Aus der Medizinishen Klinik und Poliklinik III

Klinikum der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. rer. nat. Dr. med. Michael von Bergwelt

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.) an der

Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

German secondary data: Opportunities and challenges for assessing routine cancer care from a health institution's perspective

vorgelegt von:

Amal AlZahmi

aus:

Al Fujairah, United Arab Emirates

Jahr:

2022

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

First evaluator (1. TAC member):	*Prof. Dr. med. Helmut Ostermann*
Second evaluator (2. TAC member)*:*	*Prof. Dr. Ulrich Mansmann*
Third evaluator:	*Prof. Dr. med. Hans Roland Dürr*
Fourth evaluator:	*Prof. Dr. Marion Subklewe*

Dean:	**Prof. Dr. med. Thomas Gudermann**

Datum der Verteidigung:

05.12.2022

Table of content

Abstract:

Secondary data comprise information that is collected for purposes other than answering a specific scientific research question. Examples include patients' medical records, cancer registry data, hospital administrative data, death registry data and health insurance data. In this dissertation I aim to address the opportunities and challenges associated with using routinely collected data from Ludwig Maximilians University hospital (LMU hospital) to evaluate cancer care in routine care settings and the ability to link data from multiple secondary data sources to fill in information gaps.

This dissertation is composed of two parts. The first assesses the usability of a German hospital's administrative claims data for health services research to answer questions regarding epidemiology, management patterns, and quality-of-care for inpatients admitted with multiple myeloma (MM) from health care providers' perspectives. In the second part, multiple secondary data sources are used and linked using patient identifiers to evaluate the impact of molecular tumor board (MTB) on the routine cancer care of patients with pancreatic cancer. In the second part, the data sources used are hospital administrative claims data, physician letters, tumor board reports and pharmacy files from within patient medical records and the Cancer Retrieval Evaluation and Documentation System (CREDOS) data set from the comprehensive cancer center of Munich, Ludwig Maximilians University site.

In the first part, the hospital administrative claims data allowed case identification (n = 230 patients with MM, 59.1% were men). It also contained the required data variables that allow identification of some quality indicators necessary for quality-of-care assessments and benchmark evaluation studies (e.g., infections, readmissions, and platelet transfusions). Infections were recorded in 67% of the patients. MM patients had mean number of hospital admissions of 3.69 (standard deviation [SD] 2.71, range [1 – 16]). Eighty eight of the MM patients received stem cell transplantation (SCT), 89.8% of which received platelet transfusions at a mean of 1.42 (SD 0.63, range [1 – 3]). The primary limitation to the sole use of German hospital claims data for evaluating cancer care for patients with MM is that dates of diagnosis is missing; this means it is imposibile to determine disease onset or appropriately evaluate the chronological sequence of events preceding

or following disease onset. Missing dates of diagnosis also affected the proper evaluation and construction of lines of therapy.

In the second part, two groups (MTB = 88 and control = 165) were compared together in terms of lines of therapy, mortality, hospitalization, and costs. When hospital administrative claims data were linked to the CREDOS data set, a proper and comprehensive evaluation of inpatient lines of therapy, mortality, hospitalization and costs was conducted. An index date from the onset of an event (date of metastasis) was set and proper and equal follow-up periods (at 6 months, 12 months and 18 months) for the patients in the two groups were maintained. In the 6 month follow up cohort, for example, 15 (16.6%) patients in the MTB group had records of death compared to 92 (64.2%) in the control group. The MTB patients had mean number of hospital admissions of 1.67 (SD 1.165) compared to 2.39 (SD 1.882) in the control group (p-value <0.001). MTB had mean costs of 5337€ (SD 5067) compared to 9617€ (SD 10654) for the control group (p-value <0.001). The study showed that patients in the MTB group had significantly lower mortality rate, hospital admissions, length of hospitalization and costs. Therefore, by this analysis; signals could be identified that MTB may have a positive impact on the management of patients with pancreatic cancer from routine care settings.

In conclusion, the university hospital administrative claims data from the LMU hospital can be used to identify cases, events of interest, and quality-of-care indicators necessary for quality of hospital care assessments and benchmark evaluations. However, comphrensive evaluation of the lines of therapy is not feasible using the hospital administrative claims data alone. By contract,the CREDOS data set when linked to the hospital administrative claims data allowed the proper evaluation of the impact of the MTB on the routine care of patients with pancreatic cancer in terms of lines of therapy, mortality rates, admissions and cost. Therefore, the routine clinical data sources allowed me to fill in information gaps I encountered in the hospital administrative claims data. This disseartion addresses the opportunities and challenges of using routine data from the LMU hospital. I believe it can serve as a blueprint for researchers aiming at utilizing routine care data in health services research by addressing the required data variables and their corresponding data sources.

List of figures

List of tables

List of abbreviations

CCC	Comprehensive cancer center
CGP	Comprehensive genomic profiling
CREDOS	Cancer Retrieval Evaluation and Documentation System
CT	Computed tomography
DKTK	German Cancer Consortium
DRG	Diagnosis-related group
ICD-10-GM	International Classification of Diseases, Tenth Revision, German Modification system
ID	Identifier number
IPTW	Inverse probability treatment weigthing
ISS	International staging system
LMU	Ludwig Maximilian University
MM	Multiple myeloma
MRI	Magnetic resonance imaging
MTB	Molecular tumor board
OPS	Operation and procedure codes
PET/CT	Positron Emission Tomography/Computed Tomography
PSE	Propensity score
RCT	Randomized control trial
SCT	Stem cell transplantation
SD	Standard deviation
SHI	Statutory health insurance

1. Introduction

Real-world data sources can be considered either primary or secondary [1]. Primary data are collected to answer a scientific research question. Secondary data, by contrast, are transactional data that have been collected for other purposes such as clinical documentation (e.g., electronic medical records) or administrative use (e.g., claims data). Researchers consider secondary data as a valuable source of information that can represent real-world routine healthcare provisions where their results can complement that of randomized controlled trials (RCTs) [2]. Primary and secondary real-world data have been used in research on health services research, epidemiology and health economics [2].

Secondary data analyses can be used for descriptive, exploratory or correlational designs or non-parametric statistical tests, but they cannot be used to confirm causal relationships [3]. However, they possess advantages that make them attractive for use in health research studies. Secondary data analyses assure economic saving in time, money and labor, helps in hypothesis generation, or provides subtle assessment of primary results from original studies [3,4]. An advantage of secondary data is the ability to combine multiple data sources to gather information on as many variables as needed to fill in knowledge gaps [3]. Secondary data sources also include large study populations and specific subpopulations that are difficult to recruit for prospective observational studies, rendering them potentially convenient alternatives for health research [5,6]. Additionally, the use of secondary data in health research can prevent study repetition, and over-research of sensitive topics or populations [3]. Compared to primary data, secondary data sources include a large number of variables that are timely updated and available for use [6]. However, to what extent this can be applied to secondary data sources from the German health care context has yet to be evaluated.

In the field of oncology, secondary data sources play an integral role in allowing representation of health provisions from routine care settings [7–13]. Electronic medical records (EMRs), administrative claims data, and cancer registry and health insurance databases are examples of real-world data used in oncology research [2,5,6,13,14]. Cancer is a disease that is often rare, complex, chronic, or incurable. Cancer treatment too can be highly complex, involving

many drug combinations, frequent changes in drug regimen over the course of the disease, and treatment duration that vary from patient to patient [6]. Cancer patients treated in routine care settings differ in characteristics, disease complexity and treatment options from those treated in RCTs. Thus, it is important to identify a study population representative of the general population and evaluate its clinical and economic outcomes from routine care settings using secondary databases.

In North America, utilization of secondary data dates back to the 1980s [2]. Many oncology research studies have utilized secondary data, the majority of which are based on US secondary databases [8,10,11,13–18]. In Germany, legal changes in 2011 (Versorgungsstrukturgesetz) that allowed utilization of routinely collected data (as well as allowed better access to it and funded related research) positively influenced interest in using this data for health studies [19]. Publications based on secondary data sources in the German context had increased over the past few years compared to the years before 2000 [2,5,19–25]. In Germany, routinely collected data are present in various formats, including hospital-based administrative claims data, statutory health insurance claims data (SHI), office-based physicians' association data, cancer registry data, death registry data and data from federal databases [2,5,19,20,24,26]. Though the aforementioned data sources differ slightly in terms of their granularity and depth, they complement one another. Thus, the linkage of multiple data sources in a research study could provide a comprehensive perspective on research questions and fill in information gaps driven by the use of a single data source. Data linkage of multiple data sources is not a concern in the US setting [12,13]. However, in the German setting, only a few recent studies in the oncology field have linked routinely collected data [27].

In this PhD dissertation, I aim to address the opportunities and challenges associated with using routinely collected data from Ludwig Maximilians University hospital (LMU hospital) to evaluate cancer care in routine care settings and the ability to link data from multiple secondary data sources to fill information gaps. In this context, "data linkage" refers to linking hospital administrative claims data to data from the cancer registry using a unique patient identifier number (Patient's ID). This dissertation includes two parts. The first addresses the use of hospital administrative claims data to evaluate multiple myeloma (MM) care in inpatient

settings. Information gaps and limitations associated with the use of a single data source are discussed. In the second part, I address the use of routinely collected data in the form of patient medical records, cancer registry data collected via the Cancer Retrieval Evaluation and Documentation System (CREDOS) within the Comprehensive Cancer Center - LMU site (CCC-LMU) and hospital administrative claims data to evaluate routine care of patients with pancreatic cancer managed under the so called “molecular tumor board (MTB)”.

These projects provide a practical examples of the use of routinely collected data and data linkage to evaluate cancer care in a single hospital. This method of data linkage could then be applied to create a broader, more comprehensive perspective from which to evaluate cancer care from different health care institutions or through the use of multiple data sources. Doing so would improve evaluation of routine cancer care, quality-of-care assessments, and benchmark evaluations.

2. Real-world data from a German university hospital

2.1 German hospital administrative claims data

Claims data analysis in Germany was influenced by the historic evolution of the German healthcare system [2]. Claims data have been used in research studies to help with a) research purposes and health related political decisions, b) health policy development, and c) evaluation of health service delivery and quality-of-care assessment [2]. Claims data comprise cross-sector data on billable medical interactions between insured individuals and the healthcare system [6]. The use of claims data in health services studies to evaluate health economics in Germany has increased over the past few years [19]. Based on a systematic review by Gansen, most such studies (24 of 35) used claims data from SHI ; 4 out of 35 used data from other sources of health insurance data and 5 out of 35 used hospital administrative claims data, German trauma registry data and the International Marketing Services Health database [19].

Despite the fact that SHI databases include a wide range of information, they contain limited information for quality assessment of hospital care from healthcare provider's perspective [28]. However, hospital administrative claims data are collected and provided in the format of the §21 dataset [29], contain more granular information on all health services provided to the patient during their hospitalization regardless of their health insurance than claims data from SHI or other health insurcance data. The amount of information they cover extend to include all therapeutics and diagnostic interventions that are reimbursed beyond the diagnosis-related group (DRG) system from inpatient healthcare providers perspectives. Thus, they could serve as a better alternative for quality-of-care assessments and benchmark evaluation studies from the health care providers perspective.

2.1.1 Multiple myeloma – use case

2.1.1.1 Etiology and Epidemiology of MM

MM is a rare, and incurable malignant disease characterized by abnormal accumulation of plasma cells in the bone marrow [13,15,30–32]. It accounts for 1% of all cancers and 10 – 15% of all hematologic malignancies worldwide

[30,33,34]. MM primarily affects older people with a median age at diagnosis of 72 years [13,15,30,31]. Only 2% of those diagnosed with the disease are under the age of 45 years [34,35]. In most patients, MM evolves from a precancerous stage, called “monoclonal gammopathy of undetermined significance” [30,35]. In others, it emerges from an intermediate but more advanced asymptomatic, pre-malignant stage called “smouldering MM” [30].

In Germany, MM is the third-most-common hematologic neoplasm after leukemia and non-Hodgkin lymphoma with approximately 6500 new cases diagnosed each year [34]. In 2016, nearly 3,000 women and 3,910 men were diagnosed with MM [35]. In 2017, according to the age-standardized mortality rate of MM, MM accounted for 1.9 of every 100,000 deaths in women and 3.2 of every 100,000 deaths in men [35].

