016/j.cllc.2019.04.004

91. Di Maio M, Perrone F, Conte P. Real-World Evidence in Oncology: Opportunities and Limitations. *Oncologist*. 2020;25(5):e746-e752. doi:10.1634/theoncologist.2019-0647

92. Marino M, Masella R, Bulzomi P, Campesi I, Malorni W, Franconi F. Nutrition and human health from a sex-gender perspective. *Mol Aspects Med.* 2011;32(1):1-70. doi:10.1016/j.mam.2011.02.001

93. Principles of Gender-Specific Medicine - 2nd Edition. Accessed August 6, 2021.
https://www.elsevier.com/books/principles-of-gender-specific-medicine/legato/978-0-12374271-1

94. Franconi F, Campesi I. Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women. *Br J Pharmacol*. 2014;171(3):580-594. doi:10.1111/bph.12362

95. Franconi F, Campesi I. Sex and gender influences on pharmacological response: an overview. *Expert Rev Clin Pharmacol.* 2014;7(4):469-485. doi:10.1586/17512433.2014.922866

96. Franconi F, Rosano G, Campesi I. Need for gender-specific pre-analytical testing: the dark side of the moon in laboratory testing. *Int J Cardiol.* 2015;179:514-535.

doi:10.1016/j.ijcard.2014.11.019

97. Assessing the economic impact of chronic conditions in postmenopausal women. Accessed August 6, 2021. https://www.analysisgroup.com/Insights/publishing/assessing-theeconomic-impact-of-chronic-conditions-in-postmenopausal-women/

98. Gulbins H, Vogel B, Reichenspurner H. Gender effects on health care costs in cardiovascular medicine-a black box? *Thorac Cardiovasc Surg.* 2013;61(1):74-78. doi:10.1055/s-0032-1328931

99. Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. *J Manag Care Pharm*. 2008;14(3 Suppl):2-6. doi:10.18553/jmcp.2008.14.S6-A.2

100. Rubtsova K, Marrack P, Rubtsov AV. Sexual dimorphism in autoimmunity. *J Clin Invest.* 2015;125(6):2187-2193. doi:10.1172/JCI78082

101. Yuan Y, Liu L, Chen H, et al. Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell*. 2016;29(5):711-722. doi:10.1016/j.ccell.2016.04.001

102. Kligerman S, White C. Epidemiology of lung cancer in women: risk factors, survival, and screening. *AJR Am J Roentgenol*. 2011;196(2):287-295. doi:10.2214/AJR.10.5412

103. Bain C, Feskanich D, Speizer FE, et al. Lung cancer rates in men and women with comparable histories of smoking. *J Natl Cancer Inst.* 2004;96(11):826-834. doi:10.1093/jnci/djh143

104. Yang SH, Mechanic LE, Yang P, et al. Mutations in the tyrosine kinase domain of the epidermal growth factor receptor in non-small cell lung cancer. *Clin Cancer Res.* 2005;11(6):2106-2110. doi:10.1158/1078-0432.CCR-04-1853

105. Dogan S, Shen R, Ang DC, et al. Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related

KRAS-mutant cancers. *Clin Cancer Res.* 2012;18(22):6169-6177. doi:10.1158/1078-0432.CCR-11-3265

106. Sullivan I, Planchard D. ALK inhibitors in non-small cell lung cancer: the latest evidence and developments. *Ther Adv Med Oncol.* 2016;8(1):32-47. doi:10.1177/1758834015617355

107. Wang Y, Wang S, Xu S, Qu J, Liu B. Clinicopathologic features of patients with nonsmall cell lung cancer harboring the EML4-ALK fusion gene: a meta-analysis. *PLoS One*. 2014;9(10):e110617. doi:10.1371/journal.pone.0110617

108. Ferguson MK, Wang J, Hoffman PC, et al. Sex-associated differences in survival of patients undergoing resection for lung cancer. *Ann Thorac Surg.* 2000;69(1):245-249; discussion 249-250. doi:10.1016/s0003-4975(99)01078-4

109. Alexiou C, Onyeaka CVP, Beggs D, et al. Do women live longer following lung resection for carcinoma? *Eur J Cardiothorac Surg*. 2002;21(2):319-325. doi:10.1016/s1010-7940(01)01114-9

110. Minami H, Yoshimura M, Miyamoto Y, Matsuoka H, Tsubota N. Lung cancer in women: sex-associated differences in survival of patients undergoing resection for lung cancer. *Chest.* 2000;118(6):1603-1609. doi:10.1378/chest.118.6.1603

