Aus dem Institut für Gesundheitsökonomie und
Management im Gesundheitswesen
Helmholtz Zentrum München -
Deutsches Forschungszentrum für Gesundheit und Umwelt
Leiter: Prof. Dr. Reiner Leidl

External validation of decision-analytic models based on claims data of health insurance funds

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Alina Christa Annemarie Brandes

aus
München

2016

Mit Genehmigung der Medizinischen Fakultät

der Universität München

Berichterstatter:	Priv. Doz. Dr. Wolf Rogowski
Mitberichterstatter:	Prof. Dr. Konstantin Strauch
	Prof. Dr. Klaus A. Kuhn
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	16.03.2016

To

Trixi, Rudi, Theresa, and most of all Daniel

~

Without their support I wouldn't have come this far

Abstract

Background: Decision-analytic models are used in the context of economic evaluation to bring together the best available evidence and to support the decision on the adoption of a health technology. A decision model's credibility is, however, diminished by uncertainty which, to large part, stems from parameter uncertainty. Especially when novel technologies are evaluated, high quality evidence may not be available at the point of coverage decision making. A decision model incorporating uncertain parameter values eventually simulates uncertain effectiveness and cost outcomes.
To enhance credibility of decision models, external validation of uncertain parameter values is vital. Data sources for external validation should be able to reflect the model's study design and patient cohort, and estimate real-world effectiveness and costs.

Objective: This study assesses whether claims data of health insurance funds are suitable to externally validate decision-analytic models.

Methods: To answer the research question, a validation approach is developed which highlights critical steps in the validation of decision models based on claims data. The validation steps are: 1) selection of the validation level, 2) selection of the claims dataset, study design, and patient cohort, 3) selection of disease-relevant health technologies and costs, 4) statistical analysis of claims data, 5) changes to the decision model, 6) comparison between model and claims data, and 7) sensitivity analyses.
The validation approach is exemplarily applied in the validation of a Markov model comparing treatment of localized prostate cancer (active surveillance and radical prostatectomy) in a German health care context, based on claims data of the German AOK statutory health insurance fund. An external validation of resource use, probability of utilization, and cost parameters is chosen, because these parameters are afflicted by a high degree of uncertainty in the decision model.
Two different approaches to the analysis of relevant health technologies for prostate cancer treatment are presented in claims data analysis: an excess approach and a disease-related approach.

Results: The decision model assumes that resource use and unit costs are identical in the two treatment groups; this is, however, not observed in claims data analysis.

Excess cost analysis and disease-related cost analysis of AOK claims data as well as model analysis show that, overall, active surveillance is the less costly strategy compared to radical prostatectomy, with total incremental costs of €-6,611, €-6,260, and €-7,486 respectively. When testing differences between model and outcomes of claims data analysis, p-values of 0.61 (excess approach) and 0.18 (disease-related approach) indicate an agreement that is sufficient to assume that the decision model simulates real-world costs validly.

Discussion: This study reveals general strengths and limitations of claims data based model validation.

Claims data are able to provide evidence on real-world resource utilization and, with limitations regarding clinical information, effectiveness of a wide range of indications and treatments for a large patient cohort. Validation based on claims data is especially suitable when the decision maker, interested in the validity of the model in question, is the insurance fund providing access to the claims data.

Suitability of claims data based validation is, however, limited concerning the replication of decision models' structure and patient cohort. For one, the identification of distinct health states is limited, because clinical information is not included in sufficient detail. Secondly, due to non-randomization and a restricted number of variables available to adjust for confounding, comparability of treatment groups is limited in claims data analysis. Thirdly, distinct identification of health technology utilization and corresponding costs is not possible if the technology of interest is not specifically coded. Finally, claims data are, generally, collected for billing purposes; diagnoses and technology utilization are only coded if they are relevant for reimbursement by the insurance fund, which biases outcomes of model validation in cases where treatment is not covered by the insurance fund.

Conclusion: The presented validation approach indicates critical aspects of the validation based on claims data, which may support researchers and decision makers in their decision on the suitability of claims data for model validation.

The suitability of claims data for the external validation of a decision model ultimately depends on the ability of the claims data source to reflect the model's patient cohort and outcome measures.

Zusammenfassung

Hintergrund: Entscheidungsanalytische Modelle kommen im Rahmen der gesundheitsökonomischen Evaluation von Gesundheitstechnologien zum Einsatz, um die beste verfügbare Evidenz zusammenzuführen und damit die Erstattungsentscheidung zu unterstützen. Bei der Evaluation von innovativen Technologien ist allerdings häufig zum Zeitpunkt der Erstattungsentscheidung keine hochwertige Evidenz, etwas aus klinischen Studien, verfügbar. Diese Parameterunsicherheit spiegelt sich letztlich in der im Entscheidungsmodell simulierten Kosteneffektivität der jeweiligen innovativen Technologien wieder. Für den Entscheidungsträger ist somit die Glaubwürdigkeit von Entscheidungsmodellen eingeschränkt.
Um die Glaubwürdigkeit eines Entscheidungsmodells zu erhöhen, ist eine externe Validierung der unsicheren Parameterwerte von entscheidender Bedeutung. Datenquellen für eine externe Validierung sollten in der Lage sein, das Studiendesign und die Kohorte des Entscheidungsmodells zu reflektieren sowie reale Effekte und Kosten der evaluierten Technologie zu schätzen.

Fragestellung: Im Rahmen dieser Studie wird untersucht, in wie weit sich Abrechnungsdaten von Krankenkassen für die externe Validierung von entscheidungsanalytischen Modellen eignen.

Methoden: Um die Forschungsfrage zu beantworten, wurde ein Validierungsansatz entwickelt, welcher entscheidende Schritte bei der Validierung von Entscheidungsmodellen auf der Basis von Abrechnungsdaten beschreibt. Die einzelnen Validierungsschritte sind: 1) Auswahl der Validierungsebene, 2) Auswahl des externen Datensatzes, des Studiendesigns und der Patientenkohorte, 3) Definition von krankheitsrelevanten Gesundheitstechnologien und Kosten, 4) Auswahl der statistischen Methoden zur Analyse der Abrechnungsdaten, 5) Anpassung des Entscheidungsmodells, 6) Auswahl von Methoden zum Vergleich zwischen Modell und Abrechnungsdaten, und 7) Sensitivitätsanalysen.
Der Validierungsansatz wird beispielhaft für die Validierung eines Markov-Modells angewendet, welches Behandlungsmethoden des lokalisierten Prostatakarzinoms (Active Surveillance und radikale Prostatektomie) in einem deutschen Versorgungskontext vergleicht. Zur Validierung werden Abrechnungsdaten einer deutschen gesetzlichen

Krankenkasse, der AOK Baden-Württemberg, herangezogen. Es werden Parameterwerte des Entscheidungsmodells zum Ressourcenverbrauch, zur Inanspruchnahmewahrscheinlichkeit und zu Kosten validiert, da diese Parameter die größte Unsicherheit aufweisen. Dabei werden zwei verschiedene Vorgehensweisen zur Analyse der Abrechnungsdaten der AOK herangezogen: ein Excesskosten-Ansatz und ein Krankheitskosten-Ansatz.

Ergebnisse: Im Entscheidungsmodell wird davon ausgegangen, dass Ressourcenverbrauch und Stückkosten in beiden Behandlungsgruppen identisch sind; in den Abrechnungsdaten der AOK ist diese Annahme allerdings nicht wiederzufinden.
Sowohl die Excesskosten-Analyse und die krankheitskostenbezogene Analyse der AOK-Daten als auch die Modellanalyse zeigen, dass Active Surveillance insgesamt die kostengünstigere Strategie mit einer Ersparnis von jeweils 6.611€, 6.260€ und 7.486€ gegenüber der radikalen Prostatektomie ist. Der statistische Test der Kostendifferenz aus Modell und AOK-Daten ergibt p-Werte von 0,61 (Excesskosten-Ansatz) und 0,18 (Krankheitskosten-Ansatz), die auf eine signifikante Übereinstimmung der Schätzer aus Modell und AOK-Daten schließen lassen. Die Übereinstimmung der Schätzer lässt vermuten, dass das Entscheidungsmodell in der Lage ist, die Kosten der Behandlung des lokalisierten Prostatakarzinoms valide zu simulieren.

Diskussion: Die beispielhafte Validierung des Markov-Modells anhand von Abrechnungsdaten der AOK Baden-Württemberg zeigt allgemeine Stärken und Schwächen der Kassendaten-basierten Modellvalidierung auf.
Abrechnungsdaten sind in der Lage, Evidenz zur tatsächlichen Utilisierung von Gesundheitsleistungen und, mit Einschränkungen in Bezug auf klinische Informationen, Wirksamkeit einer Vielzahl von Behandlungsoptionen für eine große Patientenpopulation zu liefern. Die Validierung auf Basis von Abrechnungsdaten ist vor allem sinnvoll, wenn die Modellvalidierung aus der Perspektive einer Krankenkasse durchgeführt werden soll.
Die Eignung von Abrechnungsdaten für die Modellvalidierung ist jedoch hinsichtlich der Nachbildung der Modellstruktur und der Patientenkohorte des Entscheidungsmodells limitiert. Erstens ist die Identifikation von Gesundheitszuständen in Kassendaten begrenzt, da klinische Informationen nicht ausreichend detailliert enthalten sind. Zweitens ist die Vergleichbarkeit der Behandlungsgruppen eingeschränkt, da eine Randomisierung nicht möglich ist und nur eine begrenzte Anzahl an Variablen zur Verfügung steht, um für

Confounder zu adjustieren. Drittens ist eine eindeutige Identifizierung von Gesundheitsleistungen und deren Kosten schwierig, wenn die Leistung nicht explizit in den Abrechnungsdaten kodiert ist. Viertens werden Kassendaten zu Abrechnungszwecken gesammelt und deshalb werden auch nur solche Diagnosen und Gesundheitsleistungen kodiert, die für die Erstattung durch die Krankenkasse relevant sind. Für Gesundheitsleistungen, die nicht von der Krankenkasse vergütet werden, ist unter Umständen keine valide Schätzung zu Ressourcenverbrauch und Kosten möglich.

Fazit: Der entwickelte Validierungsansatz zeigt kritische Aspekte der Modellvalidierung auf Basis von Abrechnungsdaten von Krankenkassen auf. Er soll Wissenschaftler und Entscheidungsträger bei der Entscheidung über die Eignung von Abrechnungsdaten für die externe Validierung eines Modells unterstützen.

Die Eignung von Abrechnungsdaten für die externe Validierung eines Entscheidungsmodells hängt letztlich von der Fähigkeit ab, Modellstruktur, Kohorte und Zielparameter des Modells abzubilden.

Table of contents

List of tables

List of figures

Abbreviations

ADT	Androgen-deprivation therapy
AOK	Allgemeine Ortskrankenkasse [Local health insurance fund]
AS	Active surveillance
ATC	Anatomical Therapeutic Chemical Classification System
BPH	Benign prostate hyperplasia
CCS	Charlson Co-morbidity Score
CM	Conservative management
DAK	Deutsche Angestellten-Krankenkasse [German health insurance fund for employees]
DGU	Deutsche Gesellschaft für Urologie [German Association of Urology]
DRG	Diagnosis Related Group
EAU	European Association of Urology
EBM	Einheitlicher Bewertungsmaßstab [Uniform Value Scale]
EBRT	External-beam radiotherapy
ED	Erectile dysfunction
EU	European Union
GP	General practitioner
H0	Null hypothesis
H1	Alternative hypothesis
HR	Hazard ratio
IC	Urinary incontinence
ICD-10 GM	German application of the International Classification of Disease, Version 10
ICER	Incremental cost-effectiveness ratio

IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen [Institute for Quality and Efficiency in Health Care]
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MEA	Managed entry agreement
OPS	Operationen- und Prozedurenschlüssel [inpatient operation and procedure codes]
PCa	Prostate cancer
PSA	Prostate-specific antigen
PZN	Pharmazentralnummer [uniform pharmaceutical identification key]
QALY	Quality-adjusted life year
RCT	Randomized controlled trial
RP	Radical prostatectomy
SEER	Surveillance, Epidemiology, and End Results
SGB V	Social Code Book V
SHI	Statutory Health Insurance
TNM	Tumor Node Metastasis
TURP	Transurethral resection of the prostate
UK	United Kingdom
US	United States
WW	Watchful waiting

1 Background and objective

1.1 Decision-analytic modeling

1.1.1 Modeling in health economic evaluations

Economic evaluation in health care is defined as the comparison of alternative health technologies regarding their costs and consequences (1). The term 'health technologies' refers to a range of health care resources including medical devices, pharmaceuticals, procedures, organizational support systems, screening, and health promotion programs (2).

The most frequently used types of economic evaluation include cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis (1). In cost-effectiveness analysis consequences are measured in natural units, for example the life years gained by an intervention. The alternative technologies are then compared by the calculation of an incremental cost-effectiveness ratio (ICER). Cost-utility analysis allows comparison of interventions with different health outcomes by a single parameter – the utility. The standard measure for utility in health economics are quality-adjusted life years (QALY). QALY allows incorporation of life time and quality of life effects in a single outcome measure. Alternatives are, again, compared by the ICER (costs per QALY). Cost-benefit analysis measures consequences in monetary units. The decision to implement a health technology is positive if its monetary benefit exceeds the costs (1).

Economic evaluation is used to inform decision makers about which health technologies to fund from available resources. This requires an incorporation of all adequate evidence in the evaluation to compare new technologies with all relevant alternative technologies. To come to a valid decision, it is also necessary to reveal uncertainty in evidence which can be addressed in future research (1). Randomized controlled trials (RCT) or observational studies can hardly accomplish these requirements independently, which provides a strong rationale for decision-analytic modeling as a framework for economic evaluation.

Models bring together the best available evidence and systematically address uncertainty in threshold and sensitivity analyses (2). Decision-analytic models also allow evaluation of cost-effectiveness of novel health technologies at an early stage of the development process where evidence from clinical trials is not available yet (3).

The two most frequently used types of decision-analytic models are decision-trees and Markov models. A decision-tree is the simplest form of a decision-analytic model. It is the visual representation of all possible options and the consequences following these. However, in case of indications with recurring events and lifetime analysis – such as prostate cancer – decision-trees become very complex. Markov models address this complexity by modeling events as transitions between defined health states (Markov states) in defined time intervals (cycles). Markov models take dates of events into account; time-dependent event probabilities as well as costs can be modeled (4).

1.1.2 Evidence sources

The best available evidence incorporated in a decision model usually stems from a variety of evidence sources and is used with different purposes within the model. Primary and secondary data sources can be incorporated in a decision-analytic model.

In the design and specification of the model, evidence is used which describes the epidemiology of the underlying indication and its clinical care as well as clinical outcomes and resource use. Information on this is drawn from a range of evidence sources including RCTs, clinical guidelines, and administrative data sources (5). Evidence on clinical outcome is often found in RCTs and meta-analysis. In addition to primary evidence sources, national reimbursement catalogues (e.g. the German diagnosis related group (DRG) catalogue for inpatient costs) are frequently used to inform cost parameters (5, 6).

The relevance of administrative data and reference sources, such as reimbursement catalogues, is often specific to the decision-making context, concerning geographical or reimbursement process features (5). Expert opinion is a common method used in the population of model parameters if no other evidence source is available. Expert estimates derived from formal methods such as Delphi or Nominal Group techniques are preferable to non-formal methods (6, 7). Reasonable effort to obtain new additional data prior to modeling should be considered. However, cost and delay in obtaining the data must be weighed against the benefit of reduction in uncertainty (7).

1.1.3 Uncertainty

Combining evidence on alternative health technologies from different evidence sources in decision models is a process that is inherently uncertain, because there will hardly ever be

complete information on all the possible cost and consequences of a technology in a given population (8). Especially at the time of introduction of a novel health technology, evidence which is relevant for the design of the model, such as treatment practice, patients' compliance, as well as effectiveness and cost outcome, is usually scarce (9).

The following types of uncertainty in decision models are described in the literature:

- Methodological uncertainty includes uncertainty about the decision-making perspective, time horizon, or discount rate assumed in the model (2).
- Structural uncertainty refers to uncertainty about the extent to which the structure of the model adequately captures relevant characteristics of the disease and health technology being investigated. This includes, for example, uncertainty about which disease states to incorporate or whether transitions between particular disease states are possible from a clinical point of view (8).
- Parameter uncertainty describes the uncertainty about the value of each parameter of the model. This uncertainty arises from various sources. Uncertainty due to variability between individuals occurs by chance and cannot be addressed by generating further evidence. Parameter uncertainty may also arise from a lack of high quality evidence or an imprecision in measuring parameters which can be mitigated by further evidence collection (2). This is especially the case for novel health technologies which are not yet regularly used in clinical practice and for whom no evidence on effectiveness and costs is available.

The overall uncertainty a decision maker is faced with, when deciding on the adoption of a novel technology, arises from the extent of methodological, structural, and parameter uncertainty in the decision model. To quantify the influence of parameter uncertainty on outcomes, probabilistic analysis methods have been developed, where variability of parameters is reflected by distribution functions; single or multiple parameters may be changed in the analysis (2). Further methods for quantifying uncertainty in decision-analytic models have been described in detail elsewhere (8).

The credibility of the model in simulating outcomes which are relevant for the decision maker is, however, not established by quantifying uncertainty. This lack of trust is one of the

greatest challenges facing decision-analytic modeling; if it is not overcome decision makers may dismiss models as generally untrustworthy (9, 10).

To assess how believable a model is and whether a coverage decision can be based on it, a sound validation of the model is necessary. Types of model validity and validation techniques are described in the following chapter.