2.1.1.2 Diagnosis of MM

The International Myeloma Working Group updated the diagnostic criteria for MM in 2014 [30]. Diagnosis of MM requires the presence of at least 10% bone marrow plasma cell infiltration, a bony or extramedullary plasmacytoma confirmed by a biopsy, and any of the myeloma-defining events (MDE) [30]. MDE consist of the presence of one or more of the CRAB criteria (hypercalcemia, renal insufficiency, anemia or bone lesions) and detectable biomarkers of malignancy (≥60% clonal bone marrow plasma cells or an involved/uninvolved serum free light chain ratio >100%) [30]. MM has a variable course and heterogenous clinical behavior. The prognostic feature of MM are based on the levels of serum β2-microglobulin, albumin, C-reactive protein, and Lactate dehydrogenase (LDH) [30]. Risk assessment and prognostic evaluation of MM could be improved by combining the revised-International Staging System (R-ISS), cytogenetic testing and LDH evaluation [30]. The ISS is a three-stage system that defines disease stages as I, II or III, where stage III is associated with poor outcomes [30].

2.1.1.3 Treatment of MM

The MM treatment landscape is continuously changing. MM primarily aims to provide symptomatic relief, control the disease, and increase the overall patients survival rates [13,36–38]. Patients with MM are treated based on their clinical fitness and age. Elderly patients not fit for stem cell transplantation (SCT) are treated with front-line combination therapy [30]. Bortezomib/melphalan/

prednisone (VMP) and lenalidomide/low-dose dexamethasone (Rd) have been approved by the European Medicines Agency (EMA) as the standard of care in this setting [30]. Other combination therapies are also used as frontline therapy in routine care settings for this group (e.g., bortezomib/cyclophosphamide/dexamethasone [VCD] and bendamustine/prednisone) [30]. By contrast, the standard of care for younger and clinically fit patients is 4 – 6 cycles of induction therapy followed by high-dose therapy with autologous SCT (ASCT) [30]. As standards of care, the following combination regimens are commonly used in Europe as induction therapy: bortezomib/dexamethasone/thalidomide and bortezomib/dexamethasone/cyclophosphamide [30]. Currently, there is insufficient evidence for the benefits of consolidation therapy, so it is not considered the standard of care in routine care settings. However, as the use of lenalidomide as maintenance monotherapy in adult patients following ASCT has been approved by the EMA.

2.1.1.4 Rational for MM as a Use Case

Despite improvements to the overall survival of patients with MM due to the use of novel agents, MM poses an economic burden that must be evaluated and addressed within routine care settings [2]. Previous studies on MM in Germany conducted using real-world data were based primarily on surveys [20], patient charts [21] and SHI claims data [20]. Depending on their quality and granularity, German hospital databases may provide valuable information for epidemiologic and health services studies, benchmark evaluations, and quality-of-care assessments. For rare conditions such as MM, secondary databases can serve as ideal sources of evidence concerning management patterns and healthcare resource utilization in routine care settings.

2.1.2 Study objectives

The aim of this study is to assess the usability of German hospital administrative claims data to determine inpatient management patterns, healthcare resource utilization, and quality-of-care for patients with MM.

2.1.3 Materials and Methods

2.1.3.1 Study Design and Data Source

I conducted a retrospective observational study based on hospital administrative claims data from the LMU hospital [41]. Data from 2015 – 2017 was used for this study. The LMU hospital is a tertiary university hospital with a specialized hematology-oncology department that includes a specialized MM treatment center.

Hospital administrative claims data in German hospitals are collected using a uniform structure of §21 dataset [29], which is a performance and flat-rate data set based on the German DRG (G-DRG) system and the International Classification of Diseases (Tenth Revision, German Modification; ICD-10-GM) system. The data set contains information on all health services (e.g., diagnostic and therapeutic procedures reimbursed beyond the DRG system) provided to patients during their hospitalization irrespective of their health insurance. The dataset contains information on patient IDs, case numbers, pay area (e.g., DRG, additional fees, fees new examinations and interventions), health insurance ID, demographics (e.g., age, gender), reasons for admission (primary vs. secondary diagnosis), admitting department, diagnosis codes, localization of diagnoses, procedure codes, dates of procedures, admission and discharge dates, and reasons for discharge or transfer. The data set was available as a spreadsheet in four distinct data tables named FAB (stands for Fachabteilung which refers to the admitting department), FALL (stands for the case), operation and procedure codes (OPS) and ICD. The hospital administrative claims data set was anonymized by the Trust Center and processed by the Medical Data Integration Center (MeDIC), both of which are located at LMU-hospital and operate under the Data Integration for Future Medicine (DIFUTURE) consortia of the Medical Informatics Initiative (MII) which is funded by the German Federal Ministry of Education and Research (BMBF) [42]. In this context, the hospital administrative claims data set was used for the MII's national, cross-consortia demonstrator study after approval by the LMU faculty of Medicine Ethical Review Board of LMU's Faculty of Medicine and LMU hospital Data Protection Officer.

2.1.3.2 Study Sample

The sample consisted of patients with MM with inpatient records during 2015 – 2017. Patients older than 18 were included if they fulfilled at least one of

the following conditions: (1) had at least one inpatient MM diagnosis (ICD-10 = C90.0) as the primary reason for hospitalization or (2) had received anti-MM therapy. The ICD-10 code for identifying patients with MM was validated elsewhere [7].

2.1.3.3 Data Completeness Tool and Outcome Measures

Based on narrative literature review of papers that investigated MM care using administrative claims data, I constructed a list of necessary data elements (Appendix A). I first evaluated the research questions, methods, and data required to answer each research question, and I noted prominent findings in the identified reports. The list also contained the required OPS and specific ICD-10-GM codes for identifying medications used, procedures performed, and diseases diagnosed. The list should represent a nearly complete list of variablest required to answer healthcare research questions, and it should also serve as a blueprint for future studies aiming to linking multiple secondary data sources by providing the sources for each variable. Based on this list, I constructed a data completeness tool to evaluate the extent of data variable availability in the LMU hospital administrative claims data.

After conducting a data availability check on the hospital administrative claims data set, I analyzed the comprehensiveness and usability of datases elements. First, I examined the demographic characteristics of patients with MM, including age and gender. Second, I examined their clinical characteristics in terms of disease stage and severity, comorbid conditions, disease- and treatment-related complications, and in-hospital mortality. Third, I examined management patterns in terms of prescribed medications, lines of therapy, and therapeutic and diagnostic procedures. Fourth, I identified anti-MM therapy documented in the data set. Anti-MM therapy included administration of bortezomib (OPS code 6-001.9), lenalidomide (OPS code 6-003.g), or combination therapy (OPS codes 8-542, 8-543, and 8-544). Finally, I evaluated healthcare utilization in terms of the number of hospital readmissions that lasted more than 24 hours, hospitalization length, and therapeutic and diagnostic procedures performed.

In a subgroup analysis, I evaluated patients who had undergone autologous stem cell transplantation (SCT) because, for this group, the start date of the procedure could be used as an an index date. Thus, the chronological sequence

of events following the first SCT procedure could be evaluated. I then assessed the possibility of evaluating these patients' clinical characteristics in terms of complications following SCT, management pattern in terms of treatment received, and reason for hospitalization after the procedure.

2.1.3.4 Statistical Analysis.

Patients were identified using their patient IDs. The patient's ID variable was present in the four data tables, and it was used as a linking variable to merge the tables together. First, I merged the FALL and ICD data tables to identify patients with MM. Second, I created a binary variable for gender (1= male, 2 = female). I then created three age categories (<65, 65 – 70 and ≥70 years). After that, I used the data variable ICD-10-GM to identify patients with an ICD-10-GM code equal to C90.0- as patients with MM. I then merged the above table with the OPS table to identify treatments and procedures performed during hospitalization, and I applied the inclusion and exclusion criteria to include only patients who had an ICD-10-GM code equal to C90.0- as a primary reason for admission or who had received anti-MM therapy during their hospitalization. Finally, I merged the this data table with the FAB table to evaluate admissions and length of hospital stays.

I then performed descriptive analysis. Categorical variables (age, gender, comorbid conditions, complications, in-hospital mortality, management pattern, diagnostic and therapeutic interventions) are presented descriptively as counts and percentages. Continuous variables (age, number of hospital admissions and hospitalization length, and platelet transfusions) are presented as means and standard deviations (SD). Statistical analyses were conducted using SAS 9.4 software (X64 10HOME platform, Copyright [c] 2002–2012 by SAS Institute Inc., Cary, NC, USA). I used Rstudio 3.6.1 (Version 1.2.500© 2009–2019, Inc.) to produce a sunburst chart to illustrating the treatment provided to SCT patients during their first three admissions following the SCT procedure.

2.1.4 Results

2.1.4.1 Data Availability

The hospital administrative claims data set contained variables required for case identification and evaluation of age and gender distribution among patients with MM (Table 1). It included some but not all information required to evaluate patients' clinical characteristics. Additionally, it contained variables required for identifying possible comorbid conditions and disease- and treatment-related complications based on ICD-10-GM codes, and it also included information for identifying in-hospital mortality with a variable termed “discharge/transfer reason.” Diagnosis date was not recorded in the data set; thus, I could not maintain the same follow-up period for all patients. It also limited our ability to rigorously evaluate the chronological sequence of events and distinguish between unrelated comorbid conditions, and disease- and treatment-related complications. In other words, I could not set an index date from disease onset and examine patients' clinical history over time to identify the occurrence and development of other conditions or complications. Moreover, details regarding laboratory and radiological findings and disease stage and severity were unavailable, which hindered the evaluation of disease stage and severity as well as disease risk assessment.

To evaluate management patterns, I identified prescribed medications and diagnostic and therapeutic procedures performed using pre-specified procedure codes. This approach allowed me to evaluate the treatment provided and the diagnostic and therapeutic procedures performed in terms of documentation frequency. However, because of the missing diagnosis dates, I was unable to construct a line of therapy. Treatment initiation date, therapy duration and dose, and evidence of treatment discontinuation or switching were not recorded, limiting the evaluation of treatment patterns.

The hospital administrative claims data set contained information for assessing healthcare utilization in terms of the number of hospital admissions and length of hospital stays. However, the data set was limited to a single hospital department, and no data on outpatient or emergency department (<24 hours) visits were available. Therefore, the data set did not capture admissions to other departments within the same hospital.

The following Table 1 – Table 4 and Figure 1 – Figure 2 of this study have been published in PLOS ONE [41].

Table 1: Data completeness evaluation tool

Outcome measures	Question	Data availability	
		Yes	No
Demographics	Can patients be identified using the data set?	X	
	Can the age distribution of the identified group of patients be determined?	X	
	Can the sex distribution of the disease group be identified?	X	
Clinical characteristics	Can the diagnosis of MM be confirmed using the data set?		X
	Can the disease stage of patients with multiple myeloma be assessed using the data set?		X
	Can the disease risk in the identified group of patients be assessed using the data set?		X
	Can comorbid conditions documented in the MM group be identified?	X	
	Can disease- and/or treatment-related complications documented in the MM group be identified?	X	
	Was in-hospital mortality of patients with MM documented in the data set?	X	
Management	Can medications used for the following applications be identified? • Front-line therapy • Relapsed/refractory multiple myeloma • Consolidation therapy • Maintenance therapy • Supportive care		X
	Can therapeutic procedures performed on the disease group be identified? • Therapeutic plasmapheresis • Hemodialysis • Blood transfusion • Thrombocyte transfusion • SCT	X	

	Can diagnostic procedures performed on the disease group be identified? • Computed tomography (CT) • Magnetic resonance imaging (MRI) • Positron emission tomography (PET)/CT • Conventional radiographs • Immunocytochemical detection of circulating tumor cells • Bone marrow biopsy • Genetic testing • Pulmonary function test • Endoscopy	X	
Health resource utilization	Can the frequency with which patients with MM were admitted to the hospital be identified?	X	
	Can the length of time patients with MM stayed in the hospital be identified?	X	
	Can the health services used by the disease group be identified? • Laboratory: (Complete blood count, serum/urine protein electrophoresis, cytogenetic test, bone marrow aspiration/biopsy) • Radiology: (CT, MRI, PET-CT) • Therapeutic: (chemotherapy, radiotherapy, immunotherapy, SCT, blood transfusion, thrombocyte transfusion, plasmapheresis, dialysis, antiviral, antifungal, antibiotic and supportive therapy*)	Not complete	

Abbreviations:
CT: Computed tomography
MRI: Magnetic resonance imaging
PET/CT: Positron emission/computed tomography
SCT; Stem cell transplantation

2.1.4.2 Description of the Study Sample

I identified 325 patients with an MM diagnosis code, of which 222 (68.3%) had MM as the primary reason for admission. An additional eight patients who received anti-MM therapy but were not admitted primarily for MM were included. Overall, 230 patients with MM were included in the study.

Patients' mean age at first admission was 65 years (SD = 12). One-hundred thirty-six of the MM patients were (59.1%) men. In total, 196 (85.2%) were readmitted to the same hospital within 1 year (Table 2). Hypertension (50.0%), chronic kidney disease (32.6%), and other tumors (21.7%) were the most documented comorbid conditions (Table 1). Infection (67.0%), neutropenia (50.0%), and thrombocytopenia (50.3%) were the most documented disease- or treatment-related complications. In-hospital mortality was reported in 12% of patients.