111. de Perrot M, Licker M, Bouchardy C, Usel M, Robert J, Spiliopoulos A. Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. *J Thorac Cardiovasc Surg*. 2000;119(1):21-26. doi:10.1016/s0022-5223(00)70213-3

112. Fu JB, Kau TY, Severson RK, Kalemkerian GP. Lung cancer in women: analysis of the national Surveillance, Epidemiology, and End Results database. *Chest*. 2005;127(3):768-777. doi:10.1378/chest.127.3.768

113. Nakamura H, Ando K, Shinmyo T, et al. Female gender is an independent prognostic

factor in non-small-cell lung cancer: a meta-analysis. *Ann Thorac Cardiovasc Surg*. 2011;17(5):469-480. doi:10.5761/atcs.oa.10.01637

114. Visbal AL, Williams BA, Nichols FC, et al. Gender differences in non-small-cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997 and 2002. *Ann Thorac Surg.* 2004;78(1):209-215; discussion 215. doi:10.1016/j.athoracsur.2003.11.021

115. Paesmans M, Sculier JP, Lecomte J, et al. Prognostic factors for patients with small cell lung carcinoma: analysis of a series of 763 patients included in 4 consecutive prospective trials with a minimum follow-up of 5 years. *Cancer*. 2000;89(3):523-533. doi:10.1002/1097-0142(20000801)89:3<523::aid-cncr7>3.0.co;2-6

116. Spiegelman D, Maurer LH, Ware JH, et al. Prognostic factors in small-cell carcinoma of the lung: an analysis of 1,521 patients. *J Clin Oncol*. 1989;7(3):344-354. doi:10.1200/JCO.1989.7.3.344

117. Scagliotti GV, Parikh P, von Pawel J, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol*. 2008;26(21):3543-3551. doi:10.1200/JCO.2007.15.0375

118. Wakelee HA, Chang ET, Gomez SL, et al. Lung cancer incidence in never smokers. *J Clin Oncol.* 2007;25(5):472-478. doi:10.1200/JCO.2006.07.2983

119. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. *New England Journal of Medicine*.
2015;373(17):1627-1639. doi:10.1056/NEJMoa1507643

120. Herbst RS, Baas P, Kim D-W, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387(10027):1540-1550. doi:10.1016/S0140-6736(15)01281-7

121. Pinto JA, Vallejos CS, Raez LE, et al. Gender and outcomes in non-small cell lung cancer: an old prognostic variable comes back for targeted therapy and immunotherapy? *ESMO Open.* 2018;3(3). doi:10.1136/esmoopen-2018-000344

122. World Health Organization. PRGlobocanFinal.pdf. Accessed May 22, 2019. https://www.who.int/cancer/PRGlobocanFinal.pdf

123. Donington JS, Colson YL. Sex and Gender Differences in Non-Small Cell Lung Cancer. *Seminars in Thoracic and Cardiovascular Surgery*. 2011;23(2):137-145. doi:10.1053/j.semtcvs.2011.07.001

124. Wagner AD, Oertelt-Prigione S, Adjei A, et al. Gender medicine and oncology: report and consensus of an ESMO workshop. *Annals of Oncology*. 2019;30(12):1914-1924. doi:10.1093/annonc/mdz414

125. Schuemie MJ, Gini R, Coloma PM, et al. Replication of the OMOP experiment in Europe: evaluating methods for risk identification in electronic health record databases. *Drug Saf.* 2013;36 Suppl 1:S159-169. doi:10.1007/s40264-013-0109-8

126. Giometto S, Baglietto L, Conte M, et al. Use of antiseizure medications and safety of branded versus generic formulations: A comparative study on Tuscan administrative databases. *Epilepsy Behav.* 2021;117:107876. doi:10.1016/j.yebeh.2021.107876

127. Roberto G, Spini A, Bartolini C, et al. Real word evidence on rituximab utilization: Combining administrative and hospital-pharmacy data. *PLOS ONE*. 2020;15(3):e0229973. doi:10.1371/journal.pone.0229973

128. ICPE All Access conference abstracts. *Pharmacoepidemiology and Drug Safety*.2020;29(S3):3-634. doi:10.1002/pds.5114

129. Caldarella A, Crocetti E, Comin CE, Janni A, Pegna AL, Paci E. Gender differences in non-small cell lung cancer: A population-based study. *European Journal of Surgical Oncology*. 2007;33(6):763-768. doi:10.1016/j.ejso.2007.01.001

130. Forsythe ML, Alwithenani A, Bethune D, et al. Molecular profiling of non-small cell lung cancer. *PLoS One*. 2020;15(8). doi:10.1371/journal.pone.0236580