1.2 Validation of decision-analytic models

1.2.1 Levels of model validity

Levels of model validity have been identified in the literature, but definitions do not always overlap (11-16). Here the definition, proposed by the joint modeling good research practices task force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Society for Medical Decision Making, is used (16):

- Face validity refers to the extent to which a model and its assumptions and applications correspond to current science and evidence as judged by experts.
- Internal validity describes the extent to which the mathematical calculations are performed correctly and are consistent with the model's specifications.
- Cross validity determines to which extent the decision model calculates same results as other models evaluating the same technology.
- External validity refers to the extent the model is able to simulate real-world outcomes from actual event data. External validity can refer to the model as a whole or to some components of it.
- Predictive validity describes whether the model is able to forecast actual events which are reported in observational studies and RCTs.

External and predictive validity of a decision model are most critical for the decision maker because they closely correspond to the model's purpose to anticipate what will occur if decision makers adapt a novel technology (16).

1.2.2 Levels of model validation

Philips et al. (2006) propose a framework for the quality assessment of decision-analytic models, which is not only a supporting tool for readers, but also a guideline for researches building models (17). It presents a literature overview of recommended validation techniques.

Face validity may be assessed by the modeling team in accordance with clinical experts to assess whether the model's structure corresponds to current research.

It is recommended that the mathematical logic of the model (internal validity) is tested by sensitivity analyses, where null or extreme values are used for some parameter inputs and

the direction of results is examined based on these. The model might also be programmed in an alternative software package or by another researcher (6, 18).

Comparing model outcomes with other decision-analytic models (cross validation) increases confidence. However, these tests may not be helpful if models are not independent and are all built on the same flawed assumptions (19).

External validity can be addressed by comparing final or intermediate model results with available, independent evidence. External validation is performed based on formal data sources, which are generated for the purpose of model validation, or informal (secondary) data sources. Depending on the extent to which the original model is based on information from the data source that is used for validation, the validation is called dependent or independent. Data should not be withheld from the model for the purpose of external validation, though (6, 19). Published validation examples show that observational data from medical records or claims data as well as RCT data are used for external validation (20-25).

In predictive validation the model is used to forecast events and, after some time, the forecasted outcomes are compared to the actual ones. Whether predictive validity of a model should be tested is discussed controversially in the literature. Some guidelines suggest that a model should demonstrate predictive validity (26, 27). Other authors conclude that some decision-analytic models are intended to support decisions at a particular point in time and not necessarily to predict future outcomes (6, 18).

Generally, the value of the information obtained by external and predictive validation must be weighed against the costs of obtaining it (19). External and predictive validity are the most difficult levels of validation as real-world data are often scarce or costly and time-consuming to collect (16). Also, external and predictive validation are methodologically the most critical part of model validation, because it involves the comparison of either outcomes of different models or comparison of model outcomes with outcomes found in the external data source. In the following, techniques for the validation of decision models based on external data sources are described.

1.2.3 External validation techniques

Two general approaches to external validation of decision models are identified in published model validations. In the first approach, the decision model simulates interventions, randomization into treatment groups, and follow-up according to the protocol of the study

that is employed for external validation. The decision model is run with parameter values extracted from the external trial; finally observed trial outcome and predicted model outcome based on trial data are compared (21-24). The second approach compares outcomes independently observed in the trial and predicted in the decision model. Evidence from the external trial is not incorporated in the decision model (20, 25).

Furthermore, the decision on validation based on published evidence or evidence generated for the purpose of model validation influences the validation approach. In validation studies, where external data stem from published studies, the decision model has to be adapted to the external data source; study design and cohort selection cannot be influenced in the external data source (22-25). Published studies are reproduced in the model by recreating cohorts in terms of demographics, baseline risk factors and complications, treatment patterns and patient management strategies (21-24).

When external validation is based on evidence generated for the purpose of model validation, the decision model's structure and outcome can be replicated in the external data source. The study by Ishida et al. (2008), for example, analyzes medical records according to the patient characteristics and study design of the validated decision model (20). More comprehensive information on patient characteristics, diagnostics, and resource utilization in the external data source allows a more precise replication of model assumptions.

ISPOR guidelines for model validation stress the importance of quantitative assessment of how well the model's results match the externally reported outcomes (16). The motivation is to determine whether differences observed between the model and external data source are significant enough to affect any conclusions derived from the model. A variety of quantitative measures are proposed in the literature, but the specific choice of measure applicable to the present validation is not readily apparent. Additionally, the interpretation of some quantitative measures of goodness of fit is unclear. Percentage errors estimated for decision models, for example, need to be interpreted in relation to other model outcomes, because from the absolute numbers no inference on goodness of fit is possible (22, 28).

A simple way to compare estimates of the external data source with the decision model is a graphical display of observed and simulated outcomes. To compare survival outcomes the comparison of Kaplan-Meier curves is applied in the literature (20, 23). For comparison of

cost outcomes, the output of the external data source and the decision model may be plotted in form of histograms such that the horizontal axis denotes costs and the vertical axis denotes the observed and simulated values, respectively (29, 30).

Statistical tests are proposed in addition to graphical analysis to obtain quantitative information about the validity of the decision model. Several studies test the null hypothesis of no difference between observed and predicted outcomes with the Student t-test (22, 29). The problem with the t-test is, however, that on the one hand the difference between observed and simulated estimates can never exactly be zero; on the other hand, the bigger the sample size is chosen, for example during bootstrapping, the smaller the critical value of the t-statistic becomes, which in turn means that a simulation model has a higher chance of being rejected as its sample size gets bigger. The t-statistic may, consequently, show a significant difference of observed and simulated estimates and yet make no statement about the validity of the decision model in simulating observed outcomes. If the sample is very large the t-statistic is nearly always significant for the difference of estimates being unequal to null. This is also called a type I or alpha error; the model is rejected while the model is valid (29).

Test-statistics other than the t-statistic are also described in the literature. These include the corrected Chi^2-test, as well as nonparametric statistical tests such as the Wilcoxon signed-rank test as a test of systematic error and Spearman's correlation coefficient for continuous variables (23, 31).

Alternatively, a hypothesis test based on bootstrapping can be employed, where it is tested whether the point estimate of the external data source is included in the model's confidence interval (CI) and the other way round (22, 25, 30). That way the problem of the t-test is avoided and no assumptions on the distribution of estimates are necessary, as in the Chi^2-test.

Also, the difference in outcome may be tested against a threshold – defined as the difference which is just acceptable to conclude that outcomes of claims data analysis and model simulation are comparable – because even in the best model results between actual treatment and model will deviate to a certain degree. However, determination of this threshold can only be based on subjective criteria (e.g. a deviation of less than 10% of outcomes estimated in claims data).

Furthermore, linear regression analysis is proposed in the literature to compare observed and simulated outcomes (22, 24, 32). Closeness of fit is assessed by plotting outcomes predicted by the model versus observed outcomes reported in the external data source, by fitting a linear curve through the points with the intercept set at zero. A squared linear correlation coefficient is obtained which provides an index of the degree to which the paired measures co-vary. It is then tested whether the linear correlation coefficient is larger than null; the simulated and real response may not necessarily have the same mean, but they may be positively correlated. A prerequisite for regression analysis is that pairs of patients (paired observations) are present in model simulation and observed data (e.g. by matching exact pairs from model and claims data).

Studies comparing outcomes from different models apply goodness of fit tests to assess validity of the validated model (22, 33). These include the mean absolute percentage error and the root mean square percentage error.

As described previously, external validation can be based on published data sources or on data sources which are generated for the purpose of model validation. A published external data source is usually not a gold standard when compared with model outputs. This is because the purpose of the model is to support decision making, which is usually not the purpose of a clinical or observational study (6, 19).

Data sources generated for the purpose of model validation may origin in primary data collection, such as RCTs, or in secondary data sources such as registries or administrative data bases. The information included in different data sources can be the gold standard for different model perspectives; registries might, for example, provide evidence for models with a focus on effects in the general population. On the other hand, administrative data, especially claims data, might be valuable to provide evidence for models with a health insurance perspective.

In the following chapters the use of claims data as a secondary data source in health economic research is described.

1.3 Claims data as a secondary data source

1.3.1 Definition

Secondary data in research are data which have not been collected with a specific research purpose. Such data are collected for management, administration or planning purposes, for the evaluation of activities within health care, for control functions, and in line with registries for surveillance purposes (34, 35).

Claims data are a secondary data source because they are collected with the purpose of billing health care provision at the expense of health insurance funds. Claims data are in the following defined as routinely collected data from various health care service categories (inpatient care, outpatient care, pharmaceuticals, physiotherapy, assistive technologies, and rehabilitative care), as well as basic information on characteristics of insured individuals, such as age, gender, and insurance status (36).

1.3.2 Claims data in health economic research

The ISPOR task force on real-world data developed a framework to assist health-care decision-makers in dealing with real-world data, including claims data. The report states that real-world data are essential for sound coverage and reimbursement decisions. Decision analytic models are the primary tool for combining clinical and claims data. Real-world data are also needed in assessing the post-launch cost-effectiveness of novel technologies to update the modeling outcomes which are made to support the initial coverage and reimbursement decisions (37).

The strength of secondary data generally lies on the generation of effectiveness evidence for the population that is of interest for the decision maker. While RCT data typically are considered as the gold standard for evidence generation on efficacy in a particular group or subgroup, they are insufficient to project the size of the effectiveness impact on the whole population (38, 39). The particular strength of claims data is the data's ability to display utilization of medical technologies in routine care and real-world cost data incurred by insurance funds (40).

Another advantage of using secondary data sources is that they already exist and the time spent on the study is therefore likely to be considerably less than on studies based on

primary data collection (35). Especially in the coverage decision-making process, where a timely generation of evidence is of importance, use of secondary data sources is useful (37). Furthermore, study costs are significantly reduced compared with collection of primary data (35).
A particular advantage of the use of claims data is the size of the covered population, which allows researchers to identify rare events and assess economic impact of various interventions (37). Furthermore, collection of claims data does not impose an additional burden on patients and clinicians, in contrast to clinical trials and registries (41).

Limitations of secondary data sources regarding data quality include missing data, coding errors, and the lack of comprehensive data across health care service categories. Claims data are especially limited in the extent of clinical information on inpatient stays, health outcomes, health status, and symptoms. A distinction between costs and charges might in some cases be not possible based on claims data (37). Completeness of claims data is limited to those who seek care in the first place and additionally obtain care through the insurance fund (42). Generally, claims data are not collected for clinical research and coding practice might follow economic incentives of the underlying reimbursement system (40, 43).
The most common and challenging methodological issue arises from treatment selection bias. Due to non-randomization, estimates of effects and costs can be biased because of a correlation between unobserved factors associated with treatment selection and outcome, for example the baseline health status (37). Matching-techniques and other statistical methods can be used to mitigate this bias (44).
General limitations of claims data compared to primary data collection and techniques to handle these have been discussed extensively (40, 44-49).

1.3.3 German statutory health insurance claims data

In Germany, claims data of statutory health insurance (SHI) funds are a valuable source of evidence for research, because medical care, resource use, and reimbursement are documented in detail and over a long period of time for a large cohort of patients.

About 85% of the German population is insured within the social security system of SHI. The remaining 15% of the population are covered by private insurance. The German statutory

health care system is characterized by pay-as-you-go financing and income-based insurance contributions. All SHI schemes are regulated by the Social Code Book V (SGB V). The SHI funds are responsible for negotiating prices, quantities, and quality assurance measures with providers on behalf of their members (50).

SHI covers inpatient and outpatient care, pharmaceuticals, medical devices, assistive technologies, physiotherapy, (ambulatory) rehabilitative care, and sickness benefits. Co-payments of patients are compulsory, especially for pharmaceuticals and assistive technologies. A small part of health-related social services is covered by accident insurance, retirement insurance, and long-term care insurance. Services covered by these funds are not documented in SHI claims data (50).

In SHI claims data, diagnostic information is coded by the German application of the International Classification of Disease, version 10 (ICD-10 GM). In outpatient care, the certainty of the diagnosis is stated by an additional code (secured: G, tentative: V, exclusion: A, status post: Z). In inpatient care information is given on the point in time during the hospital stay and the department by which the diagnosis was coded and whether it is the primary or secondary diagnosis. Inpatient procedures are coded by operation and procedure codes (OPS) and reimbursement of these is reflected in DRGs. Outpatient procedures and reimbursement of these are found in the uniform value scale (Einheitlicher Bewertungsmaßstab, EBM) for SHI physicians. Pharmaceuticals are distinctly coded by a uniform pharmaceutical identification key (Pharmazentralnummer, PZN); assistive technologies and physiotherapeutic procedures have distinct codes as well (51).

Due to the administrative nature of the data, only SHI relevant hospital episodes and outpatient visits are coded (52). Additionally, SHI claims data contain basic information on date of birth, gender and place of living, whereas more sophisticated socio-demographic information (household size and income, education, occupation) is not comprehensively documented for all insured individuals. Reasons for dropout are also coded which could be death or transition to another insurance fund.

Cost data are available for inpatient and outpatient care, as well as pharmaceuticals, physiotherapy, assistive technologies, and rehabilitation on a patient level in SHI claims data. Complete inpatient data have to be reported to sickness funds immediately after discharge (53). Outpatient data are usually delayed by about six months as they are transferred over

the association of SHI physicians to sickness funds. Diagnostic information in outpatient data is, in general, summarized on a quarterly basis per patient, whereas inpatient data are reported on a hospital-episode basis per case (54, 55).

Confidentiality issues are to be considered when patient data are not anonymous or pseudonymization is not possible. In such cases ethics approval and/ or the approval of the data protection agency as well as informed consent of affected patients is necessary to use claims data for research (56).

1.4 Objective

Decision-analytic models are developed to inform decision-makers about adoption and reimbursement of novel health technologies in treatment practice. Uncertainty about the validity of decision models and generalizability of simulated outcomes, however, limits the credibility of models for decision making. Validation is essential to establish trustworthiness of decision-analytic models in the decision making process.
Independent external data sources for validation are, however, often scarce. During model building not only a trade-off between incorporation of all best available evidence and exclusion of data sources for the purpose of external validation has to be made. Also, generation of primary data with the aim of model validation is usually costly and time-consuming, especially evidence from clinical trials.

This study assesses an alternative to external validation of decision-analytic models with primary data sources – external validation based on claims data of health insurance funds.
The underlying assumption is that claims data represent the gold standard for real-world resource use and costs incurred by health insurance funds. Outcomes from claims data are representative for the study population most relevant to health insurance funds, which make the decision to reimburse a novel technology. Analysis of claims data for the purpose of model validation is, also, less costly and realized in a timelier manner than clinical trials.

A validation approach is developed which highlights critical steps in the external validation based on claims data. The validation approach is exemplarily applied to the external validation of a Markov model comparing treatment options of localized prostate cancer in a German health care context. It is validated based on claims data of a large German SHI fund.
The claims data set is used to build a cohort reflecting the model population as closely as possible. Input parameters for unit costs as well as resource use and treatment costs accumulated over the study duration are estimated in the claims data analysis. Comparison with model outcomes is based on statistical tests, with the underlying hypothesis that costs simulated by the model are equal to the costs observed in claims data.

The focus of this study lies on the methodological approach to the external validation based on claims data. Application of the validation approach to other health care systems and

medical indications is discussed in detail. Strengths and limitations of claims data based validation are presented to answer the research question:

'*Are claims data of health insurance funds suitable to externally validate decision-analytic models?*'

1.5 Structure of the dissertation

Chapter 2 gives an introduction to the example of use – localized prostate cancer (PCa). Disease background including epidemiologic and socioeconomic burden as well as diagnostics and disease classification are presented. Treatment options for localized PCa, evidence on comparative effectiveness, and cost studies are summarized. Finally, the structure, basic assumptions, and data sources of the validated Markov model, comparing costs and utilities of treatment of localized PCa by Koerber et al. (2014), are described.

The main focus of this study lies on chapter 3, where the methods of validation are described. Chapter 3.1 presents an overview of the proposed step-wise validation approach, whose implementation is described on the exemplary validation of the model by Koerber et al. (2014) in the chapters 3.2 to 3.8.

First, the selection of the validation level is presented. Then, study design and methods for cohort selection in the claims dataset are described, which allow creation and follow-up of a patient cohort comparable to the Markov model. Special attention is given to the selection of procedure and diagnostic codes in the claims data that reflect disease-related resource use and costs assumed in the model. Statistical methods for calculation of resource use and costs in claims data are described separately for an excess cost and a disease-related cost approach. Changes to the Markov model's runtime and age structure of the cohort are described in line with additional model analyses conducted to gain resource use and cost outcomes in a comparable format to outcomes from claims data analysis. In chapter 3.7 methods for comparison of input parameter values for unit costs in the Markov model and unit costs calculated in claims data is presented. Also, methods to compare costs and resource use simulated by the model against observed outcomes of claims analysis are described. Finally, sensitivity analyses are presented which are constructed to test robustness of outcomes to modification of the claims data cohort. Further sensitivity analyses relate to the impact of a change of the age structure of the model cohort on the agreement of model and claims data outcomes.

In chapter 4 results of the validation of the PCa-model are presented; first, results of the claims data analysis. Secondly, results of the additional model analyses conducted for this study are displayed. Input parameters as well as simulation outcomes of model and claims

data are compared; unadjusted and simulated results of the comparison are presented in tables and graphs. The impact of sensitivity analyses on outcomes is described, as well.

Chapter 5 interprets and contrasts results presented here to published validation approaches. Strengths and limitations of claims data in respect to the research question are discussed. The practical implications of results for model validation but also application of the validation approach in industry is debated. Special focus lies on the critical assessment of the proposed validation approach. The applicability of the validation approach to administrative data from health care systems other than the German SHI and applicability to indications other than PCa are discussed.

Chapter 6 contains concluding remarks and indicates areas requiring further research.