Table 2 Demographics, clinical characteristics, and management patterns of patients with multiple myeloma

Patients with MM	**N = 230** **n (%)**
I. Patient demographics:	
a. Sex:	
Male	136 (59.1%)
Female	94 (40.9%)
b. Age group	
<65 years	101 (43.9%)
65–70 years	37 (16.1%)
≥70 years	92 (40.0%)
Mean age at first admission (SD)	65 (12)
(min–max)	(34–89)
II. Readmission episodes following first admission/year	
Within 1 year	196 (85.2%)
Within 2 years	22 (9.6%)
Within 3 years	4 (1.7%)
III. Comorbid conditions:	
Hypertension	115 (50.0%)
Congestive heart failure	36 (15.6%)
Cerebrovascular disease	15 (6.5%)
Chronic kidney disease	75 (32.6%)

Other tumors†	50 (21.7%)
Diabetes mellitus type 2	36 (15.7%)
Ischemic heart disease	29 (12.6%)
IV. Disease or treatment-related complications:	
Skeletal-related events	76 (33.0%)
Anemia	51 (22.2%)
Drug-induced anemia	89 (38.7%)
Renal complications	44 (19.1%)
Gastrointestinal bleeding	9 (3.9%)
Infections‡	154 (67.0%)
Urinary tract infections	51 (22.2%)
Thrombocytopenia	116 (50.3%)
Peripheral neuropathy	26 (11.3%)
Neutropenia	115 (50.0%)
End-stage renal disease	54 (23.5%)
Underweight§	7 (3.0%)
V. In-hospital mortality	28 (12.2%)
VI. Management pattern	
a. Anti-MM therapy	
Combination therapy	162 (70.4%)
Bortezomib	83 (36.1%)
Lenalidomide	25 (10.9%)
Immune therapy	29 (12.6%)
b. Supportive therapy	
Pain medication	14 (6.1%)
Lipegfilgrastim¶	36 (15.7%)
Antifungal medications	42 (18.3%)
c. Therapeutic procedures	
SCT	88 (38.3%)
Blood product transfusion	164 (71.3%)
- Platelet transfusion	117 (50.9%)
Stem cell collection	73 (31.7%)
Hemodialysis	23 (10.0%)
d. Diagnostic procedures	
Computed tomography scan	187 (81.3%)
Pulmonary function test	120 (52.2%)
Bone marrow biopsy	93 (40.4%)
Magnetic resonance imaging (MRI)	62 (27.0%)
Diagnostic endoscopy	31 (13.5%)

†Primary benign, malignant, and unspecified tumors as well as secondary tumors were grouped together under one category.

‡Infections included cholera, typhoid and paratyphoid, salmonella infections (enteral salmonella, sepsis salmonella, localized salmonella, and unspecified salmonella); shigellosis; bacterial stomach infection (*E.coli*); bacterial enteritis; foodborne bacterial illness; amebiasis; intestinal diseases caused by protozoa; viral-induced gastroenteritis; other unspecified infectious gastroenteritis and colitis of unspecified origin; meningococcal infection; streptococcus infections; unspecified sepsis, bacterial infections of unspecified localization; other viral encephalitis not otherwise classified; unspecified viral encephalitis; viral meningitis; other unspecified viral infection of the central nervous system; viral infections of unspecified localization; streptococci and staphylococci as the cause of infections classified in other chapters; other specified bacteria as the cause of diseases classified in other chapters; viruses as the cause of diseases classified in other chapters; other specified infectious agents as the cause of diseases classified in other chapters; herpes simplex infection; varicella infections; herpes zoster infection; smallpox; rubeola; and viral-induced skin and mucosal diseases.

§ICD-10 codes for underweight include R63.4= abnormal weight loss and R64= cachexia.

¶Lipegfilgrastim is a medication used to treat neutropenia in patients with cancer.

Abbreviations:

MM: Multiple myeloma

SCT: Stem cell transplantation

Approximately 70.4% of patients received combination therapy. Of the novel therapies, bortezomib (36.1%) and lenalidomide (10.9%) were most frequently administered, and 38.3% of patients underwent SCT. Of the therapeutic modalities, blood transfusion (71.3%) was the most frequently documented (Table 2). Computed tomography (81.3%) and pulmonary function tests (52.3%) were the most frequently documented diagnostic modalities. Patients with MM were admitted a mean of 3.69 (SD = 2.71) times for a mean duration of 12.52 (SD = 9.55) days per hospital stay (Table 3).

Table 3: Health resource utilization by patients with multiple myeloma

Health resource utilization (per patient) in patients with MM (N = 227) †	**Mean**	**SD, (Min–Max)**
Number of admissions‡	3.69	2.71 (1–16)
Average duration of each hospital stay (in days)	12.52	9.55 (1–68.5)
Total duration of hospital stays (in days)	40.25	34.99 (1–247)

†Admissions <24 h were excluded from the analysis.

‡Number of admissions calculated over the 3-year study period.

For subgroup analysis, procedure dates of the first documented procedure were set as an index dates, and patients were followed up prospectively. Among patients who had undergone SCT (n = 88), 71.6% were under 65 years old, with a mean age of 58 during the first SCT procedure (Table 4). The first SCT procedure was performed a mean of 98.5 (SD = 83) days after the first recorded admission. Sixty-seven patients (76.1%) underwent a single SCT, 20 (22,7%)

received two SCTs, and one patient received three SCTs during the 3-year study period. Regarding possible disease- or treatment-related complications, neutropenia (100%), thrombocytopenia (87.5%), and infection (78.4%) were most frequently documented (Table 4). After SCT, 27 (30.7%) patients were readmitted at least once, with MM (56.9%) being the primary reason, followed by other tumors in only two of the patients (21.7%; Fig 1). For the first three post-SCT readmissions, combination therapy (100%) and blood transfusions (96.3%–100%) were the most frequent (Fig 2). Bortezomib (40.7%) and lenalidomide (14.8%) were used post-SCT.

Table 4: Characteristics of patients who underwent stem cell transplantation

SCT-MM patients	**N = 88** **n (%)**
a. Age distribution	
≤65years	63 (71.6%)
>65years	25 (28.4%)
Mean age at the time of first SCT (min–max)	58 (34–74)
Mean number of days from first admission to first SCT (SD)	98.6 (83)
b. Number of SCT procedures	
1	67 (76.1%)
2	20 (22.7%)
3	1 (1.1%)
c. Platelet transfusion	79 (89.77%)
Mean (SD [range])	1.42 (0.63 [1–3])
d. Treatment-related or disease-related complications in SCT patients	
Neutropenia	88 (100%)
Thrombocytopenia	77 (87.5%)
Infections	69 (78.4%)
Hypokalemia	56 (63.6%)
Drug-induced anemia	46 (52.3%)
Abbreviations: SCT: Stem cell transplantation	

Figure 1: Reason for and duration of hospital readmission after SCT

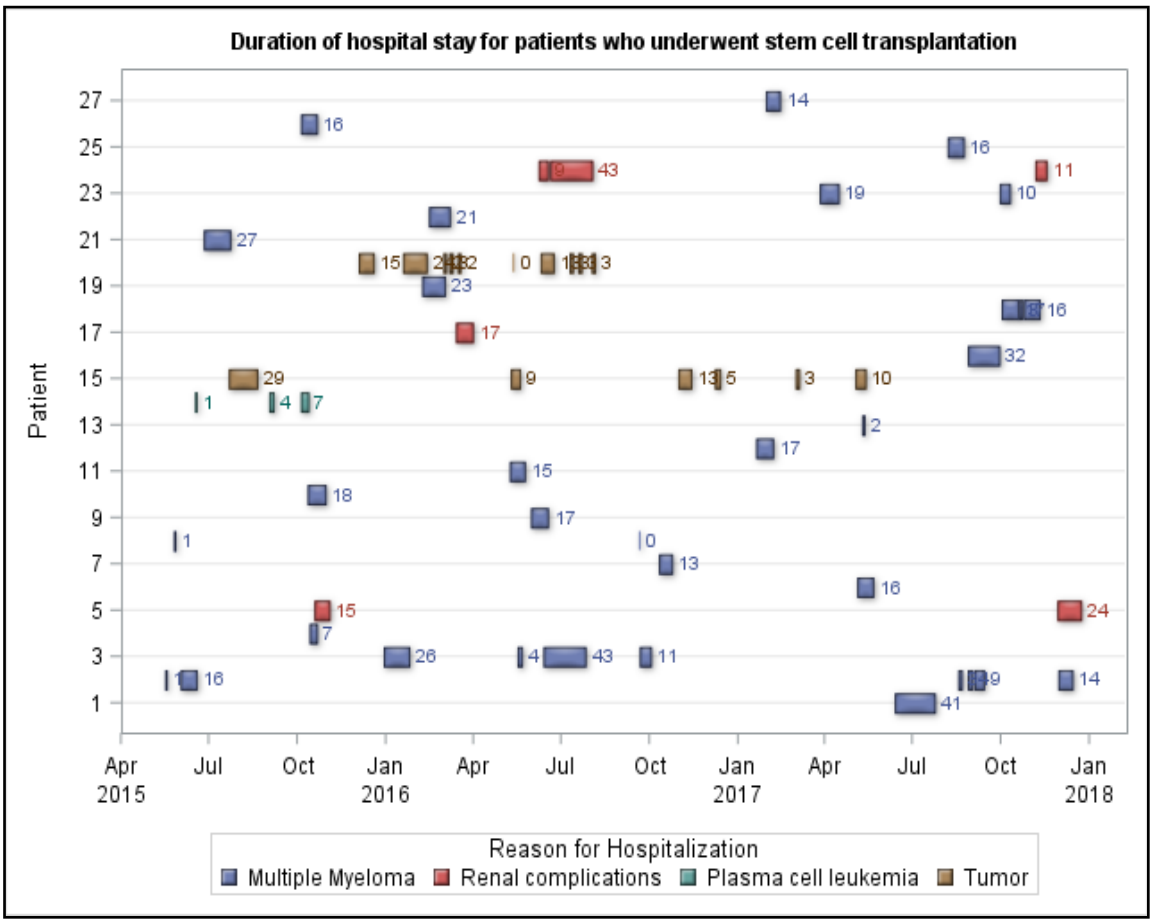

Figure 1 shows the duration of hospital stays (in days) and reason for each admission after a SCT

Figure 2: Treatment received during each subsequent admission following stem cell transplantation

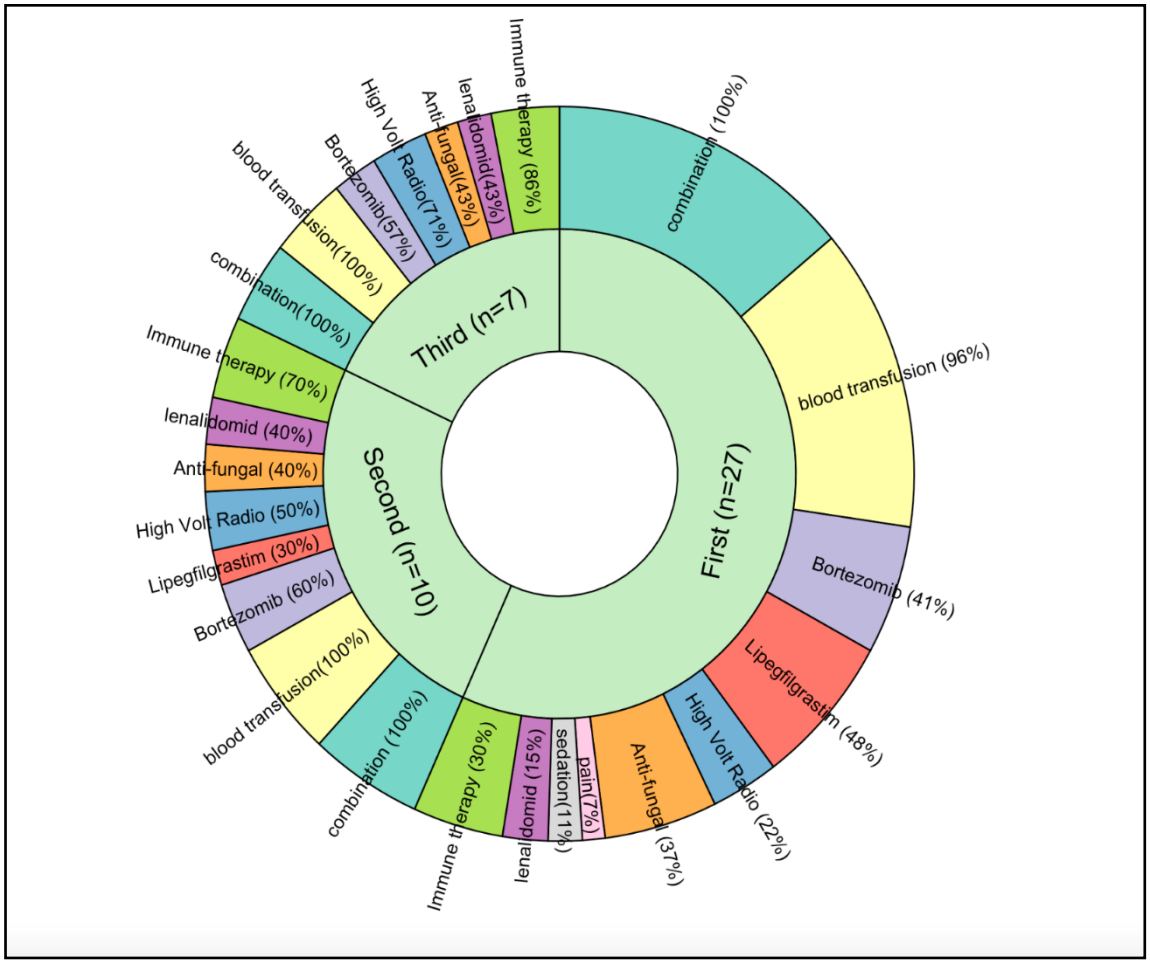

Figure 2 shows the treatment provided to SCT patients during the first three readmissions after a SCT procedure.