131. Bell DaphneW, Brannigan BrianW, Matsuo K, et al. Increased Prevalence of EGFR-Mutant Lung Cancer in Women and in East Asian Populations: Analysis of Estrogen-Related Polymorphisms. *Clin Cancer Res.* 2008;14(13):4079-4084. doi:10.1158/1078-0432.CCR-07-5030

132. Barrera-Rodriguez R, Morales-Fuentes J. Lung cancer in women. *Lung Cancer* (*Auckl*). 2012;3:79-89. doi:10.2147/LCTT.S37319

133. Ward MM, Ullrich F, Matthews K, et al. Who Does Not Receive Treatment for Cancer? *J Oncol Pract*. 2013;9(1):20-26. doi:10.1200/JOP.2012.000829

134. David EA, Daly ME, Li C-S, et al. Increasing Rates of No Treatment in Advanced-Stage Non–Small Cell Lung Cancer Patients: A Propensity-Matched Analysis. *Journal of Thoracic Oncology*. 2017;12(3):437-445. doi:10.1016/j.jtho.2016.11.2221

135. Frenkel M. Refusing Treatment. Oncologist. 2013;18(5):634-636.doi:10.1634/theoncologist.2012-0436

136. Roviello G, D'Angelo A, Sirico M, Pittacolo M, Conter FU, Sobhani N. Advances in anti-BRAF therapies for lung cancer. *Invest New Drugs*. 2021;39(3):879-890. doi:10.1007/s10637-021-01068-8

137. Facchinetti F, Mazzaschi G, Barbieri F, et al. First-line pembrolizumab in advanced non-small cell lung cancer patients with poor performance status. *Eur J Cancer*. 2020;130:155-167. doi:10.1016/j.ejca.2020.02.023

138. Amrane K, Geier M, Corre R, et al. First-line pembrolizumab for non–small cell lung cancer patients with PD-L1 \geq 50% in a multicenter real-life cohort: The PEMBREIZH study. *Cancer Med.* 2020;9(7):2309-2316. doi:10.1002/cam4.2806

139. Velcheti V, Hu X, Piperdi B, Burke T. Real-world outcomes of first-line

pembrolizumab plus pemetrexed-carboplatin for metastatic nonsquamous NSCLC at US oncology practices. *Sci Rep.* 2021;11(1):9222. doi:10.1038/s41598-021-88453-8

140. Howlader N, Forjaz G, Mooradian MJ, et al. The Effect of Advances in Lung-Cancer Treatment on Population Mortality. *New England Journal of Medicine*. Published online August 12, 2020. doi:10.1056/NEJMoa1916623

141. Sagerup CMT, Småstuen M, Johannesen TB, Helland Å, Brustugun OT. Sex-specific trends in lung cancer incidence and survival: a population study of 40 118 cases. *Thorax*. 2011;66(4):301-307. doi:10.1136/thx.2010.151621

142. Radkiewicz C, Dickman PW, Johansson ALV, Wagenius G, Edgren G, Lambe M. Sex and survival in non-small cell lung cancer: A nationwide cohort study. *PLOS ONE*. 2019;14(6):e0219206. doi:10.1371/journal.pone.0219206

143. FDA approvals in lung cancer treatment. Lung Cancer Research Foundation. Accessed October 20, 2021. https://www.lungcancerresearchfoundation.org/research/whyresearch/treatment-advances/

7. PUBLICATIONS

7.1 Publications and congress communications related to thesis

Publications

Spini A, Hyeraci G, Bartolini C, Donnini S, Rosellini P, Gini R, Ziche M, Salvo F, Roberto G. Real-World Utilization of Target- and Immunotherapies for Lung Cancer: A Scoping Review of Studies Based on Routinely Collected Electronic Healthcare Data. *Int J Environ Res Public Health*. 2021 Jul 19;18(14):7679.

Spini A, Rosellini P, Bellan C, Furiesi F, Giorgi S, Donnini S, Gini R, Ziche M, Salvo F, Roberto G. Development and validation of a case-finding algorithm for the identification of non-small cell lung cancers in a region-wide Italian pathology registry. *Submitted July 2021*.

Spini A, Gini R, Rosellini P, Singier A, Bellan C, Pascucci A, Leoncini L, Mathieu C, Martellucci I, Furiesi F, Giorgi S, Donnini S, Roberto G, Ziche M, Salvo F. First-line pharmacotherapies and survival among patients diagnosed with non-resectable NSCLC: a real-life setting study with gender prospective. *Submitted October 2021*.