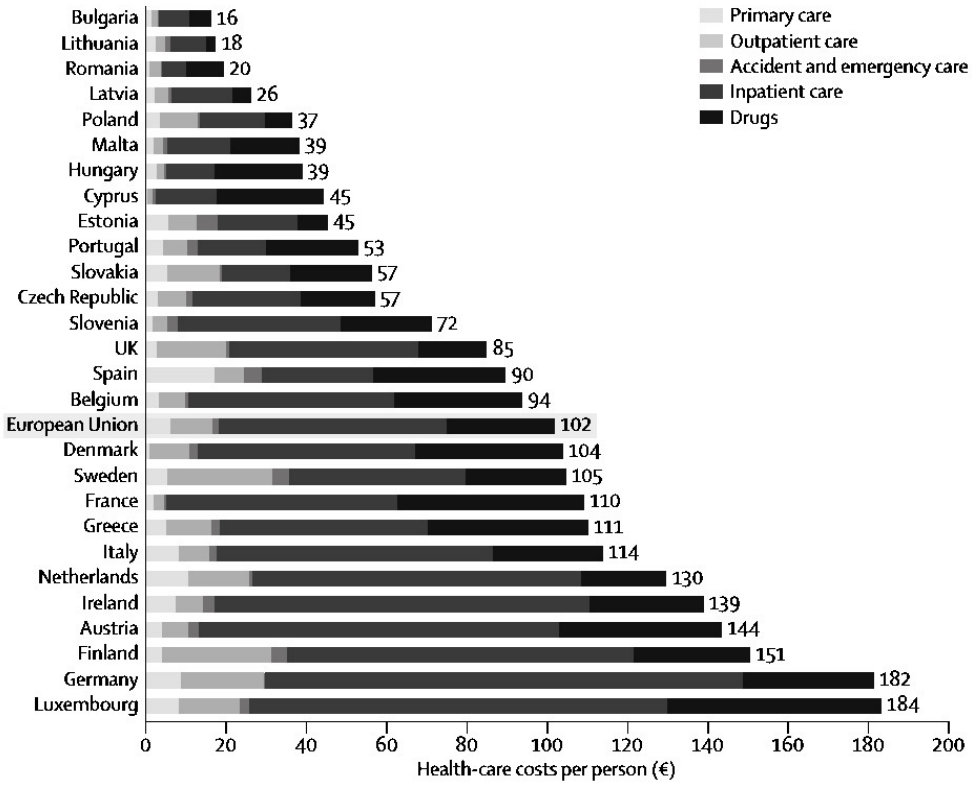
2 Example of use: localized prostate cancer

2.1 Disease background

2.1.1 Epidemiology and socioeconomic burden

PCa is the second most common cancer in men worldwide and the most common in Germany, making it a major health concern (57, 58). Reported worldwide incidence rates vary widely, being highest in the regions of the United States (US) and Europe where prostate specific antigen (PSA) testing and subsequent biopsies have become widespread, leading to a continuous increase in incidence rates (59).

PCa-specific mortality rates, on the other hand, are decreasing since the 1990s in Germany and other European countries (60). Today, 5-year relative survival rates are around 83% for men newly diagnosed with PCa (61). This is, in part, due to the fact that screening methods allow earlier tumor detection (62).

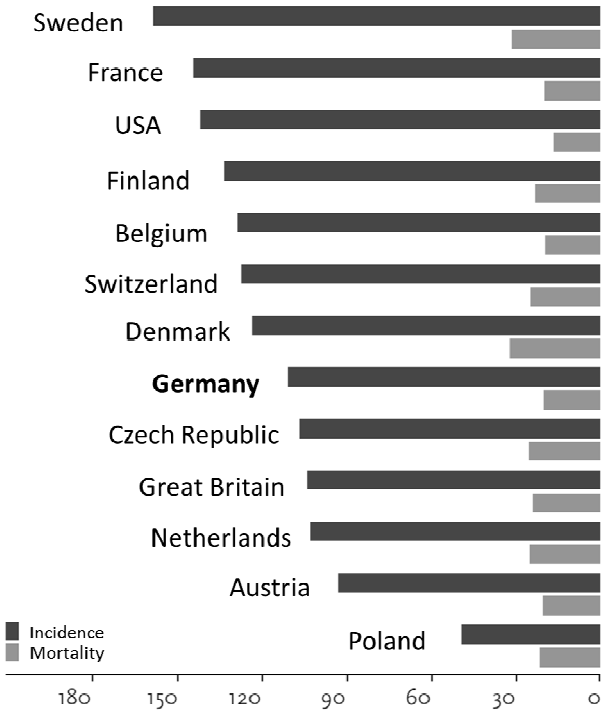


Source: according to Luengo-Fernandez et al. 2013 (63)

Figure 2-1: Health care costs of prostate cancer per person in European Union countries in 2009, by health care service category

PCa is a diagnosis which occurs mostly in older men; mean age at diagnosis is 69 years (59). In the light of demographic change – it is estimated that around 34% of the population will be over 65 years of age in 2060 in Germany – a further increase in incidence of PCa is expected (64). In turn, the discrepancy between incidence and mortality rate will likely increase further.

With the expected demographic change and increase in the incidence of PCa, the economic burden in Europe is also expected to increase significantly. Even today, annual health care spending due to PCa is substantial and accounts for €5.43 billion in the European Union (EU), with a high proportion of the health care costs occurring in the first year after diagnosis. Germany exhibits the highest PCa-related health care costs per person in the EU, mostly due to inpatient expenditures (Figure 2-1). The correlation between health care expenditure and cancer mortality is, however, unclear (63). A cross-country comparison of PCa incidence and mortality rates indicates that health outcome in Germany is slightly better than in other European countries (Figure 2-2) (59).



Source: according to Robert-Koch-Institut et al. (2013) (59)

Figure 2-2: Cross-country comparison of age-standardized prostate cancer incidence and prostate cancer-specific mortality rates (per 100,000), 2009/2010

2.1.2 Diagnosis and tumor classification

The diagnosis of early stage PCa is usually suspected based on digital rectal examination (DRE) and PSA levels. PSA is a serum marker, with higher values indicating a higher probability of PCa (65).

Confirmation of the diagnosis requires histopathological verification of the tumor, usually adenocarcinoma, in the prostate biopsy cores or operative specimens.

T - Primary tumor		
TX		Primary tumor cannot be assessed
T0		No evidence of primary tumor
T1		Clinically inapparent tumor not palpable or visible by imaging
	T1a	Tumor incidental histological finding in 5% or less of tissue resected
	T1b	Tumor incidental histological finding in more than 5% of tissue resected
	T1c	Tumor identified by needle biopsy
T2		Tumor confined within the prostate
	T2a	Tumor involves one half of one lobe or less
	T2b	Tumor involves more than half of one lobe, but not both lobes
	T2c	Tumor involves both lobes
T3		Tumor extends through the prostatic capsule
	T3a	Extracapsular extension (unilateral or bilateral) including microscopic bladder neck involvement
	T3b	Tumor invades seminal vesicle(s)
T4		Tumor is fixed or invades adjacent structures other than seminal vesicles: external sphincter, rectum, levator muscles, and/or pelvic wall
N - Regional lymph nodes		
NX		Regional lymph nodes cannot be assessed
N0		No regional lymph node metastasis
N1		Regional lymph node metastasis
M - Distant metastasis		
MX		Distant metastasis cannot be assessed
M0		No distant metastasis
M1		Distant metastasis
	M1a	Non-regional lymph node(s)
	M1b	Bone(s)
	M1c	Other site(s)

Source: according to Sobin et al. 2009 (66)

Table 2-1: Tumor Node Metastasis (TNM) classification of PCa

The standard diagnostic procedure is to use ultrasound to guide prostate biopsies and to administer oral antibiotics before the procedure to prevent an infection (65). Complications of prostate biopsy include hematuria and hematospermia as well as post-procedural infections. Reported rates of severe infections have increased due to the evolution of antibiotic resistant strains (67).

Tumor grade is based on the pathology of prostate core biopsies or surgery specimens. The (modified) Gleason grading system, as defined by the 2005 consensus conference of the International Society of Urological Pathology, is used to grade prostatic adenocarcinoma. In the Gleason system, grades are based on the architectural pattern of the tumor (grade 1-5). The Gleason score is reported as the sum of the two most common grade patterns (score 2-10) (68).

Tumor stage describes the extent of the prostatic carcinoma. It is usually assessed by DRE and PSA testing and supplemented by bone scan or computed tomography (65, 69). It is standard in clinical trials and routine care to define tumor stage by the Tumor Node Metastasis (TNM) classification proposed by the International Union Against Cancer (Table 2-1) (66). 'Localized PCa' is present in stages T1-2 N0 M0, whereas 'locally advanced PCa' in stages T3-4 N0 M0. Stages N1 and M1 indicate 'metastatic PCa'.

Based on PSA level, Gleason score, and TNM stage, localized tumors are further categorized regarding the risk of recurrence (70):

- Low-risk: PSA ≤ 10 ng/ml and Gleason score ≤ 6 and T-stage 1/ 2a
- Intermediate-risk: PSA > 10 ng/ml - 20 ng/ml or Gleason score 7 or T-stage 2b
- High-risk: PSA > 20 ng/ml or Gleason score ≥ 8 or T-stage 2c

Risk factors for PCa are still unclear; however there is a strong correlation with age and family history of PCa. Untreated prostate tumors are, generally, characterized by a slow natural progression where patients predominantly die of other causes than PCa (69). Detected tumors are often localized, clinically insignificant cancers, which do not require any treatment in their lifetime (Figure 2-3) (59, 71).

In this light, early detection and screening programs for PCa are controversial among experts as they bear the risk of over-detection and, following that, overtreatment. A Cochrane review of five randomized controlled trials, representing more than 341,000 randomized men, could not show any PCa-specific survival benefit of screening (72). In Germany, no

population screening program for PCa based on PSA testing is recommended by health authorities; regular DREs are covered by SHI in line with PCa screening. However, if a patient wishes an early detection examination in form of PSA testing, this is covered by SHI as well (69).

Source: according to Robert-Koch-Institut et al. 2013 (59)

Figure 2-3: Distribution of tumor stage at diagnosis, Germany 2009/ 2010

2.1.3 Treatment of localized prostate cancer

Treatment of localized PCa can be divided into two groups: treatment strategies with a curative intention and conservative management (CM) strategies with deferred treatment.

Standard curative treatment of localized PCa is radical prostatectomy (RP), which can be performed with a retropubic, perineal, or (robot-assisted) laparoscopic approach. The entire prostate gland is removed between the urethra and bladder, and both seminal vesicles, along with sufficient surrounding tissue to obtain a negative margin, are removed (65). The aim of RP is eradication of disease, while preserving urinary continence and if possible erectile function (73). Still, post-operative urinary incontinence (IC) and erectile dysfunction (ED) are common complications following RP. Recent systematic reviews report mean continence rates of 80-97% with retropubic RP versus 89-100% with robot-assisted RP and mean potency rates of 26-63% versus 55-81%, respectively, 12 months after surgery (74, 75).

Radiotherapy is an alternative to surgery for curative therapy of localized PCa. Definitive radiotherapy includes external-beam radiotherapy (EBRT) and brachytherapy. Intensity-modulated radiotherapy, with or without image-guidance, is the gold standard for EBRT (65). Brachytherapy describes a therapy option where a radiation source is implanted via a transperineal approach (69). Complications of radiotherapy are late genitourinary (including IC) or gastrointestinal toxicity as well as ED (65, 76). Radiotherapy affects potency, however, to a lesser degree than retropubic RP: a meta-analysis from 2002 reports 12-month probability for maintaining erectile function of 55-76% after radiotherapy and 25-34% after RP (77).

Other curative therapies, with less evidence regarding outcomes, are not recommended in guidelines as primary treatment of localized PCa. These include proton beam therapy, cryosurgery, and high-intensity focused ultrasound therapy (65, 69).

Data suggest that many men with localized PCa do not benefit from curative treatment in terms of survival, because tumor progression is so slow that no treatment is required (78). The CM strategies active surveillance (AS) and watchful waiting (WW) have been proposed as alternatives to curative treatment to reduce overtreatment and subsequent complications. It is estimated that about 45% of men with a PSA-detected PCa are candidates for CM (79).

Based on published evidence, the German Association of Urology (Deutsche Gesellschaft für Urologie; DGU) and the European Association of Urology (EAU) define eligibility for AS strategy by the following criteria (69, 80, 81):

- PSA-level ≤ 10 ng/ml
- Gleason score ≤ 6
- Tumor stage T1 and T2a
- Life expectancy ≥ 10 years

The recommended surveillance scheme consists of PSA testing, DRE, and regular biopsies (69). PSA tests and DRE should be performed every 3 months for the first 2 years, after that every 6 months if the PSA level is stable. The first biopsy is supposed to be taken 6 months after surveillance initiation and in the first 3 years, follow-up biopsies are taken every 12-18 months. If tumor status is stable, further biopsies are performed every 3 years. Studies suggest, though, that in addition to complications described previously, regular biopsies might affect erectile function (82, 83).

Curative treatment is initiated if the inclusion criteria described above are no longer met. 5 to 10 percent of men under AS choose a curative treatment although tumor progression does not require it (84).

If life expectancy is less than 10 years or co-morbidity does not allow any other form of PCa-treatment, watchful waiting (WW) is suggested. This strategy has no standardized follow-up scheme and does not monitor tumor progression closely. Symptom-oriented, palliative therapy is initiated if disease progresses (69).

Comparative effectiveness of localized PCa treatment is usually reported as overall mortality, PCa-specific mortality, or health-related quality of life. At the time of literature search, only two RCTs report comparative effectiveness outcomes of localized PCa treatment: the Scandinavian SPCG-4 trial and the US-based PIVOT trial.

The SPCG-4 trial randomly assigned 695 newly diagnosed, localized PCa cases to RP or WW. After a follow-up of 23 years, the SPCG-4 trial showed that RP was associated with a reduction of all-cause mortality compared to WW (relative risk (RR)=0.77, confidence interval (CI) [0.59;0.86], p<0.001). RP was also associated with a reduction in PCa-specific mortality (RR=0.56, CI [0.41;0.77], p=0.001) as well as development of distant metastases (RR=0.57, CI [0.44;0.75], p<0.001) and use of androgen-deprivation therapy (ADT) (RR=0.49, CI [0.39;0.60], p<0.001) (85). Health-related quality of life was better in the RP group than in the WW group, but lower in both cases compared to a general-population control. Reduction in quality of life is induced by ED and IC in case of RP, whereas men under WW are predominantly affected by symptoms of tumor progression (86).

The PIVOT trial also randomized 731 men with localized PCa to RP or WW. After a median follow-up of 10 years the trial showed, contrary to SPCG-4, that RP did not significantly reduce all-cause mortality (hazard ratio (HR)=0.88, CI [0.71;1.08], p=0.22) or PCa-specific mortality (HR=0.63, CI [0.36;1.09], p=0.09) compared to WW (87).

RCT evidence on comparative (cost-)effectiveness of RP, EBRT, brachytherapy, and AS is expected to be available when results of the ongoing ProtecT trial (United Kingdom; UK) and the German PREFERE trial are available (88, 89).

Several observational studies report comparative effectiveness outcomes. In a study with about 28,000 localized PCa-cases, identified in the Surveillance, Epidemiology, and End Results (SEER) population-based cancer registry (US), non-curative therapies (AS, WW, TURP) are significantly associated with an increased risk of PCa-specific mortality compared to curative therapies (RP, EBRT, brachytherapy) (HR=1.05, CI [1.02;1.08], p<0.001) over a 2.5 year study duration (90). First results of the German observational study, HAROW, indicate that after 2 years of follow-up, AS exhibits the highest tumor progression rate (31%) and EBRT the lowest (7%) compared to RP, ADT, and WW. All-cause and PCa-specific mortality have not been reported yet (91). Results of both studies, however, are limited in their validity due to a short follow-up and non-randomization.

In the CaPSURE-registry (US) health-related quality of life outcomes are shown for several localized PCa therapies (RP, EBRT, brachytherapy, ADT, WW/ AS). There is no significant difference in physical and mental health, assessed by the Medical Outcomes Studies Short Form-36 (SF-36), after 10 years of follow-up (92). Similarly, the PCOS study, a US-based registry including about 3,500 men newly diagnosed with PCa and treated with either RP or radiotherapy, found no significant differences between the groups regarding SF-36 results. (93).

A meta-analysis examining curative treatment of low-risk PCa reports higher progression-free survival rates for brachytherapy than RP and EBRT, which have comparable rates (94). The PRIAS study recruits men under AS worldwide and reports effectiveness outcomes, but does not compare these to any other PCa-specific treatments (95).

2.2 Cost studies

One economic study reports disease-related resource use and costs for treatment of localized PCa based on RCT data – the SPCG-4 trial. After 12 years of follow-up, total mean costs per patient are significantly higher in the RP group than in the WW group with a difference of €6,123 (34%, $p<0.01$). This difference originates predominantly in the costs of the surgical procedure (96). A cost study based on CaPSURE registry data (US) also concludes that WW generates lowest annual disease-related costs over a period of 5 years compared to RP, EBRT, brachytherapy, cryotherapy, and ADT (97). Similar results are presented by both a cost study on data from the French network of cancer registries and a cost analysis based on SEER-Medicare linked data (98, 99).

A cost study conducted with claims data from several US health insurance funds reports that, while costs of RP and EBRT are significantly higher than costs of CM in year 1 with additional $15,200 and $18,900 respectively, after year 2 no significant cost differences can be shown between treatment strategies (100). Only one cost analysis comparing in- and outpatient urologist reimbursement for AS to RP over a period of 10 years reports that costs of AS exceed costs of surgery after 4 years of follow-up (101).

Cost analyses based on decision-analytic models present similar results to observational studies: A Markov model, comparing costs of RP, EBRT, brachytherapy, ADT, and AS, reports cost savings per patient of AS amounting to $16,039 (CI [16,039;16,042], $p<0.001$) after 5 years and $9,944 (CI [9,941;9,948], $p<0.001$) after 10 years compared to immediate curative therapy. EBRT in combination with ADT has the highest costs after 5 and 10 years follow-up (102). Eldefrawy et al. (2013) and Corcoran et al. (2010) also report that cumulative costs of AS are lowest, even though AS has higher follow-up costs than curative treatment (103, 104). When considering effectiveness and costs, the Markov model by Koerber et al. (2014) shows that AS is the dominant strategy compared to RP with 0.04 additional QALYs and a cost reduction of €6,883 per patient. Considering only life years gained without quality adjustment as outcome measure, RP is more effective with an ICER of €96,420/life year gained (105). In contrast, a Markov model based on SPCG-4 trial data reports that RP has higher costs but also higher QALY outcomes than AS (106).