Combination therapy, Blood transfusion, Bortezomib, Lipegfilgrastim, High-voltage radiotherapy, Antifungal therapy, Pain therapy, Sedation therapy, Lenalidomide, Immune therapy

2.2 Routine clinical care data from a German university hospital

The CCC-LMU runs a weekly Molecular Tumor Board (MTB) with experts from different fields in order to better understand the individual patient's cancer through comprehensive genomic profiling (CGP) to tailor targeted therapy "precision medicine" accordingly. However, the, effectiveness, and benefit of such innovative interventions in providing higher care-value for patients with pancreatic cancer as well as how such innovations are utilized in routine clinical settings in Germany remains to be evaluated. Within the CCC-LMU, clinical data are collected using the CREDOS [43]. The CREDOS is used to record all oncology-relevant data about the patients treated at the LMU hospital. In this project, I linked CREDOS data to hospital administrative claims data using patient IDs to track patient information and interaction within the hospital to evaluate the impact of MTB on the routine care of patients with pancreatic cancer. Additionally, I collected missing clinical details from physicians' notes, tumor board reports and pharmacy files from patient medical records.

2.2.1 Pancreatic cancer – use case

2.2.1.1 Precision Medicine in Pancreatic Cancer Care

Precision medicine is constantly gaining relevance for the management of different diseases. In this project, I focus on the impact of precision medicine on the field of oncology, with a special focus on pancreatic cancer. Precision medicine is a two-step process wherein patients who might benefit from this intervention are identified using CGP, and based on the results, a targeted therapy or course of therapies is prescribed. Throughout this process, the idea of administering "the right treatment to the right patient" is upheld [44,45]. CGP is important not only for identifying the genetic alterations a cancer patient might have but also for implementing in all phases of disease management [46]. Ideally, the use of CGP improves patient selection for therapeutic interventions, predicts response to treatment or disease progression, assess toxicity risks and hastens in the early detection and classification of cancer [44–47]. In other words, it helps to identify patients most likely to benefit from a treatment. Thus, it provides evidence on the clinical utility of not only the targeted therapies themselves but also of the molecular diagnostics used to evaluate the patients who received them.

Precision oncology is a multi-stakeholder field, and its stakeholders have diverse needs and interests. Therefore, the adaptation of such an innovative approach in the healthcare system requires acceptability and understandability of the value provided by such an intervention by the different stakeholders involved [47]. However, despite its importance, there are no uniform criteria to evaluate or assess the economic value based on the use of precision medicine in routine care settings [47]. From a health economics perspective, not only costs of diagnostic tests and treatment are important, but also the type and number of diagnostic tests performed after initial testing, the type and frequency of therapeutic interventions provided, the rate of adverse events (e.g., infection), the mortality rate, and the health resources utilized by patients during management should be considered. Although it has been argued that diagnostic tests do not treat patients and therefore do not directly affect patient outcomes, they in general influence about 60 – 70% of all clinical decisions at a cost not exceeding 4 – 5% of all healthcare costs [47]. However, how much of that cost is related to oncology management has not yet been evaluated. The ability to identify patients most likely to benefit from a therapeutic intervention would eventually help to appropriately allocate resources and reduce the use of unnecessary and less effective intervention; thereby lowering costs [48].

2.2.1.2 Epidemiology of Pancreatic Cancer

Pancreatic cancer is most common in North America, Europe and Australia [49,50]. It was the third-most-common cancer-related cause of death in 2016 [51] and is projected to be the second-most common cancer-related cause of death in the US in 2030 [50]. In Europe, death related to pancreatic cancer has continued to increase and is the fourth-most-common cancer-related cause of death for both men and women [52]. The same figures also apply to Germany with rising incidence and mortality rates [53]. In Germany, pancreatic cancer occurs in older patients, with a mean age at diagnosis of 72 for men and 76 for women [53]. It also tends to be more prominent in men (age-standardized rate: 5.5/100,000) than women (age-standardized rate: 4/100,000) [50]. In the majority of cases, the disese is diagnosed late when the tumor has already distally metastasized and is no longer resectable [50,51]. If left untreated, the median survival rate for patients with pancreatic cancer is 3 months [50,51]. Surgical resection is the only chance to cure the disease, however, the recurrence rate is very high [50–52]. Pancreatic

cancer has a very poor 5-year survival rate that ranges from 2 – 9% [50]. Pancreatic ductal adenocarcinoma which accounts for 80% of all pancreatic cancer, is the most common form of this disease [52]. The majority of pancreatic cancer (60 - 70%) arises from the pancreatic head, followed by 20 – 25% from the body or tail while 10 - 20% from the whole pancreatic organ [52]. Symptoms of pancreatic cancer depend on the area of the pancreas that is affected. Around 90% of pancreatic cancer are associated with common risk factors such as tobacco smoking, alcohol consumption, age, helicobacter pylori infection, diet and high bod-mass-index (BMI) while 5 -10% of cases are due to genetic alterations [52]. Among those with genetic mutations, more than 80% of the cases happen due to sporadic mutations, and less than10% arise due to inherited germline mutations [52].

2.2.1.3 Diagnosis of Pancreatic Cancer

Screening for pancreatic cancer is not the standard of care, but it can be applicable to high-risk individuals, including those with a strong family history of pancreatic cancer. [50] Endoscopic ultrasound (EUS), MRI and CT scans are commonly used to diagnose pancreatic cancer. However, EUS is more sensitive in detecting pancreatic lesions less than 2cm in size than the other two diagnostic modalities. Additionally, when combined with fine-needle aspiration cytology, EUS is able to more accurately identify features of potentially cancerous pancreatic cysts than the other two. [50,52]

Recent advancements in molecular testing have allowed the integration of CGP and molecular testing in the management of pancreatic cancer. CGP can yield important information that may eventually affect the treatment decisions. [51] The most recent National Comprehensive Cancer Network guidelines recommend molecular testing in patients with metastatic disease (the focus of this study). [51] Multiple combinations of genetic mutations in pancreatic cancer have been well studied and can be grouped as 1) mutational activation of oncogeneses (KRAS) found in 90% of pancreatic cancer, 2) inactivation of tumor suppressor gene (TP53, p16/CDKN2A, and SMAD4), and 3) inactivation of the genome maintenance genes (hMLH1, and MSH2) that controls the repair of DNA damage. [52] Tumor markers (e.g., CA19-9), by contrast, have a value as prognostic markers but not for primary diagnosis of the disease. They can be better utilized to detect disease burden and guide treatment decision. [52]

2.2.1.4 Treatment of Pancreatic Cancer

Pancreatic cancer can present in one of the following forms: 1) resectable, 2) borderline resectable, 3) locally advanced, or 4) metastatic [52]. Treatment of pancreatic cancer differs from patient to patient based on the resectability of the cancer and the patient's performance status. Previously, patients with metastatic cancer were treated with gemcitabine alone. However, recent advancement in therapeutics lead to the introduction of FOLFIRINOX regimen for metastatic disease in patients with a good performance status. FOLFIRINOX is a combination therapy that consists of folic acid, 5-fluorouracil, irinotecan and oxaliplatin. As an adjuvant therapy, FOLFIRINOX shown improvement in disease-free survival and overall survival and is recommended by the National Comprehensive Cancer Network for fit patients. [51,52]

Based on the European Society for Medical Oncology clinical practice guidelines for treatment of pancreatic cancer, patients with poor performance status of 3 or 4 with significant morbidities are only offered symptomatic treatment. [52] Those with performance status of 2 with a heavy tumor load can be offered gemcitabine-based therapy. For patients with a good performance status of 0 or 1 with bilirubin level less than 1.5 times the upper limit of normal (ULN) FOLFIRINOX or Gemcitabine-based therapy can be administered [52].

2.2.1.5 Rationale for Pancreatic Cancer as a Use Case

In Germany, the impact of applying CGP in routine cancer care on patients with pancreatic cancer has not been evaluated from an economic and epidemiologic standpoint using secondary data sources. Such evidence is of great value to stakeholders, third-party payers, policymakers, and healthcare providers, not only to inform formulary and insurance coverage decisions but also for developing preferred practice guidelines to improve patient outcomes. In fast-moving fields, it is important to track economic and epidemiological factors in a timely manner by utilizing real-world data. Conducting RCTs to demonestrate the value of innovative interventions is complex and expensive, and, a well-structured observational research study using high-quality secondary databases could provide the same evidence but rather at a lower cost in a shorter time period.

2.2.2 Study objectives

The aim of this study is to evaluate whether the MTB has an impact on routine cancer care for patients with pancreatic cancer from health economic and outcomes perspectives

2.2.3 Materials and Methods

2.2.3.1 Study design and Data sources

This retrospective descriptive cohort study utilized the tumor documentation database CREDOS of the CCC-LMU site and hospital administrative claims data in the period from 2012 to 2021. CREDOS is the local documentation system used at the site and it allows compilation and tractability of most oncology-relevant data. Documented clinical data are driven by local certification requirements (e.g. onkozert) and following the federal law – the Bayrisches Krebsregistergesetz (https://www.gesetze-bayern.de/Content/Document/BayKRegG/true). The sites are legally required to document all information necessary to meet the data standards of the Associations of German Tumor Centers (Arbeitsgemeinschaft Deutscher Tumorzentren [ADT]; these data standards allow communication of information among different cancer centers and larger registries within Germany. The CREDOS database contains more than 1,000 attributes related to patient demographics, their medical histories and tumor characteristics. It contains data variables on patient ID, diagnoses information (i.e., ICD-10-GM code, date of diagnosis, tumor grading, histology, localization of diagnosis, tumor stage), procedure codes (diagnostic and therapeutic interventions), procedure start and end dates, vital status, vital dates and follow up information.

Hospital administrative claims data contain information about IDs, admission date, discharge date, admitting department and cost of management during each admission. The CREDOS data set was linked to the hospital administrative claims data from the medical department of the LMU hospital's medical department using unique patient IDs. Claims data allowed evaluation of hospital resource utilization and management costs including the sum of costs related to DRG services, fees for novel examinations and interventions and additional fees. In cases of missing information on disease stage, metastatic status, or therapy I performed

manual data extraction from patient electronic records, physicians' notes, and tumor board reports.

2.2.3.2 Study Sample

The study sample was composed of two groups of patients diagnosed with metastatic pancreatic cancer based on the International Classification of Diseases-10th revision (ICD-10 = C25.-). The first group included patients with cases of pancreatic cancer that were discussed by the MTB in the period 2016 – 2021; this group is therefore referred to as the MTB. The second group included pancreatic cancer patients who were not discussed by the MTB; it is referred to as the control group.

To be included in the study, patients had to 1) be 18 years of age or older and 2) be diagnosed with metastasized (UICC stage IV) pancreatic cancer in the period between a) 2016 – 2021 for the MTB group or b) 2012 – 2015 for the control group.

Patients diagnosed with 1) neuroendocrine tumors or, 2) other cancer diagnoses or syndromes or those who were 3) younger than 18, 4) not discussed in the MTB (for the MTB – group), 5) had missing information on metastatic status or date of metastasis, or 6) had missing information on therapy were excluded from the study.

2.2.3.3 Statistical Analysis

The CREDOS data set contains multiple data tables in .txt format. I imported the required data tables into SPSS and performed data cleaning, variable renaming, variable recoding, and category creation. I then identified patients with pancreatic cancer using the ICD-10 code of C25.-. The data variable "mol.TB" was used as a group identifier because it was given a value when the patient was discussed in the MTB. I renamed this variable to group and split it into two categories (1 = MTB and 2 = control).