Oral communications

Spini A, Salvo F, Roberto G, Ciccone V, Pascucci A, Rosellini P, Francini F, Donnini S, Gini R, Ziche M. Drug-utilization pattern of Non-Small Cell Lung Cancer patients related to gender at the University Hospital of Siena, Italy *The 9th Congress of the International Society of Gender Medicine* – Vienna 12-13 September 2019.

Spini A, Roberto G, Bellan C, Pascucci A, Rosellini P, Caldarella A, Barchielli A, Francini F, Donnini S, Furiesi F, Leoncini L, Giorgi S, Bartolini C, Zappa M, Ziche M, Salvo F, Gini R. Linking regional administrative healthcare data with the pathology registry of the University

Hospital of Siena (Italy) to describe treatment patterns of patients with non- small-cell lung cancer. 35° International Society of Pharmacoepidemiology – Szeged 4-7 March 2020.

Posters

Spini A, Roberto G, Bellan C, Pascucci A, Rosellini P, Caldarella A, Barchielli A, Francini F, DonniniS, Furiesi F, Leoncini L, Giorgi S, Bartolini C, Zappa M, Ziche M, Gini R. Linking regional administrative healthcare data with the pathology registry of the University Hospital of Siena (Italy) to describe treatment patterns of patients with non-small- cell lung cancer. *39° National congress Società Italiana di Farmacologia* - Florence 19-22 September 2019.

Spini A, Hyeraci G, Pellicano F, Donnini S, Giorgi S, Bartolini C, Gini R, Ziche M, Salvo F, Roberto G. Re-use of routinely collected electronic healthcare data for the execution of real-world studies on the utilization of target and immunotherapies in lung cancer patients: a systematic review. *39° National congress Società Italiana di Farmacologia* - Florence 19-22 settembre 2019.

Rosellini P, **Spini A**, Roberto G, Gini R, Ciccone V, Francini E, Pascucci A, Donnini S, Ziche M Drug-utilization pattern of Non-Small Cell Lung Cancer patients analyzed by gender at University Hospital of Siena, Italy (UHS). *39° National congress Società Italiana di Farmacologia* - Florence 19-22 September 2019.

Spini A, Rosellini P, Bellan C, Furiesi F, Giorgi S, Donnini S, Gini R, Ziche M, Salvo F, Roberto G. Identification of cases of NSCLC in the pathology registries of the University Hospital of Siena, Italy: a validation study. *ICPE all access Meeting 35 years of real word science*; September 2020.

Spini A, Roberto G, Bellan C, Pascucci A, Rosellini P, Donnini S, Furiesi F, Leoncini L, Giorgi S, Singer A, Bartolini C, Ziche M, Gini R, Salvo F. Using administrative data and the regional pathology registry of Tuscany to describe first-line pharmacotherapy and survival of patients with

non-resectable Non-Small Cell Lung cancer from 2009 to 2019. *ICPE all access Meeting 35 years of real word science*; August 2021.

Reports for the Tuscany regional healthcare agency

Spini A, Roberto G, Bartolini C, Gini R, Donnini S, Bellan C, Furiesi F, Leoncini L, Giorgi S, Rosellini P, Pascucci A, Ziche M, Barchielli A, Caldarella A, Zappa M. Nuovi trattamenti per il tumore polmonare non a piccole cellule: esperienza nell'interpretazione dei dati dei pazienti diagnosticati nell'ospedale di Siena. *Rapporto farmaci della regione toscana 2018* – December 2018.

Spini A, Roberto G, Bartolini C, Gini R, Donnini S, Bellan C, Furiesi F, Leoncini L, Giorgi S, Rosellini P, Pascucci A, Ziche M, Barchielli A, Caldarella A, Zappa M Salvo F. Farmaci Oncologici *Rapporto farmaci della regione toscana 2019* – December 2019.

Spini A, Roberto G, Bartolini C, Gini R, Donnini S, Bellan C, Furiesi F, Leoncini L, Giorgi S, Rosellini P, Pascucci A, Ziche M, Barchielli A, Caldarella A, Zappa M Salvo F. Differenze di genere nella farmacoterapia di prima linea e nella sopravvivenza di pazienti non operati con tumore polmonare non a piccole cellule in Toscana. Farmaci Oncologici. *Rapporto farmaci della regione toscana 2020* – December 2020.

7.2 Other publications and communications

Publications

Spini A, Giudice V, Brancaleone V, Morgese MG, De Francia S, Filippelli A, Ruggieri A, Ziche M, Ortona E, Cignarella A, Trabace L. Sex-tailored pharmacology and COVID-19: Next steps towards appropriateness and health equity. *Pharmacol Res.* 2021.