The decision analytic model by Hayes et al. (2013) directly compares AS and WW in a cost-effectiveness analysis. It shows that WW is cost-saving compared to AS by $15,374 with a

quality-adjusted life expectancy gain of 2 months. AS becomes as effective as WW if less than 63% of men progress to curative treatment in their lifetime (78).
A cost-effectiveness analysis of RP compared to radiotherapy shows that surgical methods are – in terms of QALYs – significantly more effective and less expensive than radiotherapy strategies for the treatment of low-risk PCa (107).

Overall, evidence suggests that CM strategies, especially WW, save health care costs which arise due to unnecessary curative therapy and treatment of its adverse effects, and additionally achieve better quality of life outcomes. Most studies base cost analysis on US reimbursement values, though, which are not representative for European health care systems, mainly due to cost differences between public and private health care systems (108).

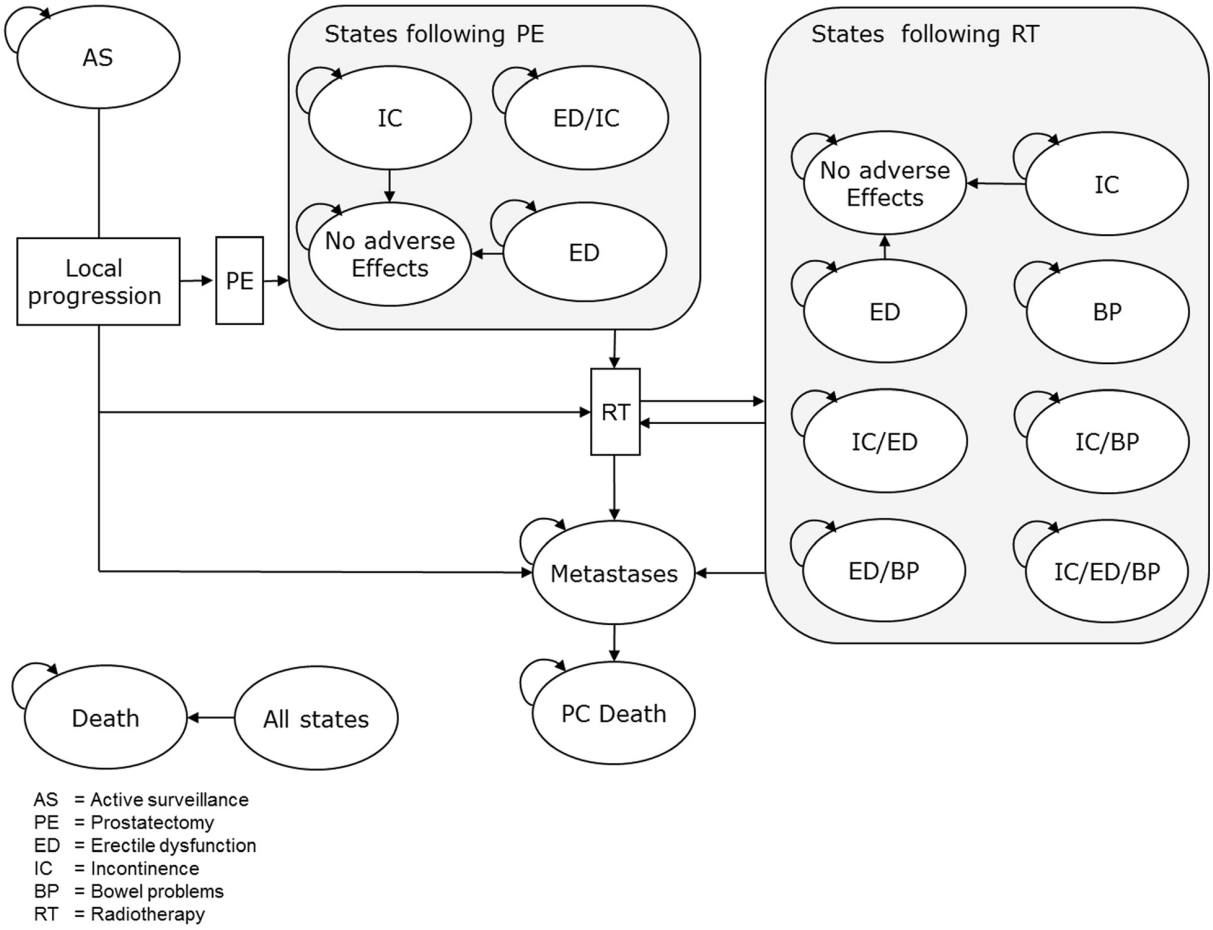
2.3 Validated model

A lifetime Markov model by Koerber et al. (2014), comparing the cost-utility of AS and RP for a cohort of men newly diagnosed with low-risk PCa, was chosen as the basis for validation (105). The PCa-model was selected from 3 decision models on different indications, developed at Helmholtz Zentrum München, because it features structural assumptions and data sources that are commonly used in decision-analytic modeling. Also, it includes resource use and cost parameters from a wide range of health care service categories. These characteristics make the model suitable for a validation example; results of the validation can be transferred to other indications. Also, the decision model is highly relevant for German SHI funds because PCa has a high (socioeconomic) burden of disease in Germany and treatment incurs substantial costs at the expense of SHI funds. Thus, external validation based on SHI claims data seems suitable for this model, as it is assumed that a patient cohort and outcomes relevant for SHI funds can be analyzed in SHI claims data.

In the decision model by Koerber et al. (2014) men enter the model at the age of 65 years and are assumed to have a life expectancy greater than 15 years as well as no severe co-morbidities, including benign prostate hyperplasia (BPH).
Utility is measured in QALYs; direct medical costs, including out-of-pocket payments, from a broad range of health care service categories (in- and outpatient care, pharmaceuticals, assistive technologies, physiotherapy) are considered. The model adopts the perspective of the SHI scheme insurant population as recommended by the German Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; IQWiG).
Figure 2-4 provides an overview of the Markov model. In case of local recurrence after RP, EBRT is the primary treatment option. In case of local progression under AS, RP is the standard treatment for men < 72 years of age and EBRT for older men, respectively. Metastatic disease is treated by ADT and in case of refractory disease by chemotherapy. For bone metastasis treatment with zoledronic acid and radiation therapy is assumed.
Adverse effects of RP include short- and long-term ED, IC, and a combination of both. After EBRT men may, additionally, develop bowel problems which include abdominal pain,

bloating, and diarrhea. Regular biopsies performed under AS may lead to urosepsis requiring hospitalization. PCa-specific or other-cause death may occur in any state.

The model is implemented in Treeage Pro 2013.



Source: Koerber et al. 2014 (105)

Figure 2-4: Structure of the validated model by Koerber et al.

A variety of data sources provide evidence for the model. Treatment pathways are based on recommendations from DGU (109). Health state specific utility rates are derived from published literature (110-112). In the absence of suitable evidence, it is assumed that AS exhibits the same utility as life after curative treatment without side effects. Comparative PCa-specific mortality is based on evidence from the SPCG-4 trial comparing RP and WW (113). To adapt results to AS strategy it is assumed that half the treatment effect of RP is maintained compared to AS. The transition probability of developing metastases under AS is

also adapted from the SPCG-4 trial. All other transition probabilities are derived from a meta-analysis comparing AS and RP (114).

Resource utilization is based on treatment guidelines and estimation by clinical experts (109). Outpatient unit costs are based on values of the physicians' fee catalogue and inpatient costs on DRG weights from the German DRG catalogue (115, 116). Pharmaceutical prices are extracted from the German formulary (117); unit costs of assistive technologies and physiotherapy are based on market prices provided by health care providers.

Parts of the model were validated during model building. Face validity of the model structure and of major assumptions was screened within the modeling team and by a clinical expert. This resulted in adaptation of the model regarding development and treatment probabilities of adverse effects as well as assumptions on resource utilization. Results of the model were cross validated with results of two existing decision-analytic models comparing effectiveness and one cost-utility analysis comparing effectiveness and costs of AS and RP. All studies report that more QALYs are generated under AS than with initial RP (78, 112, 118). The QALY advantage of AS reported by Koerber et al. (2014) is, however, smaller because an age related decline in quality of life is considered. Regarding costs, Hayes et al. (2013) also found that AS is a cost-saving strategy. The probability of PCa-specific death was externally validated with data from the PIVOT RCT, which did not change strategy rankings (87).

3 Methods

3.1 Validation approach

Based on the theoretical literature and published examples of external model validation, described in chapter 1.2, an approach to the external validation of decision-analytic models based on claims data is proposed which highlights critical steps in the validation process. The individual validation steps include the selection of 1) validation level, 2) claims dataset, study design, and patient cohort, 3) relevant health states and health technologies, 4) statistical methods for claims data analysis, 5) changes to the decision model, 6) comparison between model and claims data, and 7) sensitivity analyses.

Implementation of this step-wise approach is described exemplarily for the validation of the Markov model on treatment of localized PCa by Koerber et al. (2014) based on claims data from a German SHI fund in the following chapters 3.2 to 3.8.

Step 1: Validation level

The decision on the validation level depends on the uncertainty present in the validated model. An external validation based on claims data is useful for decision-analytic models, which present substantial parameter uncertainty regarding resource use, probability of utilization, or costs. Efficacy or quality of life outcome measures are difficult to validate based on claims data, because clinical information is typically very limited and diagnostic codes are, in most cases, not sufficient to describe severity of disease and quality of life. Effectiveness measures concerning utilization of health technologies, such as number of hospital episodes, on the other hand, are validly represented in claims data.

To compare outcomes of the decision model and the external data source two general approaches are possible. In the first approach, the decision model simulates cohort characteristics and study design of the external data source, which is usually a published RCT or observational study; the decision model is then run with the parameter values extracted from the external data source. In the second approach, outcomes are independently estimated in the external data source and the decision model; parameter values of the external data source are not incorporated in the decision model.

External validation can be applied to input parameters fed into the model (such as resource use and unit costs) and simulation outcomes generated by the model (such as probability of

utilization and accumulated costs over the study period). Validation of input parameters is useful to explain differences between model simulation outcomes and claims data outcomes. Differences in outcome might be due to differing input parameters, the model calculation itself, or a combination of both.

Step 2: Claims dataset, study design, and patient cohort

The most important criterion for the selection of the claims dataset used for the validation is that study design, assumptions, and patient cohort of the decision model can be replicated as closely as possible. To achieve this, the insurance fund(s) providing the dataset should be located in a health-care system with similar treatment pathways and reimbursement arrangements as assumed in the decision model.

The indication or treatment of interest should be present in the claims dataset in a population large enough to estimate valid outcomes. Furthermore, characteristics of the model cohort (such as age structure, co-morbidity, or socioeconomic status) should be found in the claims dataset. The coding of diagnoses and procedures in the claims dataset should make it possible to select a cohort according to the inclusion criteria of the model (e.g. distinction between localized and metastatic PCa-disease).

It should also be ensured that the claims dataset allows a long enough follow-up of patients to be comparable to the whole model span or, at least, parts of it representing relevant disease stages. A unique identification number is necessary to merge claims from different health care service categories and years on an individual insurant level.

Step 3: Relevant health technologies and costs

To identify within the claims data set the utilization and costs of treatment strategies compared in the model, a definition of disease-related health technologies is necessary. This definition can be set more or less closely to the decision model's definition of disease-related health care utilization. It can either strictly follow the definition used in the decision model by, for example, using the same procedure codes on which the identification of disease-related health technology utilization is based. Or a wider definition of codes may be applied, if no information on the definition of disease-related health care utilization is provided for the model or codes used in the model do not exist in claims data.

Codes for calculation might also be defined irrespective of the model's definition of disease-related technology utilization. In addition to parameter uncertainty, structural uncertainty regarding the definition of resource utilization in the model is addressed in this case.

An alternative approach to estimation of disease-related technology utilization is an excess approach. In this case, all health care utilization incurred at the health insurance fund, irrespective of the relevance for the treatment of interest, is summarized and compared between treatment alternatives. The absolute value of resource utilization and corresponding costs of a treatment strategy estimated via an excess approach has no informative value and, thus, only validation of incremental outcomes of the decision model is possible.

Resource use and costs may be defined by codes of inpatient (OPS) and outpatient procedures (EBM), pharmaceuticals (PZN), assistive technologies, and physiotherapeutic procedures in SHI claims data. For the identification of health technologies relevant for the treatment of adverse effects of PCa-treatment, identification of health states that describe these adverse effects may be necessary. Health states are usually defined as diagnostic ICD-codes in claims data.

Depending on the perspective of the decision model only costs incurred by the SHI fund or costs additionally including co-payments of insured individuals (SHI scheme insured community) can be considered in SHI claims data analysis.

Step 4: Statistical methods for claims data analysis

For one, statistical methods are used to reflect cohort assumptions of the decision model in the claims data analysis. An important assumption in decision models is that treatment groups are equal in their baseline characteristics, such as age, gender, and co-morbidity. In claims data analysis, as in all observational studies, no randomization into treatment groups is possible which would ensure equal distribution of baseline characteristics. Statistical methods are available to mitigate selection bias due to non-randomization in claims data analysis. These include matching techniques, adjustment in regression analysis, and stratified analysis.

Secondly, standard statistical methods for health economic evaluation should be applied for cost analysis in claims data. These include methods which take the skewed distribution of cost data into account, such as generalized linear models with a gamma-distribution and log-

link as well as two-part models. Application of the recycled predictions method is useful to estimate an absolute cost difference between treatment strategies which is needed for the comparison with model outcomes.

Step 5: Changes to decision model

Depending on the characteristics of the claims data set, it might be necessary to change model assumptions to make comparison with claims data possible – provided that access to the model is granted. This may include adaptation of the runtime of the model to the study period of the claims data analysis.

In addition, it might be difficult to replicate the age structure of the model cohort in claims data. Especially, if the model assumes that patients all have the same age at treatment initiation, treatment groups in claims data can become very small. Adaptation of the age structure of the model to the claims data cohort is necessary in this case.

Additional analyses may be necessary, because decision models usually only report total costs per strategy. For the validation, however, it might also be interesting to simulate probability of utilization and costs of single treatments; differing total costs in claims data and model may be explained by differing utilization and costs of single treatments. Furthermore, it is of interest to analyze the validity of model simulations in different health care service categories, such as inpatient and outpatient care.

The model structure should not be changed to conduct additional analyses; the aim of external validation is the comparison of input parameters and simulation outcomes of the model as it was originally designed with an external data source, and not the adaptation of the model structure to the external data source.

Step 6: Comparison between model and claims data

Absolute costs of each treatment strategy might be compared between model and claims data. Alternatively, incremental costs as estimated by claims data and model are compared, which is of greater importance for the decision maker.

As a first step, resource use and costs estimated in claims data analysis may be compared with outcomes generated by a microsimulation of the decision model; in microsimulation a cohort including the same number of patients as in claims data analysis is simulated in the decision model. Microsimulation pictures variability due to the alternative pathways through

the decision model and is similar to the population variability of outcome in the claims data analysis. Descriptive (mean, standard deviation, median, interquartile range) and graphical presentation is useful to show distribution and variability of results.

In a second step, uncertainty of parameters in claims data and model analysis may be considered. Outcomes generated by regression analysis and bootstrapping of claims data are compared with outcomes of Monte Carlo simulation of the decision model. Overlapping of simulated outcomes can be shown graphically. Additionally, the hypothesis that simulated incremental costs of the model are equal to observed incremental costs of claims data analysis may be tested by statistical methods described in chapter 1.2.3.

Step 7: Sensitivity analyses

Sensitivity analyses are useful to explain differences between model and claims data. For one, sensitivity analyses can address changes to the claims data cohort. This might include changes to the inclusion criteria in the claims data cohort to replicate the assumptions of the decision model more closely.

Apart from sensitivity analyses conducted with claims data, it might be useful to vary model assumptions as well, for example the age structure in the decision model, to assess whether differences in outcome are due to differing assumptions in model and claims data analysis.

3.2 Validation level

As described above, the first step in the external validation of the PCa-model by Koerber et al. (2014) is the selection of the validation level.

In this study an external validation of resource use, probability of utilization, and cost parameters is chosen, because in the decision model these parameters are mainly based on expert estimates and national reimbursement catalogues, which are afflicted by a high degree of uncertainty. Effectiveness parameters, on the other hand, are based on high quality evidence from clinical trials and should be more valid. Also, effectiveness outcome is measured in QALYs in the PCa-model, which cannot be assessed in the AOK claims data set because clinical information is not included in sufficient detail to establish quality of life.

Resource use and costs are independently calculated in AOK claims data analysis and PCa model; outcomes of AOK claims analysis are not incorporated in the model. Instead PCa model's cohort and study design are replicated in AOK claims data analysis to achieve comparable results.

Furthermore, both input parameters (resource use and unit costs) and simulated outcomes (probability of utilization and per capita costs) of the PCa-model are validated in this study.