Next, I evaluated the completeness of the data set in terms of information on the following variables: tumor stage, metastatic status, date of metastasis, drug names, and line of therapy. For patients with missing information for any of the aforementioned variables, I conducted manual data extraction from physicians' notes, tumor board reports, and pharmacy files from within patients' medical

records. I then applied the inclusion and exclusion criteria and excluded patients who were not eligible for the study.

To maintain the sample size and account for bias that could result from confounding factors such as age and gender, I applied inverse probability treatment weighting (IPTW) method [54]. I ran a logistic regression to calculate a propensity score (PSE) by including the confounding factors that I wanted to balance between the groups (age and gender) and used the group as an outcome variable. I then calculated the weight for the two groups as follows: weight_MTB = 1/PSE, and weight_control = 1/(1-PSE). I used these weights to run weighted descriptive analysis between the two groups. Categorical variables (e.g., age groups, line of therapy, therapy regimen and mortality status) were presented descriptively as counts and percentages. Continuous variables (e.g., age, time-metastatsis-to-MTB, time-metastasis-death, number of hospital admissions, length of hospital stays, costs) were presented as mean and (SDs).

In an additional step I merged the CREDOS data set with the hospital administrative claims data using patient IDs. For both groups, I restricted the analysis to patients with complete follow-up data for 6 months,12 months, and 18-months. I then evaluated hospitalization as number of hospital admissions and length of hospital stays for stays that lasted over 24 hours. Next, I calculated the total cost of management per patient for the two groups including the sum of costs related to DRG, costs of novel examination and interventions, and additional fees. All analyses were performed using IBM SPSS statistics 27 @copyright 2020.

2.2.4 Results

Initial inspection and evaluation of the CREDOS data set, administrative claims data and patients' medical records allowed evaluation of data variables availability in the three data sources (Table 5). In the CREDOS data set, 55.2% and 60.8% of the information on metastatic status and disease stage was missing. Therefore, data on missing information was manually extracted from physicians' notes, tumor board reports and pharmacy files from within the patients' medical records. The data variables available in the CREDOS data set and the patients' medical records (ICD-10 diagnosis code, date of diagnosis, metastatic status and disease stage) allowed identification of cases; and construction of follow-up period; and evaluation of demographic characteristics, clinical

characteristics, and management pattern in the study sample. A total of 140 MTB and 220 control metastatic pancreatic cancer patients were identified from the data set. As shown in Figure 3, 52 of the 140 MTB patients and 55 of the 220 control patients were excluded for one of the following reasons (multiple cancer diagnoses (MTB n =11, control n = 9), neuroendocrine tumor of the pancreas (MTB n= 16, control n = 17), no distance metastasis (MTB n = 4, control = 1), or insufficient clinical data (MTB n = 21, control n = 28).

Figure 3: Flowchart – study sample selection.

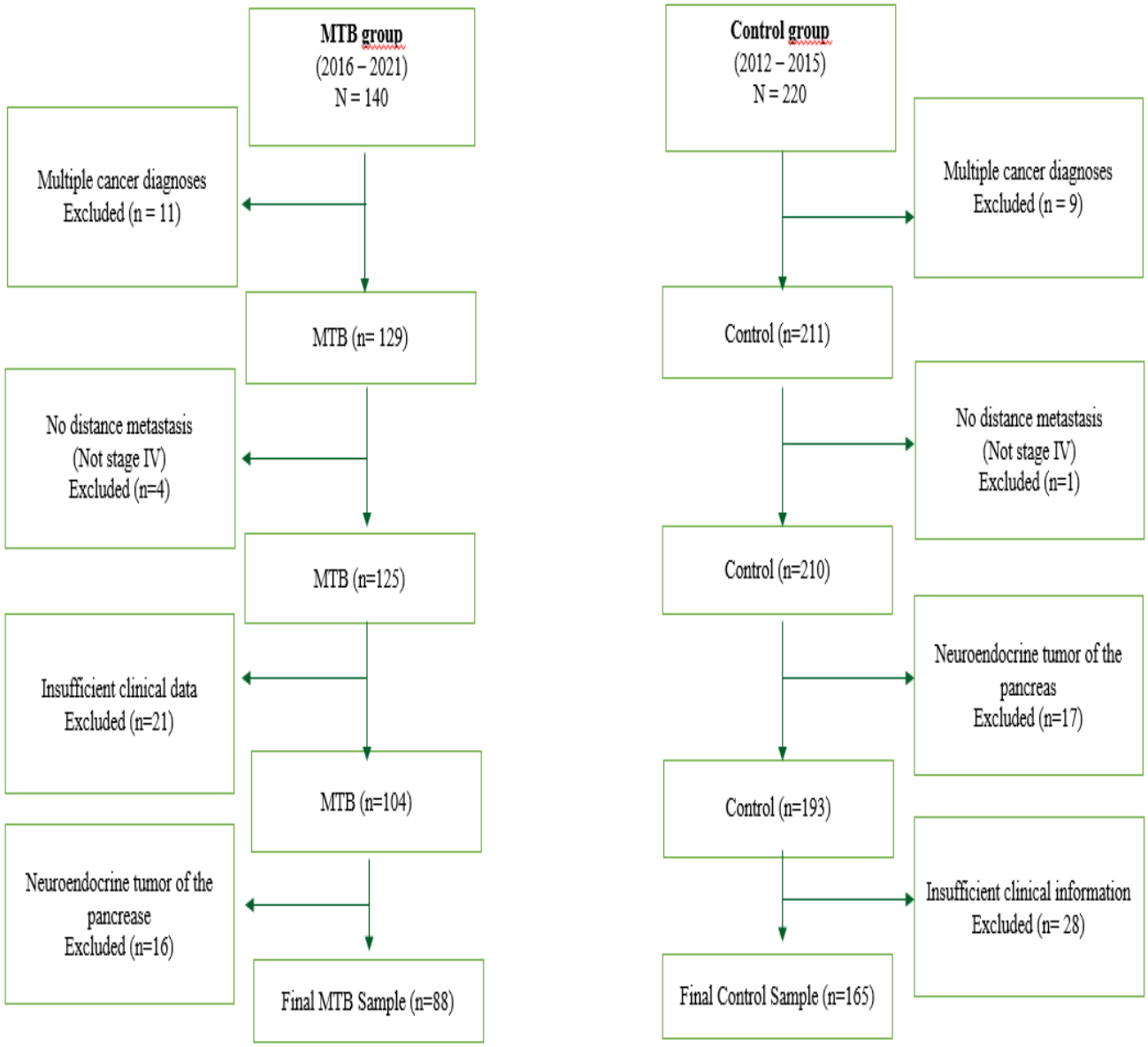

Figure 3 shows the flowchart for the study sample selection for the two groups (MTB and control).

Before I apply IPTW, there was a statistically significant difference between the two groups with in terms of age as a continuous variable (Table 6). Patients in the MTB group were slightly younger than those in the control group, with a mean age at diagnosis of 60.20 (SD 11.657) for the MTB group and 64.06 (SD 9.635) for the control group with a p-value of 0.005 for the complete study sample. The same applied to the cohorts with follow-up at 6 months,12 months and 18 months. However, after applying weight, there were no statistical difference between the two groups with regards to age (Table 7). Moreover, the distribution of age groups between the two groups in the three study cohorts did not differ significantly. Likewise, gender distribution between the two groups did not differ significantly (Table 6). Of participants in the MTB group, 59.1% were men; of those in the control group, 50.9% were men. Men made up 59.8% of the MTB group and 51.4% of the control group in the 6-month FU cohort, and 57.7% of the MTB and 52% of the control group in the 12-month FU cohort and 62.1% of the MTB and 61.7% of the control group in the 18-month FU study cohort. The median length of time from reported metastasis to discussion in the MTB for the MTB group was 3 months (SD 7.188) (Table 8).

Recorded vital status in both CREDOS and patient medical records enabled evaluation of mortality rate and calculation of the length of time from the onset of metastasis to reported death. Mortality was statistically significantly higher in the control group than in the MTB group in allstudy cohorts (p-value <0.001; Table 9).

Data on therapy before metastasis was incomplete; of the complete study cohort, data on line of therapy and therapy regimen (not shown) was available for only 15 MTB patients and 16 control patients. However, available data in both the CREDOS data set and patient medical records allowed construction of lines of therapy and evaluation of therapy regimens post-metastasis. There was no statistically significant difference between the two groups in the three study cohorts in terms of number of lines of therapy received after metastasis.

There was a statistically significant difference between the two groups for all three study cohorts in terms of therapy regimens administered as the first line of therapy after metastasis. As the first line of therapy after metastasis, gemcitabine – based therapy was frequently administered to patients in the MTB group while gemcitabine – monotherapy was frequently administered to patients in the control

group in the 6 months and 12 months study cohorts. (p-value <0.001) (Table 9). While Fluorouracil-based therapy was given as a first therapy post-metastasis more frequently to patients in the MTB group than thos in the control group in the 12-months and 18 months study cohorts, (P-value 0.035 – 0.043)

Table 5: Data availability in the secondary data sources from the university hospital.

	Administrative claims data	CREDOS	Patient medical records
Demographics			
Patients ID	✓	✓	✓
Age	✓	✓	✓
Gender	✓	✓	✓
Clinical characteristics			
Diagnosis code	✓	✓	✓
Date of diagnosis	×	✓	✓
Disease severity	×	insufficient info	✓
Disease stage	×	insufficient info (60.8%)	✓
Metastatic status	×	insufficient info (55.2%)	✓
Disease progression	×	×	✓
Therapy			
Drug name (OPS-code)	✓	✓	✓
Dose	×	×	✓
Start date of therapy	×	Missing info	✓
End date of therapy	×	Missing info	✓
Hospitalization			
Admission date	✓	×	×
Discharge date	✓	×	×
Admitting department	✓	✓	✓
Cost data	✓	×	×

Table 6: Demographics and clinical characteristics of patients with pancreatic cancer in the complete data set and the three study cohorts before IPTW.

	Complete data set			6 month FU			12 month FU			18 month FU		
	MTB N = 88	Control N = 165	No weight	MTB N = 87	Control N = 142	No weight	MTB N = 78	Control N = 125	No weight	MTB N = 29	Control N = 47	No weights
	N (%)	N (%)	P-value	N (%)	N (%)	P-value	N (%)	N (%)	P-value	N (%)	N (%)	P-value
Gender												
Men	52 (59.1%)	84 (50.9%)	0.214	52 (59.8%)	73 (51.4%)	0.217	45 (57.7%)	65 (52%)	0.428	18 (62,1%)	29 (61,7%)	0.974
Women	36 (40.9%)	81 (49.1%)		35 (40.2%)	69 (48.6%)		33 (42.3%)	60 (48%)		11 (37,9%)	18 (38,3%)	
Age groups												
30 - 49	14 (15.9%)	11 (6.7%)	0.097	13 (14.9%)	9 (6.3%)	0.113	12 (15.4%)	9 (7.2%)	0.162	6 (20,7%)	6 (12,8%)	0.723
50 - 59	25 (28.4%)	43 (26.1%)		25 (28.7%)	35 (24.6%)		23 (29.5%)	30 (24%)		7 (24,1%)	13 (27,7%)	
60 - 69	27 (30.7%)	60 (36.4%)		27 (31%)	51 (35.9%)		23 (29.5%)	44 (35.2%)		8 (27,6%)	11 (23,4%)	
>= 70	22 (25%)	51 (30.9%)		22 (25.3%)	47 (33.1%)		20 (25.6%)	42 (33.6%)		8 (27,6%)	17 (36,2%)	
Age												
Mean (SD)	60.20 (11.657)	64.06 (9.635)	0.005	60.56 (11.226)	64.56 (9.808)	0.005	60.4 (11.424)	64.64 (9.952)	0.006	60,28 (12,241)	63,30 (11,466)	0.280

Table 7: Demographics and clinical characteristics of patients with pancreatic cancer in the complete data set and the three study cohorts after applying IPTW.