Spini A, Roberto G, Gini R, Bartolini C, Bazzani L, Donnini S, Crispino S, Ziche M. Evidence of β-blockers drug repurposing for the treatment of triple negative breast cancer: A systematic review. *Neoplasma*. 2019 Nov;66(6):963-970.

Spini A, Donnini S, Pantziarka P, Crispino S, Ziche M. Repurposing of drugs for triple negative breast cancer: an overview. *Ecancermedicalscience*. 2020;14:1071. Published 2020 Jul 13.

Spini A, Crescioli G, Donnini S, Ziche M, Collini F, Gemmi F, Virgili G, Vannacci A, Lucenteforte E, Gini R, Lombardi N, Roberto G. Drug utilization of reimbursable drugs for the treatment of COVID-19: gender differences in the Tuscan nursing home population. *Submitted on June 2021*.

Spini A, Roberto G, Bartolini C, Moscatelli V, Barchielli A, Paoletti D, Giorgi S, Fabbri A, Bocchia M, Donnini S, Gini R, Ziche M. Real word evidence on rituximab utilization: Combining administrative and hospital-pharmacy data. *PLoS One*. 2020 Mar 12;15(3):e0229973.

Hyeraci G, **Spini A**, Roberto G, Gini R, Bartolini C, Lucenteforte E, Corrao G, Rea F. A Systematic Review of Case- Identification Algorithms Based on Italian Healthcare Administrative Databases for Three Relevant Diseases of the Cardiovascular System: Acute Myocardial Infarction, Ischemic Heart Disease, and Stroke. *Epidemiol Prev.* 2019 Jul- Aug;43(4 Suppl 2):37-50. Canova C, Simonato L, Barbiellini Amidei C, Baldi I, Dalla Zuanna T, Gregori D, Danieli S, Buja A, Lorenzoni G, Pitter G, Costa G, Gnavi R, Corrao G, Rea F, Gini R, Hyeraci G, Roberto G, **Spini** A, Lucenteforte E, Agabiti N, Davoli M, Di Domenicantonio R, Cappai G. A Systematic Review of Case-Identification Algorithms for 18 Conditions Based on Italian Healthcare Administrative Databases: A Study Protocol. *Epidemiol Prev.* 2019 Jul-Aug;43(4 Suppl 2):8-16.

Oral communications

Spini A, Crescioli G, Donnini S, Ziche M, Collini F, Gemmi F, Virgili G, Vannacci A, Lucenteforte E, Lombardi N, Gini R, Roberto G. Covid19: Real word evidence on drug utilization and gender. *40° National congress Società Italiana di Farmacologia* – Digital edition 9-13 March 2021.

Posters

SpiniA, RobertoG, GiniR, BartoliniC, BazzaniL, DonniniS, CrispinoS, ZicheM Preclinical and clinical evidence about the use of betablockers for the treatment of triple negative breast cancer: a systematic review. *San Antonio Breast Cancer Symposium*. San Antonio 4-8 December 2018.

Spini A[,] Bartolini C, Moscatelli V, Barchielli A, Paoletti D, Giorgi S, Fabbri A, Bocchia M, Donnini S, Gini R, Ziche M, Roberto G. Linking administrative and hospital-pharmacy data to study drug utilization, survival and adverse infectious events in patients treated with rituximab at the University Hospital of Siena, Italy. *39° National congress Società Italiana di Farmacologia* - Florence 19-22 September 2019.

Roberto G, **Spini A**, Crescioli G, Donnini S, Ziche M, Collini F, Gemmi F, Virgili G, Vannacci A, Lucenteforte E, Lombardi N, Gini R. Utilization of drugs reimbursed by the Italian Healthcare Service for the treatment of COVID19 during the first pandemic wave: gender differences in the

general and nursing home population of Tuscany. *ICPE all access Meeting 35 years of real word science*; Digital edition August 2021.

Reports for the Tuscany regional healthcare agency

Roberto G, Bartolini C, Gini R, **Spini A**, Donnini S, Bellan C, Furiesi F, Leoncini L, Giorgi S, Rosellini P, Pascucci A, Ziche M, Barchielli A, Caldarella A, Zappa M, Salvo F. Pattern prescrittivo dei farmaci rimborsabili per il trattamento del COVID-19 e dell'ipertensione arteriosa: differenze di genere nella popolazione generale e nelle residenze sanitarie assistite della Toscana. *Rapporto farmaci della regione toscana 2020* – December 2020.