3.3 Claims data set, study design, and patient cohort

3.3.1 Dataset

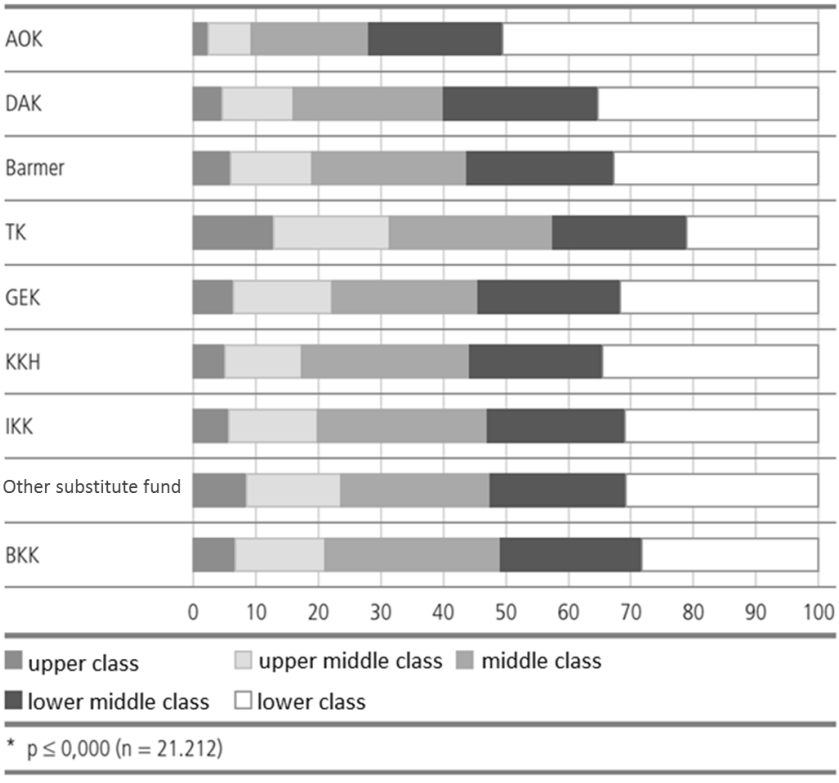
In the second step of the validation approach, a suitable claims dataset for the validation of the PCa decision model is selected. A dataset from a German SHI fund is useful for the validation of a decision model set in a German health care context with the perspective of the SHI scheme insurant population. The insurant population expected to be treated for localized PCa was estimated in the study planning phase, based on prevalence data for the general German population for 2 SHI funds, AOK Baden-Württemberg and Deutsche Angestellten-Krankenkasse (DAK). It was decided to perform model validation with data from AOK Baden-Württemberg, as the PCa-cohort was estimated to be larger than in DAK data (estimated number of incident PCa-cases AOK Baden-Württemberg: 3,817, DAK: 929).

AOK Baden-Württemberg is the largest SHI fund in the south-western German federal state of Baden-Württemberg with about 3.9 million insured individuals in 2014, which corresponds to 43% of the SHI scheme insured community in Baden-Württemberg. About 44% of insured individuals are mandatory members, 25% of insured individuals are retired, the remaining individuals are either family members or voluntarily insured at AOK Baden-Württemberg (119).
Historically, AOK insurance funds provided health insurance for blue collar workers with lower educational status and income compared to other SHI funds. This insurant structure of AOK still exists in alleviated form (Figure 3-1), although insured individuals are free to choose an insurance fund since the 1990s in Germany (120). Implications of insurant structure on generalizability of results are discussed in chapter 5.3.2.

AOK Baden-Württemberg provided access to data on all claims incurred between 2008 and 2011 in the following health care service categories: inpatient and outpatient care, pharmaceuticals, assistive technologies, physiotherapy, outpatient rehabilitation, work incapacity and sick pay. Additionally, basic insurant information such as age, gender, nationality, insurance status, date of death or date of termination of membership are provided. Earlier claims were not accessible because of a change in the database system which does not allow merging data before and after 2008. Claims after 2011 were not

available yet, at the time of data selection in July 2012. All claims are available on a patient level and can be merged by a personalized insurant identification number. Co-payments to medical services covered by SHI are included in the dataset, whereas patients' out-of-pocket payments for other services are not.

German data protection laws were considered during extraction and analysis of data and AOK Baden-Württemberg approved of the intended use of the data. An ethics committee was consulted regarding this study; ethics approval is not necessary as identification of individual persons is not possible in the dataset (see Appendix B for ethics committee statement).



Source: according to Gesundheitsmonitor 2008 (120)

Figure 3-1: Social status of insured individuals (%), by insurance fund

3.3.2 Study design

In the second step of the validation approach, also a study design for the AOK claims data is selected, which replicates model assumptions. Due to the secondary nature of the underlying data source, an observational study design is given where no form of

randomization into treatment groups or intervention is possible. To replicate the modeling approach a prospective, longitudinal study design is chosen, where a cohort of men diagnosed with early-stage PCa is followed from the point of treatment initiation (121). As in the model, the alternative treatment options RP and AS are considered. 4 consecutive years of AOK claims allows a maximum follow-up of 2.5 years, which in turn allows validation of the first 10 3-month cycles of the decision model by Koerber et al. (2014). Thus, only the part of the decision model representing treatment of localized disease is considered for validation, because follow-up is too short to include individuals with disease progression. The probability for disease progression in the first years after initial treatment, assumed in the model is extremely small (< 1%).

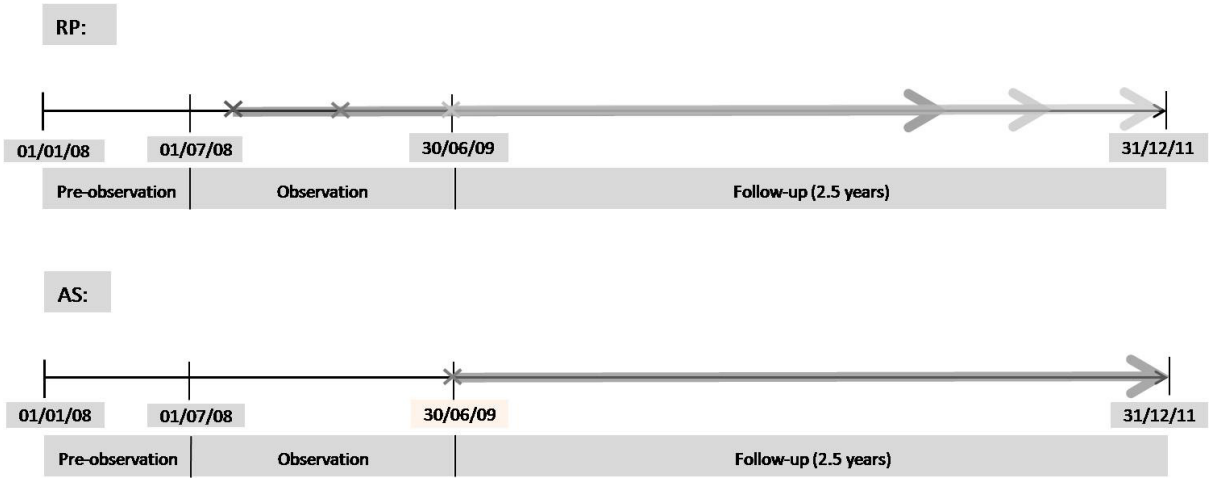


Figure 3-2: Study timeline (AOK)

The study period covers 4 years, from January 1st 2008 to December 31st 2011, and is divided into 3 sections (Figure 3-2):

1) The 6-month pre-observation period ranges from January 1st 2008 to June 30th 2008. In order to allow for co-morbidity adjustment of outcomes, this period before initial treatment is created as a basis to calculate a Charlson Co-morbidity Score (CCS) for each insured individual. CCS is calculated based on diagnoses coded before initial treatment, to ensure that diseases, which are complications of the treatment of interest (AS or RP), are not considered in the calculation (122).
2) In the 12-month observation period from July 1st 2008 to June 30th 2009 PCa-cases are identified and categorized into treatment groups. An observation period of 12 months is chosen, on the one hand, to identify a sufficient number of men treated with AS or RP to

analyze resource utilization and costs. On the other hand, the risk is decreased that men waiting for radical treatment are falsely classified as AS-patients.

3) The cohort is followed-up for a period of 2.5 years (follow-up period). In case of RP, follow-up time starts individually after the date of the initial surgery for each insured individual. Individuals under AS are followed for a fixed period from July 1st 2009 to December 31st 2011. One reason for this is that the starting point of AS cannot be established in the chosen cohort, because the time of initial PCa-diagnosis – when AS usually starts – cannot be identified in the AOK data. Thus, an artificial starting point for AS has to be created, which is the beginning of the observation period (July 1st 2008). Secondly, follow-up is not intended to start with the onset of AS. This is due to the selection process in this study, where men under AS have to be surveyed for at least 12-months (the observation period) to be included in the cohort. Thus, men under AS who die in this period are not considered for analysis. In the RP-group, on the other hand, a single event (the RP surgery) at some point in the observation period determines inclusion in the cohort, which in turn possibly includes men dying in this period. In conclusion, individuals under AS have, by definition of the selection strategy, a lower probability to die in the observation period. To account for this bias, follow-up of AS is offset by 12 months.

3.3.3 Cohort selection

Furthermore, in step 2 of the validation process the cohort is selected in AOK claims data.
In a first step all insured individuals with an ICD-coding of ‘C61 - Malignant neoplasm of prostate’ in any health care sector were identified in the complete claims of AOK Baden-Württemberg of the year 2008. For the identified individuals claims from all health care sectors of the years 2008 to 2011 were extracted. To comply with data protection laws the baseline dataset was pseudonymized, which includes deletion of all information that allows identification of individuals (insurant identification number, name, address, date of birth) and creation of a pseudonymized identification number for each individual. Furthermore, information which allows identification of single health care providers (e.g. hospitals, physicians, pharmacies) was deleted from the baseline dataset. The pseudonymized baseline

dataset was transferred from AOK Baden-Württemberg to Helmholtz Zentrum München for data analysis.

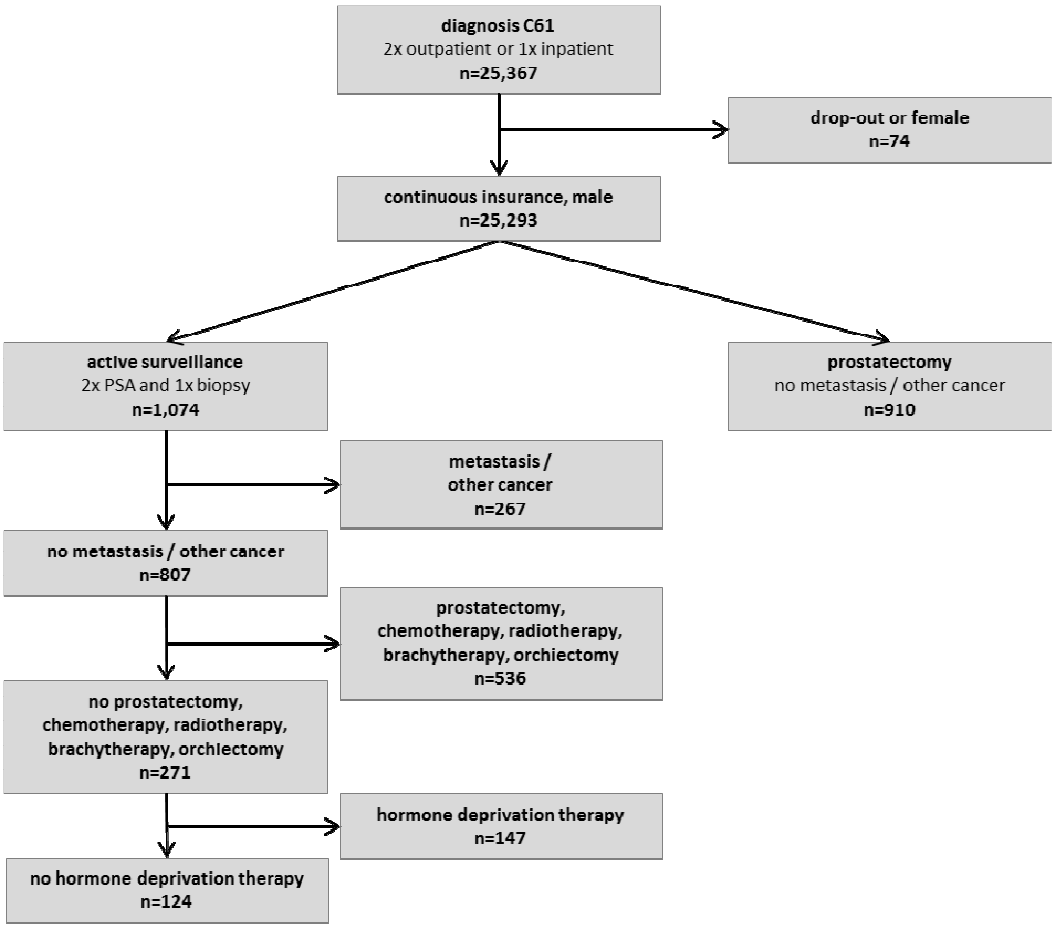


Figure 3-3: Cohort selection (AOK)

The first step of the cohort selection process (Figure 3-3) is to identify individuals with a validated PCa-diagnosis in the 12-month observation period. PCa diagnosis is validated in three ways. First, only men are considered with at least one inpatient or at least two outpatient diagnostic codes of C61 in different quarters of the observation period. Secondly, only main diagnostic codes determined at discharge are used in case of inpatient diagnoses, because these codes are most valid for hospital care (123). Thirdly, only codes with the additional diagnostic certainty code 'secured' (G, gesicherte Diagnose) are considered for outpatient diagnoses. In German outpatient care tentative diagnoses are usually added the code 'V: tentative diagnosis' or 'A: exclusion of the coded diagnosis'. When the diagnosis is clarified, the code 'G: secured diagnosis' is added. If the disease is cured the additional code is changed to 'Z: state after disease'. To include only patients in the cohort with a current

diagnosis of C61, all diagnoses with additional codes other than G are not considered for cohort selection. By applying the criterion of at least two diagnostic codes, consideration of tentative diagnoses is further minimized (124). 25,367 insured individuals apply to these criteria.

After exclusion of individuals not constantly enrolled in the study period or with female gender coding, the dataset includes 25,293 individuals.

Diagnostic ICD-10 GM codes, inpatient (OPS) and outpatient procedure (EBM) codes, as well as pharmaceutical prescriptions (coded by ATC codes) are used to define inclusion and exclusion criteria for cohort selection (Table 3-1 and Table 3-2).

Diagnosis	ICD-10 GM
Metastases	C77, C78, C79
Other cancer	C00-C14, C15-C26, C30- C39, C40-C41, C43-C44, C45-C49, C50, C51-C58, C60, C62, C63, C64-C68, C69-C72, C73- C75, C76, C80, C81-C96
PCa	C61

Table 3-1: Diagnostic codes for cohort selection (AOK)

Procedure	Inpatient (OPS)	Outpatient (EBM)	Pharmaceuticals (ATC)
PSA test	-	32351	-
Prostate biopsy	-	26341	-
RP	5-604	36276, 36277, 36287	-
EBRT	8-520, 8-521, 8-522, 8-523	25321	-
Chemotherapy	8-54	86512, 86514	-
Brachytherapy	8-524, 8-525	25333	-
Orchiectomy	5-622	-	-
ADT	-	-	L02AE

Table 3-2: Procedure codes for cohort selection (AOK)

Following the definition of AS in the PCa-model by Koerber et al., which is based on the surveillance recommendation of the DGU of 2014 (69), AS is defined by EBM codes of at least two PSA tests and at least one prostate biopsy during the observation period. DREs, which are recommended in a 3 monthly frequency, cannot be identified in SHI data, because these are covered by the urologic insurant lump sum and are, thus, not coded separately. To limit the cohort to early stage PCa, men with diagnoses on metastatic or any other kind of cancer disease are excluded. Additionally, patients undergoing any form of PCa-specific therapy other than AS (RP, EBRT, brachytherapy, chemotherapy, orchiectomy, and ADT) in the observation period are excluded to avoid misclassification of individuals waiting for radical treatment as AS cases. The inclusion and exclusion criteria mentioned above apply to the observation period only. In the follow-up period, men in the AS-group do not have to be surveyed as closely as defined in the observation period. They may move on to any other form of PCa-specific therapy or develop metastatic disease, as assumed in the decision model.

RP is defined by inpatient as well as outpatient procedure codes on open, laparoscopic, or robotic-assisted radical prostatectomy. Individuals with a diagnosis of metastatic or any other kind of cancer than PCa in the observation period are excluded. Other forms of PCa-specific treatment may be performed in addition to RP in the observation period, because inclusion criteria are not supposed to be so strict that treatment practice is not reflected in the cohort.

Application of above mentioned inclusion and exclusion criteria creates a cohort consisting of 124 individuals in the AS-group and 910 in the RP-group.

3.4 Relevant health technologies and costs

In step 3 of the validation, health technologies and corresponding costs relevant for PCa-treatment are defined. In this study, two different approaches for the definition of relevant health technologies in the AOK claims data are explored: an excess approach and a disease-related approach.

3.4.1 Excess approach

The excess analysis is a pragmatic approach to external validation of decision models based on claims data; it gets by with few assumptions in addition to study design and cohort selection in claims data, because no definition of disease-specific health care utilization is necessary. All health care utilization and corresponding costs accruing in the follow-up period are considered for analysis. When comparability of the AS- and the RP-group regarding patients' baseline characteristics can be ensured (e.g. by statistical methods described in the following), the difference in outcome is solely attributable to the difference in initial PCa-treatment (52).

However, the absolute outcome values (resource use or costs) of treatment strategies have no informative value for validation, because disease-specific costs which are considered in the PCa model are not represented. With the excess approach only the validation of incremental outcome measures is possible. All the same, excess analysis is useful for the validation, because it indicates whether AOK claims data estimate similar incremental outcomes as the decision model and whether it is reasonable to go to the time and effort to analyze AOK claims data on a disease-related basis.

3.4.2 Disease-related approach

In disease-related analysis only resource use of health technologies and corresponding costs that are relevant for the treatment of localized PCa are considered. Disease-related analysis allows validation of absolute values of resource use and costs of the 2 treatment strategies in addition to incremental outcomes. Also, comparison of utilization and costs of single health technologies is possible; by that not only total outcome of the model is validated but also outcome in different health care service categories. The validity of model outcomes might, for example, differ in in- and outpatient care. Furthermore, disease-related analysis

helps to explain difference in total outcome between model and claims data by difference in outcome of single health technology utilization or difference in unit cost estimates.