	Complete data set			6 - months FU			12-months FU			18-months FU		
	MTB N = 88	Control N = 165	IPTW	MTB N = 87	Control N = 142	IPTW	MTB N = 78	Control N = 125	IPTW	MTB N = 29	Control N = 47	IPTW
	N (%)	(%)	P-value	(%)	(%)	P-value	(%)	(%)	P-value	(%)	N (%)	P-value
Gender												
Men	53.8%	54,5%	0.858	55%	55%	1.000	60%	60.9%	0.873	55.2%	54.5%	0.885
Women	46.2%	45,5%		45%	45%		40%	39.1%		44.8%	45.5%	
Age groups												
30 - 49	10.2%	9,1%	0.290	10%	8.7%	0.620	10.7%	10.6%	0.801	10.3%	10.4%	0.469
50 - 59	24.8%	29,1%		25.3%	27.9%		25.2%	26.1%		25.5%	26.7%	
60 - 69	30.7%	34,6%		31%	34.5%		31.4%	354%		29.9%	34.2%	
>= 70	34.3%	27,2%		33.6%	28.8%		32.7%	28%		34.3%	28.7%	
Age Mean (SD)	63.08 (11.080)	62,77 (9,887)	0.741	63.11 (10.802)	63.13 (10.122)	0.988	62.72 (11.008)	62.64 (10.033)	0.945	63.07 (10.933)	63.12 (10.285)	0.962

Table 8: Time from disease metastasis to MTB discussion in the MTB group (in months)

	Full data set					
Time from disease metastasis to MTB discussion	**N**	**Mean**	**Median**	**Minimum**	**Maximum**	**SD**
	88	6.05	3.00	0	38	7.188

Table 9: Mortality status and treatment pattern in the MTB and control groups

	6 months FU			12 month FU			18 month FU		
	MTB N = 87	**Control N = 142**	**IPTW**	**MTB N = 78**	**Control N = 125**	**IPTW**	**MTB N = 29**	**Control N = 47**	**IPTW**
	N (%)	**N (%)**	**P-value**	**N (%)**	**N (%)**	**P-value**	**N (%)**	**N (%)**	**P-value**
Vital status									
Alive	72 (83.4%)	50 (35.8%)	<0.001	63 (81.3%)	43 (35.5%)	<0.001	21 (73.7%)	18 (39.5%)	<0.001
Deceased	15 (16.6%)	92 (64.2%)		15 (18.7%)	82 (64.5%)		8 (26.3%)	29 (60.5%)	
Treatment Pattern									
a. Line of therapy after metastasis	**n = 82**	**n = 140**		**n = 74**	**n = 123**		**n = 28**	**n = 46**	
First line	82 (100%)	140 (100%)		74 (100%)	123 (100%)		28 (100%)	46 (100%)	
Second line	49 (63%)	65 (48.7%)	0.214	46 (61.7%)	59 (49.7%)	0.219	20 (71.2%)	32 (70.7%)	0.876
Third line	20 (22.6%)	27 (19.6%)		20 (25.4%)	26 (21.6%)		7 (24.7%)	14 (30.7%)	
Fourth line	8 (9.6%)	9 (6.7%)		8 (13.4%)	8 (7%)		4 (13.7%)	4 (9.3%)	

Fifth line	6 (7.2%)	2 (1.3%)		6 (7.3%)	2 (1.5%)				
b. First therapy after metastasis	**n = 82**	**n = 140**		**n = 74**	**n = 123**		**n = 28**	**n = 46**	
Fluorouracil – based therapy	46 (50.9%)	59 (44%)	0.146	43 (52.3%)	49 (41.7%)	0.035	19 (65.8%)	22 (49.3%)	0.043
Gemcitabine – based therapy	26 (36.9%)	26 (18.6%)	<0.001	21 (32.6%)	22 (17.6%)	<0.001	5 (17.8%)	7 (14.9%)	0.629
Gemcitabine -monotherapy	6 (8.9%)	48 (32%)	<0.001	6 (9.8%)	45 (34.5%)	<0.001	4 (16.2%)	14 (29.3%)	0.056
Others	4 (3.3%)	7 (5.3%)	0.289	4 (4.7%)	7 (6%)	0.548	0	3 (6.8%)	0.024
c. Second therapy after metas-tasis	**n = 49**	**n = 65**		**n = 46**	**n = 59**		**n = 20**	**n = 32**	
Fluorouracil – based therapy	16 (38%)	33 (50.5%)	0.053	14 (36.1%)	31 (52.5%)	0.015	5 (28.8%)	18 (54.7%)	0.007
Gemcitabine – based therapy	23 (42.2%)	13 (21.1%)	<0.001	22 (42.9%)	11 (20.2%)	<0.001	12 (57.7%)	8 (26.9%)	0.001
Gemcitabine -monotherapy	5 (11.7%)	8 (11.9%)	0.961	5 (11.8%)	7 (10.1%)	0.696	1 (5.8%)	2 (5.8%)	1.000
Others	5 (8.6%)	11 (16.7%)	0.060	5 (9.2%)	10 (17.2%)	0.077	2 (7.7%)	4 (13.2%)	0.526

Linking CREDOS data to the hospital administrative claims data allowed evaluation of hospitalization and costs of management in the study sample. Patients in the MTB group were admitted less frequently to hospital and had shorter hospital stays than those in the control group (Table 10). For the three study cohorts, the mean number of hospital admissions were 1.44 – 1.69 (SD 0.599– 1.214) for the MTB group and 2.39 – 2.81 (SD 1.730 – 1.965) for the control group (p-value <0.001). There was also a statistically significant difference between the two groups in terms of length of hospital stays. In all three study cohorts, patients in the MTB group spent a shorter period at the hospital than those in the control group (p-value <0.001). For the three study cohorts, the mean length of hospital stay was 7.96 – 9 days (SD 6.625– 7.987) for the MTB group and between 19.81 – 21.38 days (SD 17.028 – 18.557) for the control group.

Table 10: Hospitalization

	N	Mean	Median	Min – Max	SD	SE	SMD	IPTW P-value
Number of hospital admissions								
6 month FU								
MTB	43	1.67	1	1 – 7	1.165	0.110	-0.72111	<0.001
Control	86	2.39	2	1 – 12	1.882	0.161		
12 month FU								
MTB	37	1.69	1	1 – 7	1.214	0.123	-0.76196	<0.001
Control	75	2.45	2	1 – 12	1.965	0.179		
18 month FU								
MTB	15	1.44	1	1 – 3	0.599	0.95	-1.3741	<.001
Control	29	2.81	2	1 – 7	1.730	0.254		
Length of hospital stay (days)								
6 month FU								
MTB	43	9	7	1 – 32	7.514	0.708	-11.486914	<0.001
Control	86	20.49	16	1 – 101	18.557	1.585		
12 month FU								
MTB	37	8.39	7	1 – 32	6.625	0.672	-11.4153	<0.001
Control	75	19.81	16	1 – 85	17.028	1.553		
18 month FU								
MTB	15	7.96	6	1 – 32	7.987	1,261	-13.4156	<0.001
Control	29	21.38	17,21	2 – 85	17.162	3		

The total cost of management per patient (the sum of costs related to DRG, additional fees, and fees for novel examinations and interventions) was significantly higher for the patients in the control group than for those in the MTB group for all three cohorts (Table 11). For the three study cohorts, mean cost was between 5337 – 4958€ (SD 4775 – 6196€) for the MTB group and between 8439 - 9617€ (SD 5851 - 10654€) for the control group (p-vale <0.001).

Table 11: Total cost of management (in Euros).

	N	Mean	Median	Min – Max	SD	SE	SMD	IPTW P-value
6 month								
MTB	43	5337	3876	269 – 29191	5067	483	-427913808	<0.001
Control	86	9617	7189	265 – 74243	10654	910		
12 month FU								
MTB	37	4958	3808	269 – 29191	4775	491	-3898,02	<0.001
Control	75	8856	7095	265 - 57079	8433	769		
18 month FU								
MTB	15	4626	3419	269 – 29191	6196	979	-3812,46	0.004
Control	29	8439	7372	795 - 23934	5851	860		

3. Discussion

This dissertation aims to address the opportunities and challenges associated with using routinely collected data the LMU hospital in healthcare research to evaluate cancer care in routine care settings as well as the ability to link multiple secondary data sources to fill in information gaps. In the first part of the dissertation, a single secondary data source, hospital administrative claims data from the LMU hospital was used to evaluate the management of MM in routine care settings [41]. In the second part of the dissertation, multiple secondary data sources from the LMU-hospital were used to evaluate the impact of MTB on the routine cancer care of patients with pancreatic cancer. The data sources used for the second part were CREDOS data, hospital administrative claims data and patient medical records. To gather information missing from the CREDOS data set, I performed manual data extraction from physicians' notes, tumor board reports and pharmacy files from within the patients' medical records. The CREDOS data set was linked to the hospital administrative claims data using he unique patient IDs.

As demonstrated in chapter 1, it was possible to identify cases of MM and determine their basic demographic and clinical characteristics using hospital administrative claims data. Treatment pattern in terms of frequency of documentation of therapeutic interventions was also evaluated. Using the same data set, it was possible to evaluate healthcare resource utilization in terms of number of hospital admissions and length of hospital stays. Moreover, a subset of patients who underwent SCT was easily identified using predefined procedure codes for SCT (OPS codes 5-410, 5-411 and 8-860). In this subgroup, an index date from the date of the first SCT was set and; thereby, allowing for the chronological evaluation of events that happened post-SCT such as possible complications and additional treatments received after the transplantation from inpatient all-payers perspective. However, important data variable such as date of diagnosis was not recorded in the hospital administrative claims data set. This limited my ability to appropriately evaluate the disease progression and the chronological sequence of events (e.g., complications, or treatment side-effects) following the diagnosis. I was unable to differentiate between comorbid conditions and possible disease and treatment related complications. Moreover, setting a fixed follow up period from the disease onset onward was not possible for the same reason.

Additionally, data was limited to a single hospital department, and no data from other departments within the same hospital or from other health institutions were captured. Therefore, clinical findings and mortality reports based on this analysis cannot be generalized to all patients with MM.

In chapter 2, the clinical details available in the CREDOS data set and patients' medical records (e.g., date of diagnosis, disease stage, metastatic status and vital status) enabled the identification and construction of the study groups as well as the follow up periods for the sub-cohorts. Using these data sources, it was also feasible to construct lines of therapy and evaluate therapy regimens provided to the patients during hospitalization. Such an evaluation was not feasible in part 1. When the CREDOS data set was merged with hospital administrative claims data, it was possible to evaluate number of hospital admissions, length of hospitalization and cost of management during the study period. The study showed that MTB had a positive impact on the management of patients with pancreatic cancer. MTB patients had significantly lower mortality, number of hospital admissions, length of hospital stays, and lower costs of management compared to the control group. Although I have controlled for any significant bias caused by age difference between the two groups, the MTB patients are more likely to have had better performance status compared to the control group. Usually patients with good performance status are more likely to be enrolled in MTB discussions than frail patients. Thus, bias due to the difference in the performance status between the two groups could have accounted for the significant difference in terms of hospitalization, mortality and costs. The CREDOS data set was missing considerable information on disease stage, metastatic status, and therapy; therefore, missing data were extracted manually from physicians' notes, tumor board reports and pharmacy files from within the patients' medical records. However, 19 of the MTB patients and 28 of the control patients had insufficient clinical information and were excluded from the study. One possible explanation is that these patients were referred from other healthcare centers for a second opinion and therefore did not receive active treatment at the LMU hospital. Future study to comprehensively evaluate impact of baseline clinical characteristics and effect of performance status of patients with pancreatic cancer on the treatment decision from routine care settings is warranted.

The limitations I encountered when attempting to use hospital administrative claims data to adequately evaluate clinical characteristics and disease progression align with previous studies [5,20,55]. Additionally, missing clinical details in our hospital administrative data set agrees with previous reports including the study by Kreis K. et al. on the limitations of German claims data from statutory health insurance (SHI) [2]. Such a finding is unsurprising as the main intention of these data sources is billable purposes. However, our data set differs from SHI or private insurance claims data in the fact that, hospital administrative claims data provided better and granular information on the therapeutic and diagnostic interventions beyond the DRG system that are provided to patients during their inpatient stays regardless of their health insurance. By contrast, SHI and private insurance claims data provide more detailed information on all encounters between the insured individual and the healthcare system; they provide data on all reimbursed healthcare services from outpatient and emergency department visits [6]. Thus, the use of each type of claims data in scientific research is limited to invistigating specific research questions developed based on the variables for which data are available. This indicates that there is no single claims database in the German context that can provide generalizable, representative results on all individuals, and it indicates that the various available databases complement each other. Therefore, when integrated together (hospital administrative claims data and single-payer insurance claims data) or with other secondary data sources (e.g., disease registry, death registry, patient medical records or pharmacy files), they can provide a better and more comprehensive picture on any predefined topic from broader perspectives e.g., inpatient, outpatient and across the German healthcare system in general.

Our results on the limitations of administrative claims data in terms for adequately evaluating the incidence of adverse events are likewise consistent with results of previous studies [29-31]. However, the hospital administrative claims data set provided some quality indicators, such as infections, and readmissions among SCT patients (see chapter 1), signaling possible adverse events that require further evaluation. While I was not able to prove whether these complications were disease-related or treatment related in patients with MM, I was able to evaluate the impact of an intervention on patients with pancreatic cancer in Chapter 2. In chapter 2, I noted that patients in the control group had higher admission

rates and longer hospital stays, and higher costs of management than those in the MTB group. Thus, correlation between those indicators and the intervention can be assumed. This finding signals that MTB had a positive impact on patients with metastatic cancer as it reduced the number of hospital admissions, length of hospitalization and cost of management compared to the control group. Such indicators are also essential for evaluating the economic impact of the disease from a broader perspective. These indicators could serve as crucial elements for healthcare management evaluation within a healthcare facility or for benchmark evaluations (e.g., comparing guideline adherence or post-interventions complications) across different facilities. Fonseca et al. reported that multiple admissions and treatment- or disease-related complications of MM have some effect on disease financial burden [2]. A similar evaluation should be conducted in the German context with high-quality data to more broadly address the health outcomes and economic burden of MM and pancreatic cancer. This would not be possible without linking and merging multiple high-quality secondary data sources in order to provide generalizable and representative results.