Health care utilization and corresponding costs of PCa-treatment are identified in the 2.5 years of follow-up after initial treatment. Only treatment relevant for localized PCa is considered, whereas treatment of metastatic disease and palliative therapy are not considered; only the part of the PCa-model is validated which evaluates treatment of localized PCa.

The definition of health technologies relevant for the treatment of localized PCa in claims data analysis follows the definition of the model very closely. Whenever codes for the calculation of resource use and costs are explicitly mentioned in the model, these codes are used in claims data analysis as well. In in- and outpatient care DRG and EBM codes are reported in the model to identify health technology utilization. In all other health care service categories no specific codes of health technologies are available; in these cases corresponding codes are identified in claims data in consultation with the modeling team. Costs are reported in Euro (€) as in the decision model. Discounting of costs is not necessary due to the short study period.

Some deviations from the definition of disease-related costs in the PCa-model are necessary; mainly to ensure that only PCa-relevant treatment is included in claims data analysis. For one, inpatient treatment is defined by OPS codes in claims analysis rather than by DRGs as in the PCa-model, because OPS codes are more specific. For example, to identify the utilization and costs of prostatectomy without complications, a search for a combination of DRG M01B ‘Major procedure at the male pelvic organs without complications’ and OPS 5-604 ‘Radical prostatectomy’ is used. The DRG code alone is too unspecific to identify RP in claims data, because code M01B is also used to bill procedures other than RP.

Secondly, in claims data analysis criteria for the temporal link between ICD, EBM, or OPS codes are defined. For example, treatment of IC by a general practitioner (GP) is defined by generic EBM codes, because no specific codes for treatment of IC are available. Thus, to consider this treatment relevant for the analysis a diagnosis of IC in the same quarter is required.

Table 3-3 lists all health technologies that are assumed to be relevant for the treatment of localized PCa by health care service category; corresponding codes for the selection of disease-related technologies in claims data are provided.

Regarding costs of the identified health technologies, generally, all costs are considered that are relevant for the perspective of the SHI scheme insured community, which includes reimbursement by the insurance fund as well as out-of-pocket payments of insured individuals. DRG codes for hospital episodes, in which relevant procedures are performed, are used to estimate costs in claims data analysis. In case of outpatient treatment, costs are usually reported on a quarterly case basis by EBM codes. To make costs comparable to the decision model, costs of single EBM procedures are considered and not the costs of the whole case.

To allow direct comparison of utilization probability and costs estimated in claims data with outcomes of the decision model, single health technologies are summarized in cost groups according to the cost groups defined in the decision model (Table 3-4).

Treatment	Code
Inpatient	
Prostatectomy w/ o complications	OPS 5-604 + DRG M01B
Prostatectomy with complications	OPS 5-604 + DRG M01A/ A13C/E / M38Z
Treatment stricture (RP)	OPS 5-580 / ICD N35.0
Penis prosthesis (ED)	OPS 5-649.5 -- 5-649.9
Artificial urethral sphincter (IC)	OPS 5-597
Sling surgery (IC)	OPS 5-596
Treatment urosepsis (AS)	DRG T60/ T61/ L63
Transurethral prostate resection for BPH	OPS 5-601
Outpatient	
Follow-up RP: PSA testing (urologist)	OPS 5-604 + (EBM 26211/ 26212 + EBM 32351 in same quarter)
Treatment ED (urologist)	ICD N48.4/ F52.2 + EBM 26211/ 26212
Treatment ED (GP)	ICD N48.4/ F52.2 + EBM 3111/ 3112
Treatment IC (urologist)	ICD N39.3/ N39.4/ R32/ F98.0 + EBM 26211/ 26212
Treatment IC (GP)	ICD N39.3/ N39.4/ R32/ F98.0 + EBM 3111/ 3112
EBRT after AS (radiotherapist)	EBM 34360 + 25342 + 40840 + 25321/ 25322/ 25323
AS: PSA testing and biopsies (urologist)	EBM (32351 + 26211/ 26212 in same quarter) + 26341 in same year

Table 3-3: Codes for selection of disease-related resource use and costs (AOK)

Treatment	Code
Pharmaceuticals	
Antibiotics before prostate biopsy	EBM 26341 + ATC J01MA02 (max. 30 days before biopsy)
Treatment BPH: alpha blockers, 5α-reductase inhibitors	ICD N40 + ATC G04CA/ G04CB
Treatment ED: PDE5 inhibitors	ICD N48.4/ F52.2 + ATC G04BE
Assistive technologies	
Vacuum pump, rings (ED)	Position no. 99.27.02/ 99.27.01
Incontinence aids (e.g. diapers) (IC)	Position no. 15.25.01 -- 09/ 15.25.15
Physiotherapy	
Physiotherapy pelvic floor (IC)	ICD N39.3/ N39.4/ R32/ F98.0 + position no. SO2

Table 3-3: Codes for selection of disease-related resource use and costs (AOK) (continued)

Cost group model	Treatment claims data
ED	Outpatient treatment ED (urologist)
	Outpatient treatment ED (GP)
	Pharmaceutical PDE5 inhibitors
	Vacuum pump, rings
IC	Outpatient treatment IC (urologist)
	Outpatient treatment IC (GP)
	Incontinence aids
	Physiotherapy pelvic floor
RP w/o complications	Prostatectomy w/ o complications
AS	PSA testing and biopsies (urologist)
	Antibiotics before prostate biopsy
EBRT	EBRT after AS (radiotherapist)
Urosepsis	Inpatient treatment urosepsis
TURP	Transurethral prostate resection for BPH
Surgery IC	Artificial urethral sphincter surgery
	Sling surgery
Surgery ED	Penis prosthesis surgery
RP with complications	Prostatectomy with complications
BPH	Alpha blockers, 5α-reductase inhibitors
Stricture	Inpatient treatment stricture
Monitoring RP	Follow-up RP: PSA testing (urologist)
Total	Sum of all cost groups

Table 3-4: Assignment of health technologies to cost groups

3.5 Statistical methods for claims data analysis

In step 4 of the validation of the PCa-model, statistical methods are selected which allow estimation of outcomes in AOK claims data comparable to outcomes of the decision model. These include matching of individuals in the AS- and the RP-group as well as regression analysis of effect and cost outcomes.

3.5.1 Matching

Matching is a statistical technique which, in part, corrects the treatment selection bias induced by non-randomization. Each AS-subject is individually paired with a RP-subject for variables that might confuse the comparison. Matching is only useful, though, for variables that are strongly related to both the treatment and outcome of interest (28). Age is chosen as the matching variable because it is strongly related to the treatment decision; AS is recommended for patients with a life expectancy greater than 10 years. Regarding health outcome, rates of ED and IC as well as corresponding health care utilization are also correlated with age (125). Furthermore, matching of the claims data cohort makes the comparison with the decision model more valid, because the model assumes that there is no difference in age distribution between treatment groups.

Treatment groups are matched by +/- 2 years of age and in a ratio of 1 AS-subject to 2 RP-subjects to account for the relatively smaller group of patients under AS. Especially for rare events the strength of the study can be increased by having more RP-subjects than AS-subjects in the cohort (28).

After matching, the AS-group includes 107 individuals and the RP-group 214, respectively.

All statistical analyses are performed with the software package SAS, version 9.3.

3.5.2 Descriptive analysis of patient characteristics

Patients' age and CCS as well as prevalence of ED, IC, and benign prostate hyperplasia (BPH) are calculated at baseline before and after matching. For calculation of CCS the co-morbidity group 'cancer' is set to 0, because diseases other than PCa are of interest for adjustment. Definition of ED, IC, and BPH on the basis of ICD-10 GM codes is shown in tTable 3-5.

Diagnoses coded in in- and outpatient care are used and the same criteria for validity of diagnoses as for cohort selection are applied: only inpatient discharge diagnoses and outpatient diagnoses with the additional coding ‘secured’ are used.

Comparison of patient characteristics in the 2 treatment groups allows estimation of the usefulness and performance of matching. Furthermore, characteristics of the claims data cohort and model cohort are compared based on the descriptive analysis.

Diagnosis	ICD-10 GM
Benign prostate hyperplasia (BPH)	N40
Erectile dysfunction (ED)	N48.4, F52.2
Urinary incontinence (IC)	N39.3, N39.4, R32, F98.0

Table 3-5: Diagnostic codes for effect analysis (AOK)

3.5.3 Effect analysis

Rates of short- and long-term (continuous diagnosis > 90 days after initial treatment) ED and IC as well as rates of BPH are estimated per treatment group in the follow-up period of 2.5 years. These rates are relevant for the analysis of disease-related resource utilization: identification of individuals with complications of PCa treatment is required for the identification of utilization of PCa-relevant health technologies, which are not specifically coded as treatment of PCa and are not linked to diagnostic information in claims data (e.g. outpatient treatment of IC).

‘Incident’ ED and IC diagnoses are considered in effect analysis, which means that only subjects are included which do not have a diagnosis of IC or ED before the initial RP surgery or, in case of AS, before the beginning of follow-up. This is important for the comparison with the model, because it is assumed in the model that individuals in both treatment groups do not have ED, IC, or BPH before initial treatment. The definition of diagnostic codes for effect analysis has been described previously (Table 3-5).

To address bias in complication rates due to difference in baseline co-morbidity in the 2 treatment groups, a logistic regression model is used to adjust for CCS. Logistic regression models are suitable to predict the proportion of subjects with a complication of interest by using a logit transformation. The odds ratio of an individual having the complication is predicted by the logistic regression model (28).

The method of recycled predictions is applied to estimate mean complication rates per treatment strategy (the probability of having the complication of interest) in addition to odds ratios. Recycled predictions can impart the scale of group differences better than the regression coefficients alone, because probabilities are easier to interpret than odds ratios (126). To achieve this, the predicted probabilities of complications are at first estimated under the assumption that all individuals in the sample are treated with RP and all other variables remain the same; secondly, the predicted probabilities of complications are estimated assuming all in the sample are under AS and, again, all other variables remain the same. Finally, the averages of these two predicted probabilities are compared and a new estimate of the difference between AS and RP is produced (126).

95% CIs of mean complication rates per treatment strategy and incremental complication rates are calculated via non-parametric bootstrapping with 1,000 replications. The bootstrapping method refers to a random sampling technique with replacement. The mean costs are calculated for each re-sample and these re-samples make up the empirical estimate of the distribution of mean costs. This allows calculation of accuracy measures – such as confidence limits – for estimated means, in samples with small observation numbers, where the assumption of asymptotic normality of the estimator is questionable. To calculate 95% CIs, the 2.5 and 97.5 percentile values of the sampled distribution are taken to represent the endpoints of the interval (127, 128). P-values indicating significant differences in complication rates between treatment strategies are estimated via bootstrap hypothesis testing; p-values less or equal to 0.05 are considered statistically significant.

3.5.4 Excess analysis of resource use and costs

For the excess analysis all health care utilization incurred at the SHI fund are considered, irrespective of the relevance for PCa treatment. As in the decision model, utilization of inpatient and outpatient care, pharmaceuticals, physiotherapy, and assistive technologies are considered and corresponding costs are summarized to total costs per treatment strategy. The difference in costs reflects the excess costs of AS compared to RP, assuming that matching and adjustment in regression analysis control for treatment selection bias and the difference is solely attributable to the treatment.

All direct medical costs are considered that are relevant for the perspective of the SHI scheme insured community, according to the German SGB V (§ 35b (1) SGBV) (129). Co-morbidity adjusted costs are estimated per treatment group by a generalized linear model (GLM) with a gamma distribution and log link to account for the typically skewed distribution of cost data. To individuals with zero costs a small amount of €1 is assigned to include them in the analysis (130). When individuals with zero costs account for more than 10% of the cohort a two-part model is used: at first the probability of health care expenditure is predicted with a logistic regression model. In a second step costs are estimated by a GLM, as described previously, conditional for nonzero costs. To derive unconditional per capita costs the probability of expenditure is multiplied by the predicted conditional costs (131). The GLM reports the cost difference between the AS- and the RP-strategy as a percentage. Additionally, recycled predictions method is applied to estimate absolute values of mean costs per strategy and difference in costs (52, 126). All costs are rounded to the nearest € and inflation is not considered due to the short study period.

For per capita costs and difference in costs 95%-CIs are calculated via a non-parametric bootstrap approach based on 1,000 replications, by taking the 2.5 and 97.5 percentile values to represent the endpoints (128, 132). P-values indicating significant differences in costs are estimated via bootstrap hypothesis testing; p-values less or equal to 0.05 are considered statistically significant. The CCS is included in the regression models as a continuous variable to adjust for difference in co-morbidity in the treatment strategies (133, 134). To additionally estimate the influence of ED and IC on treatment costs, complication is included in the regression models as a binary variable. Extended models, with an interaction between treatment strategy and CCS, do not improve model fit.

3.5.5 Disease-related analysis of resource use and costs

In disease-related analysis 4 different outcome measures are estimated: resource use of treated individuals, unit costs, probability of technology utilization, and per capita costs accumulated during follow-up.

- Resource use is reported as the quantity of utilization per person of a single health care technology, averaged over all individuals treated with the technology of interest,

in the follow-up period. RP surgery is, for example, expected to be performed only once per person, whereas treatment with EBRT involves several radiation sessions.

- Unit costs describe the mean costs of one procedure, hospital episode, or pharmaceutical description.
- The absolute and relative number (probability) of individuals in the cohort utilizing a health technology is reported for single treatments and for cost groups.
- Per capita costs accumulating during follow-up are reported for cost groups with standard deviation (STD) as well as median and interquartile range, to give an indication of the distribution of costs. Especially in cost groups where a small number of individuals causes costs, extremely skewed distribution of costs is possible; in that case STD has no meaningful interpretation (28).

To adjust costs for difference in co-morbidity a GLM with a gamma distribution and log-link including CCS as a continuous variable is used, as described in chapter 3.5.4. Non-parametric bootstrapping with 1,000 replications is applied to estimate 95% CIs for mean costs of all cost groups. P-values indicating significant differences in costs between treatment strategies are estimated via bootstrap hypothesis testing; p-values less or equal to 0.05 are considered statistically significant. A random number seed is set during bootstrapping to allow replication of results and to generate comparable results for all cost groups.

3.6 Changes to the decision model

In step 5 of the validation approach, model assumptions are changed and additional analyses are conducted to make the comparison of outcomes with claims data more valid. Such a change of assumptions is possible because access to the PCa-model in Treeage is granted by Koerber et al. (2014).

3.6.1 Adaptation of model assumptions to claims data

Change of model assumptions includes adaptation of the runtime of the decision model to the follow-up period of the AOK claims data analysis of 2.5 years, which corresponds to 10 cycles of 3 months in the model. A longer follow-up or lifetime perspective could not be created in the claims data analysis, because claims are only available for a period of 4 consecutive years.

Also, it is not possible to replicate the age structure of the model cohort in claims data. If inclusion in the claims data cohort requires that all patients are 65 years of age at treatment initiation as in the decision model, treatment groups in claims data would become very small. Adaptation of the age structure of the model to the claims data cohort is necessary; consequently the mean age of the SHI cohort at treatment initiation (70 years) is assumed in the PCa-model. Mortality rates and the probability to be treated with either RP or EBRT following AS – which are the only relevant age-dependent probabilities – are adapted accordingly in the PCa-model.

Furthermore, discounting of costs is not performed in claims data analysis, due to the short period of follow-up. The PCa-model as published by Koerber et al. (2014) assumes an annual discount rate of 3% for costs, which is set to 0% for the purpose of this validation.

3.6.2 Additional analyses

For the model validation it is interesting to compare not only the probability and costs of all PCa-relevant health care utilization combined but also of single treatments. It is especially interesting in this study whether the validity of model parameters is higher in inpatient care, where OPS codes allow specific identification of procedures and DRG reimbursement provides valid cost estimates, than, for example, in outpatient care where coding of procedures via EBM codes is less specific.

To simulate health care utilization and costs of single treatments in the decision model, additional analyses are necessary, because the decision model reports only total costs per treatment arm. Such additional analyses are conducted without changing the model structure.

To compare utilization probability in the decision model with utilization probability estimated in claims data analysis, a microsimulation (also referred to as first-order Monte Carlo simulation) is performed in Treeage Pro 2013. Microsimulation pictures variability due to the alternative pathways of simulated individuals through the decision model and is similar to the population variability of outcome in the claims data (135). A cohort of 321 men in each treatment arm is followed individually through the model.
To estimate probability of utilization per cost group the microsimulation is run for each cost group separately, each with all other cost groups set to 0. The utilization probability of health technologies in each cost group is estimated as the number of individuals with costs > 0.
Averaging the costs over the 321 patients gives the overall estimate of the mean costs in each treatment arm. Simulated outcomes are exported from Treeage and imported to SAS for analysis of mean costs, STD, median, and interquartile range.

To account for parameter uncertainty in cost estimates probabilistic sensitivity analysis, using second-order Monte Carlo simulation with 1,000 replications, is performed in Treeage Pro 2013 for each cost group and total costs individually. Monte Carlo simulation allows simultaneously sampling from each cost parameter distribution (135). Because the precision of simulation point estimates depends on the number of replications, a consistent approach with 1,000 replications is used for model and claims data (22). Monte Carlo simulation samples are exported to SAS for cost analysis. Mean costs estimated by each sample are averaged over all 1,000 replications to calculate 95% CIs by taking the 2.5 and 97.5 percentile values to represent the endpoints of the interval (135).
A random number seed is set in Monte Carlo simulation as well as in microsimulation to ensure that the results are replicable from one run to the next and to make outcomes comparable between cost groups.

3.7 Comparison between model and claims data

In step 6 of the validation of the PCa-model based on AOK claims data, methods for the comparison of model and claims data outcomes are selected.

3.7.1 Input parameters: resource use and unit costs

Input parameters of resource use and unit costs included in the decision model are compared with resource use and unit costs estimated in claims data analysis. Difference in resource use and unit costs is described, whereas treatments which have a presumably high influence on total difference are especially highlighted.

3.7.2 Simulation outcome: probability of utilization and per capita costs

The probability of utilization of health technologies summarized in cost groups is displayed graphically in form of bar charts, directly comparing outcomes of unadjusted claims data analysis and microsimulation of the PCa-model. Bar charts are created for the AS- and the RP-group separately. Differences between outcomes are described, focusing on results that are most relevant to explain differences in total costs.
Unadjusted mean costs estimated in claims data analysis are compared with outcomes predicted by microsimulation of the decision model. Distribution of costs is displayed graphically in form of histograms contrasting estimates of claims data analysis with estimates of model microsimulation. Histograms show results separately for the AS- and the RP-group.

To test agreement between estimates of AOK dataset and model simulation statistically, the difference between incremental costs estimated in adjusted claims data analysis and incremental costs predicted by the decision model in Monte Carlo simulation is calculated; the difference in incremental costs is calculated for all cost groups and for total costs. Calculation of differences is based on the 1,000 sampled means of bootstrapping and Monte Carlo simulation, accordingly; each bootstrapped sample is assigned to a sample generated in Monte Carlo simulation and the difference in costs is then calculated for each of this samples. Based on the 1,000 sampled differences a mean difference is estimated.

Bootstrapping of outcomes generates a non-parametric distribution of the cost differences. For mean differences 95% CIs are calculated, taking the 2.5 and 97.5 percentile values of the bootstrap generated distribution to represent the endpoints.

The research hypothesis in this study is that the model is able to simulate outcomes observed in claims data. Consequently, incremental costs estimated in claims data are expected to be the same as incremental costs simulated in the model; the difference in incremental costs of model and claims data is expected to be 0. In hypothesis testing the research hypothesis – which one wants to prove – is usually expressed in the alternative hypothesis (H1). The null hypothesis (H0) is the negation of the research hypothesis. H0 is rejected, if the probability that the observed data could have been obtained with H0 being true (the p-value), is lower than a predefined significance level, usually 0.05. Following this, H0 would have to state in this study that the cost difference between claims data and model is unequal; a test of this hypothesis is, however, statistically not possible (28).
Consequently, the test is formulated in the usual manner:

$$H0: \text{incremental costs}_{AOK} = \text{incremental costs}_{model}$$

$$H1: \text{incremental costs}_{AOK} \neq \text{incremental costs}_{model}$$

H0 is tested based on simulated samples via bootstrap hypothesis testing (136). When 0 is included in the 95% CI or the p-value is greater than 0.05, H0 is not neglected, but this does not mean that H0 is accepted or proven. P-values lower than 0.05 indicate a rejection of H0 and a significant difference between incremental costs of claims data and model (28, 29).
Additionally to hypothesis testing, the distribution of sampled incremental costs of claims data analysis and decision model is displayed in histograms for all cost groups and total costs. An overlapping of the histograms indicates an agreement between incremental costs of claims data analysis and Monte Carlo simulation.

3.8 Sensitivity analyses

In step 7 of the validation approach, sensitivity analyses are selected which test the influence of claims data and model assumptions on outcome.

3.8.1 Incident PCa-cases in claims data analysis

In a sensitivity analysis only incident PCa cases are included to estimate the influence of the cohort selection on difference in costs in claims data analysis. A more valid comparison with the PCa-model is possible, because the model assumes incidence of disease. Incidence of disease might be especially relevant in the AS group; it can be ensured that only individuals are included in the analysis who started AS in the observation period. That way, time under treatment is the same in the RP- and the AS-group.

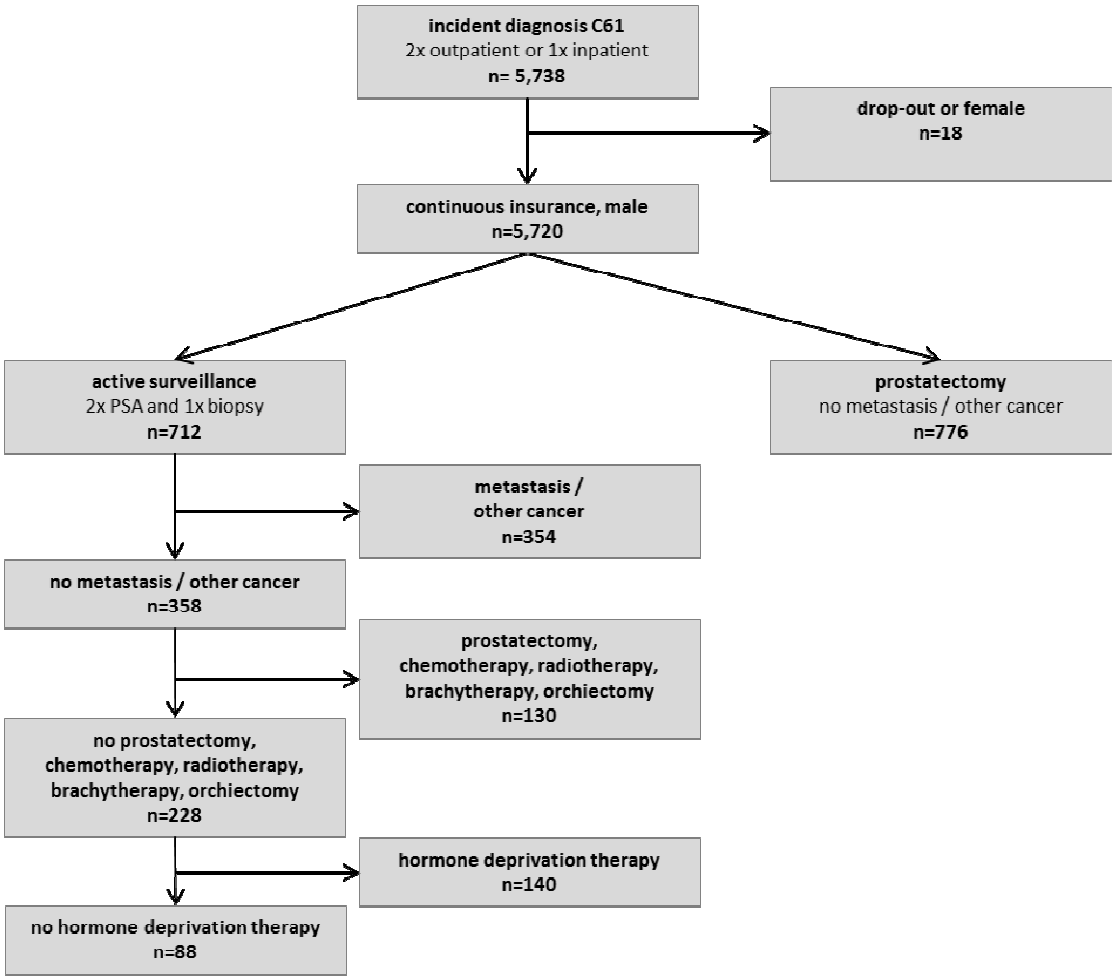


Figure 3-4: Cohort selection of incident PCa-cases (AOK)

Individuals with no coding of PCa, identified by ICD-10 GM code C61, in the pre-observation period and at least one inpatient or two outpatient diagnostic codes in the observation period are considered for analysis. After exclusion of individuals with incorrect gender coding or dropping-out during the study period, 5,720 individuals remain in the incident cohort (Figure 3-4). Applying the inclusion and exclusion criteria for treatment groups described previously, leaves 88 individuals in the AS-group and 776 individuals in the RP-group. After matching for age in a ratio of 1:2, the AS-group consists of 64 individuals and the RP-group of 128, respectively.

Excess costs and total disease-related costs are estimated in claims data analysis per treatment strategy. Adjusted and bootstrapped incremental costs are compared with outcomes of Monte Carlo simulation of the PCa-model as described in the base analysis (chapter 3.7).

3.8.2 Age distribution in the decision model

In a second sensitivity analysis the influence of the age at treatment initiation, assumed in the decision model cohort, on agreement with claims data estimates is tested. As described previously, age at treatment initiation and corresponding age-dependent variables are changed to 70 years of age, to reflect the mean age of the claims data cohort. Age in the claims data cohort, however, varies: the 2.5 percentile is 51 years of age and the 97.5 percentile is 79, respectively.

Age at treatment initiation is relevant in the model because treatment with RP or EBRT after AS depends on age; before 72 years of age RP is recommended, older men are treated with EBRT. If every individual in the base case model starts AS at the age of 70 and follow-up continues for 2.5 years, the probability of being treated with EBRT in follow-up is much lower than being treated with RP. As EBRT is less costly than RP, costs of AS-strategy may be overestimated.

To test this influence on agreement of claims data and model, the PCa-model is once run with an age at treatment initiation of 51 years and once with 79 years, respectively. In the first analysis, with mean age of 51 years of age, no patient in the AS-group is assumed to be treated with EBRT and in the second analysis, with 79 years of age, all patients in the AS-group are treated with EBRT in follow-up.

Difference in incremental costs is compared as described in base analysis (chapter 3.7).

4 Results

4.1 Patient cohort in claims data

Results of the baseline characteristics of the patient cohort in the claims data set, which is defined in step 2 of the validation approach, are presented in the following.

Mean age at baseline (before the initial treatment) in the claims data cohort is 69 years (STD 6.80) in the RP-group and 70 years (STD 7.13) in the AS-group after matching (Table 4-1). Mean CCS, representing co-morbidity in the cohort, is 0.11 in the RP-group and 0.19 in the AS-group. STD of 0.63 in both treatment strategies is comparatively high, which indicates a skewed distribution of co-morbidity in the treatment groups; a small number of men have high co-morbidity scores compared to the remaining individuals in the cohort with CCS of 0.

Prevalence of ED and IC is slightly higher in the RP-group than in the AS-group at baseline after matching (ED: 0.11 vs. 0.05, IC: 0.05 vs. 0.03). Prevalence of BPH is considerably higher in the RP-group with 0.77 as opposed to 0.68 in the AS-group. From a medical point of view, this result can be explained by the fact that men suffering from symptoms of BPH tend to surgery for treatment of PCa because BPH is cured by RP as well.

	Before matching		After matching	
	RP	AS	RP	AS
Total (n)	910	124	214	107
Age (m, STD)	66 (6.64)	70 (8.31)	69 (6.80)	70 (7.13)
CCS (m, STD)	0.13 (0.71)	0.19 (0.62)	0.11 (0.63)	0.19 (0.63)
ED (n, p)	79 (0.09)	5 (0.04)	24 (0.11)	5 (0.05)
IC (n, p)	18 (0.02)	5 (0.04)	10 (0.05)	3 (0.03)
BPH (n, p)	511 (0.56)	83 (0.67)	164 (0.77)	73 (0.68)

m = mean, n = number, p = proportion

Table 4-1: Baseline characteristics before and after matching (AOK)

Comparison of baseline results before and after matching shows that the RP-group gets older and, in line with that, prevalence rates of ED and IC as well as BPH increase. The AS-group, on the other hand, features predominantly the same baseline characteristics before and after matching. Overall, the RP- and the AS-group show a satisfactory concordance in baseline characteristics after matching, except for co-morbidity. To account for different co-morbidity structures in the treatment groups, CCS is adjusted for in regression models in the following.

In the claims data cohort 14 out of 321 individuals (4.4%) die during follow-up; the mortality rate is, however, greater in the AS-group (7.5%) than in the RP-group (2.8%).

4.2 Effect analysis of claims data

Results of the analysis of complication rates, described in step 4 of the validation approach, are presented in the following.

Unadjusted analysis of short- and long-term ED and IC rates in the follow-up period indicates that after RP complication rates are higher than after AS (0.48 vs. 0.12) (Table 4-2). However, the number of individuals with some complications is extremely small in the AS-group. For example, only 2 individuals suffer from a combination of ED and IC. Even though this result is valid from a medical point of view, it is difficult to estimate valid costs based on such a small number of individuals; this should be considered when assessing the validity of disease-related cost estimates in the AS-group.

Complication	RP (n=214)	AS (n=107)
ED < 90 d	31 (0.15)	6 (0.06)
ED > 90 d	26 (0.12)	6 (0.06)
IC < 90 d	92 (0.43)	9 (0.08)
IC > 90 d	70 (0.33)	9 (0.08)
ED & IC < 90 d	20 (0.09)	2 (0.02)
ED & IC > 90 d	16 (0.08)	2 (0.02)
Total (ED or IC or both)	103 (0.48)	13 (0.12)

d = days, n = number

Table 4-2: Mean complication rates – unadjusted (AOK)

Complication	RP		AS	
	m	95% CI	m	95% CI
ED < 90 d	0.15	0.10 to 0.20	0.06	0.02 to 0.10
ED > 90 d	0.12	0.08 to 0.17	0.06	0.02 to 0.10
IC < 90 d	0.44	0.36 to 0.50	0.08	0.04 to 0.14
IC > 90 d	0.33	0.27 to 0.40	0.08	0.04 to 0.14
ED & IC < 90 d	0.10	0.06 to 0.14	0.02	0.00 to 0.05
ED & IC > 90 d	0.08	0.08 to 0.12	0.02	0.00 to 0.05
Total (ED or IC or both)	0.49	0.42 to 0.55	0.12	0.07 to 0.19

CI = confidence interval, d=days, m = mean

Table 4-3: Mean complication rates – adjusted (AOK)

When adjusting for co-morbidity, results change very little. After RP rates of short- and long-term IC (0.44, CI [0.36; 0.50] and 0.30, CI [0.27; 0.40]) as well as short-term ED (0.15, CI [0.10; 0.20]) are significantly higher than in the AS-group (0.08, CI [0.04; 0.14]; 0.08, CI [0.04; 0.14]; 0.06, CI [0.02; 0.10]) (Table 4-3). Overall, total complication rates are significantly lower in the AS-group than in the RP-group (-0.37, CI [-0.45; -0.27], p< 0.0001) (Table 4-4).

Complication	AS - RP		
	m	95% CI	p-value
ED < 90 d	-0.09	-0.15 to -0.02	**0.008**
ED > 90 d	-0.07	-0.13 to 0.00	0.050
IC < 90 d	-0.35	-0.43 to -0.26	**< 0.0001**
IC > 90 d	-0.25	-0.33 to -0.16	**< 0.0001**
ED & IC < 90 d	-0.08	-0.12 to -0.03	**< 0.0001**
ED & IC > 90 d	-0.06	-0.10 to -0.02	**0.006**
Total (ED or IC or both)	-0.37	-0.45 to -0.27	**< 0.0001**

CI = confidence interval, m = mean

Table 4-4: Difference in complications rates – adjusted (AOK)

4.3 Excess analysis of claims data

In the following results of the excess analysis of claims data are presented; definition of health technologies relevant for excess analysis and statistical methods for analysis are described in steps 3 and 4 of the validation approach.

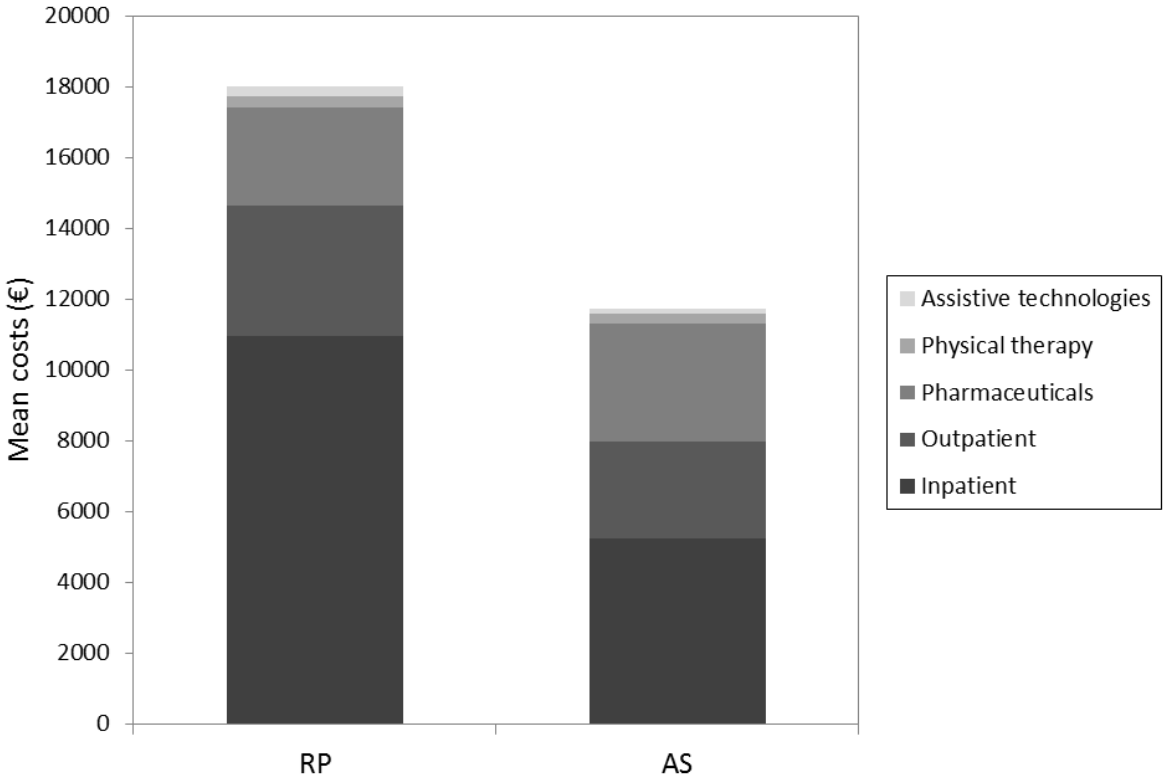


Figure 4-1: Per capita total costs (€), by treatment strategy and health care service category – unadjusted (AOK)

	RP (n=214)	AS (n=107)	AS - RP
Inpatient	10,964	5,227	-5,737
Outpatient	3,668	2,750	-918
Pharmaceuticals	2,774	3,321	547
Physical therapy	331	271	-60
Assistive technologies	281	139	-142
Total costs	18,018	11,708	-6,310

Table 4-5: Per capita total and excess costs (€), by treatment strategy and health care service category – unadjusted (AOK)

Figure 4-1 shows unadjusted per capita costs for the two treatment strategies grouped by inpatient and outpatient treatment, pharmaceuticals, physiotherapy, and assistive technologies. AS incurs lower mean costs than RP in all health care service categories, except in case of pharmaceutical prescriptions (Table 4-5).