Our findings on the age distribution of both cancers in the two study samples are consistent with published reports indicating that routinely collected data can provide representative results on demographic figures in the selected diseases [10,13,59,60]. Additionally, the data sources used enabled assessment of the integration of novel therapies (e.g., bortezomib, or lenalidomide) for the treatment of MM as well as the administration of fluorouracil-based therapy or gemcitabine (as a monotherapy or combined with other drugs) for the management of pancreatic cancer in the LMU hospital. These findings are consistent with previous reports on recent standard of care in the management of both MM and metastatic pancreatic cancer [30,61], further supporting the fact that the routinely collected data sources we used contain data required to answer certain scientific research questions on treatment pattern (therapy regimen, and lines of therapy).

To my knowledge, no previous evaluations of the CREDOS data set in healthcare studies to evaluate the care of patients with pancreatic cancer in routine care settings have yet been conducted. One possible explanation is due to patient privacy. Becuase routinely collected data sources from the LMU hospital have not yet been completely anonymized, their use for research is highly restricted, and they are accessible to only limited number of researchers. Therefore,

results of this study cannot be compared to established reports. In terms of data availability, comparison of the three secondary data sources used in this dissertation revealed that among the three data sources, there was no single source that contained all the required data elements needed to provide a comprehensive perspective on the routine cancer care of patients with MM or pancreatic cancer. It also showed that the three data sources complement one another and, when linked together are able to provide a comprehensive perspective on the routine care of cancer patients from an inpatient perspective. However, when future researchers seek to invistigare broader aspects of care, other secondary data sources should be identified and linked to existing data bases.

4. Conclusion

In conclusion, I aim in this dissertation to address the opportunities and challenges associated with using routinely collected data from the LMU hospital to evaluate cancer care in routine care settings and the ability to link data from multiple secondary data sources to fill information gaps. I created a list of data elements required to answer specific scientific research questions along with their corresponding secondardy data sources.

I find that hospital administrative claims data from the LMU hospital contain variables necessary for identification of cases, medical events, disease conditions, therapteutic and diagnostic interventions, and hospitalizations in patients with MM. Moreover, important quality-of-care indicatiors such as the frequency of infections and readmission rates are also recorded. This means that hospital administrative claims data are good data sources for assessments of quality-of-care within a healthcare institute or in benchmark evaluations across different healthcare facilites. However, dates of diagnosis, detailed clinical data on disease stage or severity, and response to treatment are missing. Therefore, hospital administrative claims data are limited for comprehensive evaluation of management patterns following the disease onset. Researchers aiming at evaluating management patterns in a particular disease entity should supplement these information from other data sources.

By contast, the CREDOS data set from the LMU hospital contain variables necessary for evaluation of disease onset, disease stage and severity, lines of therapy and mortality rates. When I linked the CREDOS data set to the hospital administrative claims data, I determined mortality rates, the number of hospital admissions, length of hospitalization and costs in patients with pancreatic cancer. I find that MTB may signal a positive impact on the management of patients with pancreatic cancer. Therefore, the usage of multiple secondary data sources allowed me to better evaluate the impact of the MTB on the management of patients with pancreatic cancer from routine care settings.

This dissertation provides a practical example on the utilization of secondary data from the LMU hospital for the evaluation of the routine care in two cancer entities (MM and pancreatic cancer). It addresses challenges each data source poses and opportunities for use in health services research. It can serve as a

blueprint for researchers aiming to employ secondary data sources in health services research to direct them through data variables identification and localization within the different available routine data. It can also help in hypothesis generation for future research studies aiming at evaluating routine cancer care in other different disease entities.

References

1. Fang Y, Wang H, He W. A Statistical Roadmap for Journey from Real-World Data to Real-World Evidence. Ther Innov Regul Sci [Internet]. 2020 [cited 2021 Mar 24];54:749–57. Available from: https://doi.org/10.1007/s43441-019-00008-2

2. Kreis K, Neubauer S, Klora M, et al. Status and perspectives of claims data analyses in Germany-A systematic review. Health Policy (New York). 2016;120(2):213–26.

3. Brant JM, Visovsky C, Wei SH, Wickham R, Wickham RJ. Secondary Analysis Research. AdvancedPractitioner.com [Internet]. [cited 2022 Jan 14];10. Available from: https://doi.org/10.6004/jadpro.2019.10.4.7

4. Cheng HG, Phillips MR. Secondary analysis of existing data: opportunities and implementation •Research methods in psychiatry• • 371 •. Heal [Shanghai Arch Psychiatry [Internet]. 2014 [cited 2018 Nov 8];26(6):371–5. Available from: http://dx.doi.org/10.11919/j.issn.1002-0829.214171]

5. Neubauer S, Kreis K, Klora M, et al. Access, use, and challenges of claims data analyses in Germany. Eur J Heal Econ. 2017;18(5):533–6.

6. Fraeman K, Nordstrom BL. Data Needs and Challenges in Cancer Epidemiology: A U.S. Real-World Database Perspective. 2016 [cited 2022 Jan 24]; Available from: https://www.evidera.com/wp-content/uploads/2016/05/Real-World-Data-Needs-and-Challenges-in-Cancer-Epidemiology.pdf

7. Palmaro A, Gauthier M, Despas F, Lapeyre-Mestre M. Identifying cancer drug regimens in French health insurance database: An application in multiple myeloma patients. Pharmacoepidemiol Drug Saf. 2017;26(12):1492–9.

8. Youn B, Trikalinos NA, Mor V, Wilson IB, Dahabreh IJ. Real-world use and survival outcomes of immune checkpoint inhibitors in older adults with non–small cell lung cancer. Cancer. 2020;

9. Ludwig H, Sonneveld P, Davies F, Blad´blad´ E J, Boccadoro M, Cavo M, et al. European Perspective on Multiple Myeloma Treatment Strategies in 2014. Oncologist [Internet]. 2014 [cited 2018 Oct 19];19:829–44. Available from: www.TheOncologist.com

10. Fonseca R, Abouzaid S, Bonafede M, et al. Trends in overall survival and costs of multiple myeloma, 2000-2014. Leukemia. 2017;31(9):1915–21.

11. Cowey CL, Liu FX, Boyd M, Aguilar KM, Krepler C, Boyd M, et al. Real-world treatment patterns and clinical outcomes among patients with advanced melanoma. 2019 [cited 2020 Feb 14]; Available from: http://dx.doi.org/10.1097/MD.0000000000016328

12. Maise EM, Evans KA, Chu B-C, et al. Temporal Trends in Survival and Healthcare Costs in Patients with Multiple Myeloma in the United States. Am Heal Drug Benefit. 2018;11(1):39–46.

13. Song X, Cong Z, Wilson K. Real-world treatment patterns, comorbidities, and disease-related complications in patients with multiple myeloma in the United States. Curr Med Res Opin. 2016;32(1):95–103.

14. Matsuno K, Ishihara R, Ohmori M, Iwagami H, Shichijyo S, Maekawa A, et al. Time trends in the incidence of esophageal adenocarcinoma, gastric adenocarcinoma, and superficial esophagogastric junction adenocarcinoma. J Gastroenterol [Internet]. 2019;54(9):784–91. Available from: https://doi.org/10.1007/s00535-019-01577-7

15. Teitelbaum A, Ba-Mancini A, Huang H, et al. Health Care Costs and Resource Utilization, Including Patient Burden, Associated With Novel-Agent-Based Treatment Versus Other Therapies for Multiple Myeloma: Findings Using Real-World Claims Data. Oncologist. 2013;

16. Ghate S, Ionescu-Ittu R, Burne R, Ndife B, Laliberté F, Nakasato A, et al. Patterns of treatment and BRAF testing with immune checkpoint inhibitors and targeted therapy in patients with metastatic melanoma presumed to be BRAF positive. Melanoma Res. 2019;29(3):301–10.

17. Griffiths RI, Gleeson ML, Mikhael J, Danese MD. Impact on medical cost, cumulative

survival, and cost-effectiveness of adding rituximab to first-line chemotherapy for follicular lymphoma in elderly patients: An observational cohort study based on SEER-medicare. J Cancer Epidemiol. 2012;2012.

18. Jagannath S, Roy A, Kish J, Lunacsek O, Globe D, Eaddy M, et al. Real-world treatment patterns and associated progression-free survival in relapsed/refractory multiple myeloma among US community oncology practices. Expert Rev Hematol [Internet]. 2016;9(7):707–17. Available from: http://dx.doi.org/10.1080/17474086.2016.1195254

19. Gansen FM. Health economic evaluations based on routine data in Germany: a systematic review. BMC Health Serv Res [Internet]. 2018 Dec 10;18(1):268. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-018-3080-3

20. Scheid C, Blau IW, Sellner L, Ratsch BA, Basic E. Changes in treatment landscape of relapsed or refractory multiple myeloma and their association with mortality: Insights from German claims database. Eur J Haematol. 2020;106:148–57.

21. Pobiruchin M, Bochum S, Martens UM, Kieser M, Schramm W. A method for using real world data in breast cancer modeling. J Biomed Inform [Internet]. 2016;60:385–94. Available from: http://dx.doi.org/10.1016/j.jbi.2016.01.017

22. Kuklik N, Stausberg J, Jöckel KH. Adverse drug events in German hospital routine data: A validation of International Classification of Diseases, 10th revision (ICD-10) diagnostic codes. PLoS One. 2017 Nov 1;12(11).

23. Reis A, Ihle P, Paulus U, Ferber L V., Diehl V, Walshe R. Cost of illness of malignant lymphoma in Germany. Eur J Cancer Care (Engl). 2006;15(4):379–85.

24. Ohlmeier C, Saum KU, Galetzka W, Beier D, Gothe H. Epidemiology and health care utilization of patients suffering from Huntington's disease in Germany: Real world evidence based on German claims data. BMC Neurol. 2019;19(1):1–8.

25. Hoffmann F. Review on use of German health insurance medication claims data for epidemiological research. Vol. 18, Pharmacoepidemiology and Drug Safety. 2009. p. 349–56.

26. Healthcare Financial Management Association. Costing healthcare in Germany. 2015; Available from: www.hfma.org.uk

27. Giersiepen K, Bachteler T, Gramlich T, Reiher J, Schubert B, Novopashenny I, et al. Zur Leistungsfähigkeit des Record-Linkage zwischen epidemiologischen Krebsregistern und dem Mammographie-Screening. Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz. 2010;53(7):740–7.

28. Maier B, Wagner K, Behrens S, Bruch L, Busse R, Schmidt D, et al. Comparing routine administrative data with registry data for assessing quality of hospital care in patients with myocardial infarction using deterministic record linkage. BMC Health Serv Res [Internet]. 2016 Oct 21 [cited 2022 Feb 23];16(1):1–9. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-016-1840-5

29. Datenlieferung gem. § 21 KHEntgG, InEK GmbH [Internet]. [cited 2020 Feb 13]. Available from: https://www.g-drg.de/Datenlieferung_gem._21_KHEntgG

30. Moreau P, San Miguel J, Sonneveld P, et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2017;28:iv52–61.

31. Turesson I, Velez R, Kristinsson SY, Al. E. Patterns of multiple myeloma during the past 5 decades: Stable incidence rates for all age groups in the population but rapidly changing age distribution in the clinic. Mayo Clin Proc. 2010;85(3):225–30.

32. Contributors MM. Haematology: Multiple Myeloma. Clin Med (Northfield Il). 2019;19(1):58–60.

33. Rajkumar SV. Multiple myeloma: Every year a new standard? 2019;

34. Gerecke C, Fuhrmann S, Strifler S, et al. The Diagnosis and Treatment of Multiple Myeloma. Dtsch Aerzteblatt Online. 2016;(1):470–7.

35. Centre for Cancer Registry Data at the Robert Koch Institute G. Multiple myeloma from "Cancer in Germany 2015/2016" [Internet]. [cited 2022 Jan 18]. Available from:

www.krebsdaten.de/cancer-sites

36. Pulte D, Jansen L, Castro FA, Al. E. Trends in survival of multiple myeloma patients in Germany and the United States in the first decade of the 21st century. Br J Haematol. 2015;171(2):189–96.

37. Denz U, Haas PS, Wä Sch R, et al. State of the art therapy in multiple myeloma and future perspectives. Eur J Cancer. 2006;42(11):1591–600.

38. Koleva D, Cortelazzo S, Toldo C, et al. Healthcare costs of multiple myeloma: An Italian study. Eur J Cancer Care (Engl). 2011;20(3):330–6.

39. Moehler TM, Merz M, Kellermann L, Goldschmidt H, Knauf W. Diagnostic and therapeutic approaches to multiple myeloma patients: 'Real-world' data from representative multicentre treatment surveys in Germany between 2008 and 2011. Oncol Lett. 2016;12(6):5043–51.

40. Yong K, Delforge M, Driessen C, et al. Multiple myeloma: patient outcomes in real-world practice. Br J Haematol. 2016;175(2):252–64.

41. AlZahmi A, Cenzer I, Mansmann U, Ostermann H, Theurich S, Schleinkofer T, et al. Usability of German hospital administrative claims data for healthcare research: General assessment and use case of multiple myeloma in Munich university hospital in 2015-2017. PLoS One [Internet]. 2022;17(7 July):1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0271754

42. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018 Jul 1;57(S 01):e50–6.

43. Nasseh D, Schneiderbauer S, Lange M, Schweizer D, Heinemann V, Belka C, et al. Optimizing the Analytical Value of Oncology-Related Data Based on an In-Memory Analysis Layer: Development and Assessment of the Munich Online Comprehensive Cancer Analysis Platform. J Med Internet Res. 2020;22(4).

44. Ferreira CG, Nicolini A, Dalurzo L, Stefani S, Teich V, Leighl N. The Value of Biomarkers in Optimizing the Use of Immuno-oncologic Therapy. Curr Drug Targets. 2018;20(1):81–6.

45. Garattini L, Curto A, Freemantle N. Personalized medicine and economic evaluation in oncology: All theory and no practice? Expert Rev Pharmacoeconomics Outcomes Res. 2015;15(5):733–8.

46. Yang H Te, Shah RH, Tegay D, Onel K. Precision oncology: Lessons learned and challenges for the future. Cancer Manag Res. 2019;11:7525–36.

47. Akhmetov I, Bubnov R V. Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. EPMA J [Internet]. 2015;6(1):1–12. Available from: http://dx.doi.org/10.1186/s13167-015-0041-3

48. Parkinson DR, McCormack RT, Keating SM, Gutman SI, Hamilton SR, Mansfield EA, et al. Evidence of clinical utility:An unmet need in molecular diagnostics for patients with cancer. Clin Cancer Res. 2014;20(6):1428–44.

49. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Vol. 18, Nature Reviews Gastroenterology and Hepatology. Nature Research; 2021. p. 493–502.

50. McGuigan A, Kelly P, Turkington RC, Jones C, Coleman HG, McCain RS. Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. World J Gastroenterol. 2018;24(43):4846–61.

51. Tempero MA. NCCN Guidelines Updates: Pancreatic Cancer. J Natl Compr Canc Netw. 2019;17(55):603–5.

52. Ducreux M, Cuhna AS, Caramella C, Hollebecque A, Burtin P, Goéré D, et al. Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol [Internet]. 2015;26(Supplement 5):v56–68. Available from: http://dx.doi.org/10.1093/annonc/mdv295

53. Robert Koch Institute. Pancreatic cancer [Internet]. 2022 [cited 2022 Apr 6]. Available

from:
https://www.krebsdaten.de/Krebs/EN/Content/Cancer_sites/Pancreatic_cancer/pancreatic_cancer_node.html

54. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, et al. An introduction to inverse probability of treatment weighting in observational research. Clin Kidney J. 2022;15(1):14–20.

55. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity Measures for Use with Administrative Data. Med Care. 1998;36(1):8–27.

56. Lovaglio PG. Benchmarking strategies for measuring the quality of healthcare: Problems and prospects. Sci World J. 2012;2012:606154.

57. Hohl CM, Kuramoto L, et al. Evaluating adverse drug event reporting in administrative data from emergency departments: A validation study. BMC Health Serv Res. 2013;13(1):473.

58. Ackroyd-Stolarz S, Bowles SK, Giffin L. Validating administrative data for the detection of adverse events in older hospitalized patients. Drug Healthc Patient Saf. 2014;6(1):101–7.

59. Ansari D, Tingstedt B, Andersson R. Pancreatic cancer-cost for overtreatment with gemcitabine. Acta Oncol (Madr) [Internet]. 2012; Available from: https://www.tandfonline.com/action/journalInformation?journalCode=ionc20

60. Ding D, Javed AA, Cunningham D, Teinor J, Wright M, Javed ZN, et al. Challenges of the current precision medicine approach for pancreatic cancer: A single institution experience between 2013 and 2017. Cancer Lett [Internet]. 2021;497:221–8. Available from: https://doi.org/10.1016/j.canlet.2020.10.039

61. Neoptolemos JP, Kleeff J, Michl P, Costello E, Greenhalf W, Palmer DH. Therapeutic developments in pancreatic cancer: Current and future perspectives. Nat Rev Gastroenterol Hepatol [Internet]. 2018;15(6):333–48. Available from: http://dx.doi.org/10.1038/s41575-018-0005-x

Appendix A:

Table : Data elements required to answer questions related to multiple myeloma.

Aspect	Question	Required variables	Additional details (Tests or ICD/OPS code)	Possible source of information
Demographics	1. How can the patients be identified?	Patient's unique identifier	-	PMC EHR CD
	2. What is the age distribution for the identified group of patients?	Age/	Date of birth	PMC EHR CD
	3. What is the sex distribution of the disease?	Sex	-	PMC EHR CD
	4. In which departments were the patients admitted?	Admitting department	- - -	PMC EHR CD
Epidemiology and Clinical characteristics	1. How can multiple myeloma be identified?	I. Documented diagnosis of multiple myeloma	C90.00 C90.01 C90.00+ C90.01+	PMC EHR CD
	2. How the diagnosis of multiple myeloma be confirmed? (other possible criteria needed for identifying the disease)	I. Laboratory tests	1. Serum or urinary protein electrophoresis (8-82). 2. Nephelometric quantification of immunoglobulins. 3. Immunofixation. 4. Bone marrow biopsies/aspiration for measurement of plasma level (1-424, 1-941). 5. Serum FLC level. 6. Complete blood count with differential serum creatinine, creatinine clearance, and calcium level.	PMC EHR
		II. Radiologic tests	1. WBLD-CT (3-20 – 3-26) 2. MRI (3-80 – 3-84)	PMC HER

			3. PET-CT (3-75) 4. Conventional radiographs.	
	3. How can disease stage of multiple myeloma patients be assessed?	I. International staging system (ISS) for multiple myeloma.	1. Serum β2M level. 2. Serum Albumin.	PMC EHR
	4. How can risk be assessed in the identified group of patients?	I. ISS II. Chromosomal abnormality detection (iFISH). III. LDH level.	Cytogenetic testing (1-991 – 1-999)	PMC EHR
	5. What are the most common comorbid conditions present in this disease group?	**Concurrent comorbidities:**		PMC EHR CD
		Hypertension	I10 – I15	
		Congestive heart failure	I50	
		Cerebrovascular disease	I60 – I69	
		Peripheral vascular disease	I70 – I79	
		Myocardial infarction	I20 – I24	
		Chronic kidney disease	N18	
		Primary tumor	C00 – C97 D00 – D48	
	6. What are the most commonly reported disease-related and/or treatment-related complications in the identified disease group?	**Disease or treatment complications:**		PMC EHR
		Bone lesions	M82.0	
		Multiple myeloma-related renal disease	N16.1*	
		Multiple myeloma-related glomerular disease	N08.1*	
		Neoplasm-induced anemia	D63.0*	
		Thrombocytopenia	D69.5, D69.6	
		Neutropenia	D70.	
		Hypercalcemia	E83.5	
		Thromboembolic event	I82	
		Gastrointestinal bleeding	K92.	
		Cerebral hemorrhage	I60. – I62.	
		Treatment complication	Y57, Y69, Y84	

		Death	R96, R98, R99	PMC EHR CD
Treatment pattern	1. What medications are used as a front-line therapy in the disease group?	Bortezomib	6-001.9	PMC EHR CD(-) Pharmacy files
		Melphalan	NA	
		Prednisone	NA	
		Dexamethasone	NA	
		Cyclophosphamide	NA	
		Carfilzomib	6-008.9	
		Thalidomide	NA	
		Lenalidomide	6-003.g	
		Pomalidomide	6-007.a	
		Combination chemotherapy	8-542 8-543 8-544 8-547	
		Panobinostat	6-009.2	
		Elotuzumab	6-009.d	
		Ixazomib	NA	
		Daratumumab	6-009.a	
		Stem cell transplantation	5-410 5-411 8-860	
		Transfusion of hematopoietic stem cell	8-805	
	2. What medications are used to treat relapsed/refractory multiple myeloma in the identified group?	Medications listed above		PMC EHR CD(-) Pharmacy files
	3. What medications are used for consolidation therapy in the identified group?	Medications listed above		PMC EHR CD(-) Pharmacy files
	4. What medications are used for maintenance therapy in the identified disease group?	Medications listed above		PMC EHR CD(-)

				Pharmacy files
	5. What medications are used for supportive care in the disease group?	Anti-fungal medications	6-002.5 6-002.p 6-002.q 6-002.r 6-003.1 6-004.5	PMC EHR CD(-) Pharmacy files
		Blood product transfusion	8-800	
		Thrombocyte transfusion	8-800.g	
		Hemodialysis	8-854	
		Therapeutic plasmapheresis	8-820	
		Sedation	8-903	
		Lipegfilgrastim	6-007.7	
		High voltage radiotherapy	8-522	
		Pain therapy	8-91	
	Further required information:	Start date of treatment	-	PMC EHR Pharmacy files
		End date of treatment	-	
		Dose of medication	-	
		Route of administration	-	
Health service utilization	1. How frequently were multiple myeloma patients admitted to the hospital?	Date of admission		PMC EHR CD
	2. How long (on average) did multiple myeloma patients stay at the hospital?	Date of discharge		PMC EHR CD
	3. To which hospital departments were the identified group of patients mostly admitted?	Admitting department		PMC EHR CD
		I. Laboratory		PMC

	4. What health services were mostly used by the identified group of patients?	II. Radiological III. Pharmacy IV. ICU V. Operation		EHR CD
	Further required information:	Reason for admission reason		PCD EHR
		Reason for discharge		PMC EHR

PMC = Patient medical chart; EHR = Electronic hospital record; CD=Claims data; FLC= Free-light chain measurement.; WBLD-CT= Whole body low dose-computed tomography; MRI = Magnetic resonance Imaging; PET-CT =Positron emission tomography with CT; β2M = β2 microglobulin; LDH=Lactate-dehydrogenase; iFISH= inter-phase fluorescent in situ hybridization; NA = not available; ICU: Intensive care unit; ICD = International Classification of Diseases; OPS = Procedure codes

Acknowledgements

I would like to thank my supervisors; Prof. Dr. Med. Helmut Ostermann, Prof. Ulrich Mansmann and Dr. PD. Karin Berger for their supervision and support during my PhD studies. Dr. Karin Berger, thank you for providing me with the opportunity to work on the MM and MTB projects and all the support and guidance thoughout these two projects.

Irena Cenzer, thank you very much for the support and guidance with real-world data analysis and the statistical support you provided me during my PhD.

I would also like to thank my sponsoring university; the United Arab Emirates University for providing me with the opportunity, financial and all possible support to pursue my postgraduate studies at the Ludwig Maximilians University.

I would like to acknowledge the editing service I received from Scribbr for proofreading and editing my dissertation. Thank you to all who pariticipated in the two projects either with clinical input or data collection to make them successful.

Affidavit

Affidavit

AlZahmi, Amal

Surname, first name

Street

Zip code, town

Country

I hereby declare, that the submitted thesis entitled

German secondary data: Opportunities and challenges for assessing routine can-cer care from a health institution's perspective

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

14.12.2022, Dubai

Place, date

Amal AlZahmi

Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office
Medical Faculty

Confirmation of congruency between printed and electronic version of the doctoral thesis

AlZahmi, Amal

Surname, first name

Street

Zip code, town

Country

I hereby declare that the electronic version of the submitted thesis, entitled

German secondary data: Opportunities and challenges for assessing routine can-cer care from a health institution's perspective

is congruent with the printed version both in content and format.

29.05.2022, Dubai

Place, date

Amal AlZahmi

Signature doctoral candidate

List of publications

1. AlZahmi A, Cenzer I, Mansmann U, Ostermann H, Theurich S, Schleinkofer T, et al. Usability of German hospital administrative claims data for healthcare research: General assessment and use case of multiple myeloma in Munich university hospital in 2015-2017. PLoS One [Internet]. 2022;17(7 July):1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0271754

2. Abstract accepted for poster presentation at the DGGO annual meeting-March. 2020: Use of German hospital claims data in research studies- Multiple Myeloma case study.