Co-morbidity adjusted analysis confirms unadjusted excess cost analysis (Table 4-6 and Table 4-7). Total costs of AS (€11,933, CI [9,430; 14,554]) are significantly lower than total costs of RP (€18,544, CI [16,867; 20,660]) by €-6,611 (CI [-9,734; -3,547], p<0.0001).

Comparison of adjusted costs in single health care service categories displays that AS has significantly lower mean inpatient (€-5,845, CI [-7,632; -3,895], p<0.0001) and outpatient costs (€-961, CI [-1,622; -361], p=0.002) as well as costs for assistive technologies (€-141, CI [-230; -50], p=0.006) than RP.

Inclusion of complication as a binary variable into the regression models does not change cost differences between treatment strategies. Concerning total costs there seems to be no significant cost difference between individuals with and without complications. In case of assistive technologies a 30% increase in costs for individuals with complications is estimated (p=0.018).

	RP		AS	
	m	95% CI	m	95% CI
Inpatient	11,123	10,157 to 12,308	5,278	3,718 to 6,977
Outpatient	3,751	3,342 to 4,276	2,790	2,360 to 3,295
Pharmaceuticals	2,893	2,267 to 3,776	3,480	2,718 to 4,490
Physical therapy	346	236 to 491	288	182 to 428
Assistive technologies	281	223 to 343	140	83 to 212
Total costs	18,544	16,867 to 20,660	11,933	9,430 to 14,554

m = mean

Table 4-6: Per capita total costs (€), by treatment strategy and health care service category – adjusted (AOK)

	AS - RP		
	m	95% CI	p-value
Inpatient	-5845	-7,632 to -3,895	**< 0.0001**
Outpatient	-961	-1,622 to -361	**0.002**
Pharmaceuticals	587	-556 to 1,718	0.274
Physical therapy	-58	-214 to 114	0.460
Assistive technologies	-141	-230 to -50	**0.006**
Total costs	-6,611	-9,734 to -3,547	**< 0.0001**

Table 4-7: Excess costs (€), by health care service category – adjusted (AOK)

4.4 Disease-related analysis of claims data

Results of disease-related analysis are presented in the following; disease-related health care utilization and methods for disease-related analysis are described in steps 3 and 4 of the validation approach.

4.4.1 Resource use and unit costs

The average quantity of resource use of treated individuals in the follow-up period and unit costs of disease-related health technologies are compared for the RP- and the AS-group in Table 4-8.

When a health technology is not utilized in one of the treatment groups, no resource use and unit costs can be estimated in claims data analysis. For example, no unit costs can be estimated for EBRT as primary treatment in the RP-group because radiation after surgery is performed in case of disease progression and not as primary treatment of localized PCa.
Inpatient treatment is predominantly utilized once; outpatient treatment is utilized several times. Follow-up after RP surgery by an urologist is performed about 2.5 times more on individuals in the RP-group (8.62) than on individuals in the AS-group treated with RP (3.50). This could, however, be due to a shorter period of follow-up after RP in the AS-group; follow-up is overall limited to 2.5 years after the initial treatment. Resource use of remaining health technologies is comparable between treatment groups, if the health technology is utilized at all.

Highest unit costs are estimated for RP surgery with complications (€10,141), artificial urethral sphincter surgery (€11,732), and penis prosthesis surgery (€7,586). However, all of these surgeries are performed in a small number of patients (number of treated individuals: 41, 1, and 1, respectively) and each in only one of the two treatment groups; thus, unit cost estimates may not be representative for the cohort (Table A-1 appendix).
If a health technology is utilized, unit costs are comparable between treatment groups. Small differences between the AS- and the RP-group in unit costs may arise because of annually varying reimbursement rates. In case of inpatient treatment, different length of stay or additional reimbursement, of for example intensive care, can explain small differences, as well.

Treatment	Resource use p.p.		Unit costs (€)	
	RP	AS	RP	AS
Inpatient				
Prostatectomy w/ o complications	1.00	1.00	6,417	6,295
Prostatectomy with complications	1.00	0.00	10,141	0
Treatment stricture (RP)	1.10	1.00	4,298	4,824
Penis prosthesis (ED)	1.00	0.00	7,586	0
Artificial urethral sphincter (IC)	0.00	1.00	0	11,732
Sling surgery (IC)	1.00	0.00	5,865	0
Treatment urosepsis (AS)	1.00	1.00	2,015	2,002
Transurethral prostate resection for BPH	0.00	1.10	0	4,715
Outpatient				
Follow-up RP: PSA testing (urologist)	8.62	3.50	25	23
Treatment ED (urologist)	6.00	7.75	19	19
Treatment ED (GP)	4.75	6.20	35	36
Treatment IC (urologist)	6.24	9.14	20	19
Treatment IC (GP)	5.54	5.80	35	36
EBRT after AS (radiotherapist)	0.00	1.82	0	1,380
AS: PSA testing and biopsies (urologist)	0.00	7.36	0	44
Pharmaceuticals				
Antibiotics before prostate biopsy	0.00	1.12	0	16
α-blockers, 5α-reductase inhibitors (BPH)	0.00	5.93	0	47
PDE5 inhibitors (ED)	1.00	0.00	23	0
Assistive technologies				
Vacuum pump, rings (ED)	1.00	1.00	205	198
Incontinence aids (IC)	6.11	8.25	39	29
Physiotherapy				
Physiotherapy pelvic floor (IC)	22.52	17.00	14	14

Table 4-8: Resource use and unit costs of single health technologies (AOK)

4.4.2 Probability of utilization and per capita costs

4.4.2.1 Unadjusted analysis

The health technology with the highest utilization in the RP-group is the RP surgery itself (80% without complications, 20% with complications) and follow-up after surgery by an urologist (72%). In the AS-group highest utilization is found for PSA testing and biopsies (100%), which constitute the surveillance scheme (Table A-1 appendix).

Utilization of remaining health technologies is low (< 10%) in both treatment groups, except for outpatient treatment of IC and use of incontinence aids in the RP-group with about 26% and 21%, respectively. Utilization of pharmaceuticals to treat symptoms of BPH is around 50% in the AS-group.

Low utilization, which corresponds to a low number of observations, influences validity of estimates of per capita costs; estimates based on a very low number of observations may not be representative of the whole cohort. Artificial urethral sphincter surgery, for example, is utilized by one individual in the AS-group only; estimated mean costs of treated individuals are extremely high with €11,732, which is about 2 times higher than the DRG (L04C, ca. €5,650) usually claimed for this procedure.

The relation of cost group utilization corresponds to utilization of single health technologies, described above. The cost group ‘total’ comprises utilization of any of the listed cost groups. A utilization of 100% in both treatment groups indicates that every individual in the cohort is utilizing at least one of the health technologies in the follow-up period. Figure 4-2 shows the probability of utilization, categorized in cost groups as defined in the decision model, for the RP- and the AS-group.

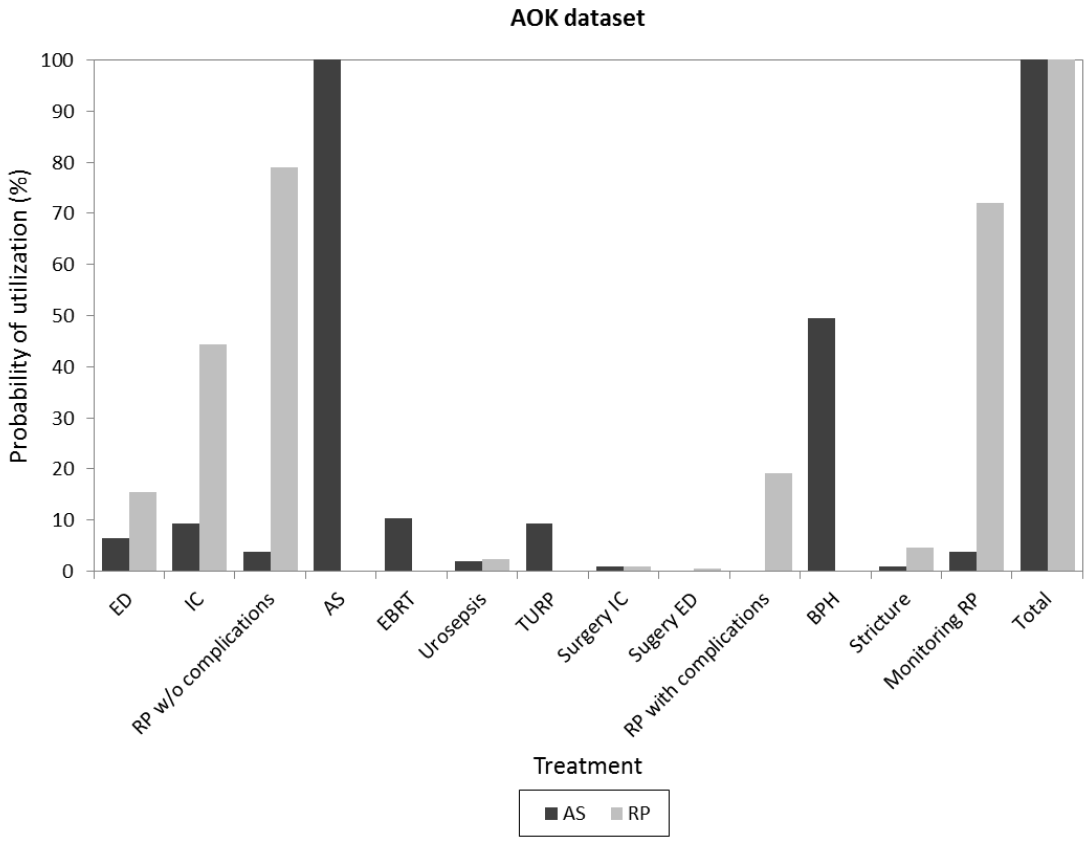


Figure 4-2: Probability of utilization (%), by treatment strategy and cost group - unadjusted (AOK)

Summarizing health technologies in cost groups mitigates the problem of small numbers of observations to a certain degree. Especially, the integration of health technologies for the treatment of ED and IC in corresponding cost groups, allows estimation of per capita costs based on larger numbers of observations (RP: 33 and 95 individuals, AS: 7 and 10 individuals) (Table A-2 appendix).

Highest unadjusted per capita costs are found for RP surgery without complications in the RP-group (€5,067, STD €2,741) (Table A-2 appendix). A median of €6,335 and interquartile range of €1,424 show a relatively even distribution of surgery costs in the RP-group. Costs of RP surgery in the AS-group, on the other hand, accumulate to €235 per person. STD (€1,207), median (€0) and interquartile range (€0) show an extremely skewed distribution of surgery costs in the AS-group, where a small number of individuals incurs high costs. This cost distribution can be found in almost all cost groups, except for per capita costs of surveillance in the AS-group (€88, STD €40) and follow-up after surgery in the RP-group (€133, STD €101).

Total per capita costs amount to €7,711 (STD €3,511, median €6,990, interquartile range €1,771) in the RP-group and €1,658 (STD €4,142, median €250, interquartile range €703) in the AS-group, respectively.

When comparing per capita costs between treatment groups in unadjusted analysis, it is shown that in 8 out of 13 cost groups the AS-group has lower costs, indicated by the minus sign (Table 4-9). The largest incremental costs are found for RP surgery without complications (€-4,832, STD 2,346). TURP and EBRT are by €685 and €258, respectively, more costly in the AS-group than in the RP-group. High STDs in both cost groups (€1,843 and €477) indicate extremely skewed distributions of incremental costs.

In total, unadjusted per capita costs in the AS-group are by €6,054 (STD 3,733) lower than in the RP-group.

Cost group	Incremental costs AS vs. RP (€)	
	m	STD
ED	-20	13
IC	-132	334
RP w/o complications	-4,832	2,346
AS	88	23
EBRT	258	477
Urosepsis	-10	306
TURP	685	1,843
Surgery IC	55	825
Surgery ED	-36	424
RP with complications	-1,943	3,624
BPH	138	135
Stricture	-176	1,231
Monitoring RP	-131	83
Total	-6,054	3,733

m= mean

Table 4-9: Disease-related, per capita incremental costs (€), by cost group – unadjusted (AOK)

4.4.2.2 Adjusted, bootstrapped analysis

When adjusting for co-morbidity in the regression model, highest per capita costs are found for RP surgery without complications (€5,051, CI [4,682; 5,400]) in the RP-group (Table 4-10). RP with complications incurs costs of €1,986 (CI [1,384; 2,655]) per person in the RP-group. In the AS-group highest per capita costs accumulate for the treatment of BPH with TURP (€771, CI [187; 1,792]); however, the wide 95% CI indicates substantial uncertainty regarding the cost estimate. Total costs amount to €7,861 (CI [7,314; 8,765]) in the RP-group and €1,601 (CI [928; 2,406]) in the AS-group, respectively.

	Costs p.p (€)			
Cost group	RP		AS	
	m	95% CI	m	95% CI
ED	39	26 to 55	20	5 to 39
IC	172	127 to 223	40	8 to 82
RP w/o complications	5,051	4,682 to 5,400	246	60 to 507
AS	0	0 to 0	88	81 to 97
EBRT	0	0 to 0	271	119 to 469
Urosepsis	50	14 to 96	42	13 to 103
TURP	0	0 to 0	771	187 to 1,792
Surgery IC	69	15 to 149	178	110 to 358
Surgery ED	83	68 to 143	0	0 to 0
RP with complications	1,986	1,384 to 2,655	0	0 to 0
BPH	0	0 to 0	140	97 to 188
Stricture	223	70 to 443	69	34 to 170
Monitoring RP	133	119 to 146	2	1 to 5
Total	7,861	7,314 to 8,765	1,601	928 to 2,406

m=mean

Table 4-10: Disease-related, per capita costs (€), by cost group – adjusted (AOK)

Adjusted incremental cost analysis shows that in the AS-group per capita costs of RP surgery without complications are by €4,805 (CI [-5,227; -4,359], p<0.0001) significantly lower than in the RP-group (Table 4-11). Costs of IC treatment (€-133, CI [-194; -71], p<0.0001) and monitoring after RP surgery (€-131, CI [-144; -117], p<0.0001] are also significantly lower in

the AS-group. In the remaining cost groups either no significant cost differences are found, or 95% CIs and p-values could not be estimated by the regression model, because in one of the treatment groups all individuals have costs of 0.

In total, adjusted per capita costs in the AS-group are estimated to be by €6,260 (CI [-7,417; -5,205], p<0.0001) significantly lower than in the RP-group.

Cost group	Incremental costs (€) AS vs. RP		
	m	95% CI	p-value
ED	-19	-42 to 3	0.088
IC	-133	-194 to -71	**< 0.0001**
RP w/o complications	-4,805	-5,227 to -4,359	**< 0.0001**
AS	88	-	-
EBRT	271	-	-
Urosepsis	-8	-64 to 63	0.606
TURP	771	-	-
Surgery IC	106	-28 to 313	0.962
Surgery ED	-34	-	-
RP with complications	-1,986	-	-
BPH	140	-	-
Stricture	-154	-389 to 24	0.060
Monitoring RP	-131	-144 to -117	**< 0.0001**
Total	-6,260	-7,417 to -5,205	**< 0.0001**

m=mean

Table 4-11: Disease-related, per capita incremental costs (€), by cost group – adjusted (AOK)

4.5 Changes to decision model

Results of changes to the decision model's assumptions and additional analyses conducted with the model described in step 5 of the validation approach are presented in the following.

4.5.1 Input parameters: resource use and unit costs

Treatment	Resource use p.p.	Unit costs (€)
Inpatient		
Prostatectomy w/ o complications	1.00	6,886
Prostatectomy with complications	1.00	9,559
Treatment stricture (RP)	1.00	2,010
Penis prosthesis (ED)	1.00	8,452
Artificial urethral sphincter (IC)	1.00	6,394
Sling surgery (IC)	1.00	3,677
Treatment urosepsis (AS)	1.00	3,075
Transurethral prostate resection for BPH	1.00	3,037
Outpatient		
Follow-up RP: PSA testing (urologist)	10.00	37
Treatment ED (urologist)	2.50	32
Treatment ED (GP)	5.00	54
Treatment IC (urologist)	2.50	32
Treatment IC (GP)	5.00	53
EBRT after AS (radiotherapist)	1.00	4,742
AS: PSA testing and biopsies (urologist)	9.00	55
Pharmaceuticals		
Antibiotics before prostate biopsy	2.50	8
Alpha blockers, 5α-reductase inhibitors (BPH)	16.25	108
PDE5 inhibitors (ED)	21.88	114
Assistive technologies		
Vacuum pump, rings (ED)	10.00	319
Incontinence aids (IC)	-	39
Physiotherapy		
Physiotherapy pelvic floor (IC)	30.00	15

Table 4-12: Resource use and unit costs of single health technologies (PCa-model)

Same input parameters of resource use and unit costs are included in the decision model for the AS- and the RP-group.

Resource use – incorporated in the model from published literature and expert interviews – is adapted to 10 cycles, corresponding to 2.5 years of follow-up. It shows that inpatient treatments are utilized once on average (Table 4-12). The largest quantity of utilization, in terms of number of treatments, is assumed for physiotherapy (30 treatments).

Highest unit costs are assumed in the model for RP surgery with and without complications (€9,559 and €6,886) and penis prosthesis surgery (€8,452). Inpatient unit costs are, generally, considerably higher than unit costs of technologies in the remaining health care service categories.

4.5.2 Simulation outcome: probability of utilization and per capita costs

4.5.2.1 Microsimulation

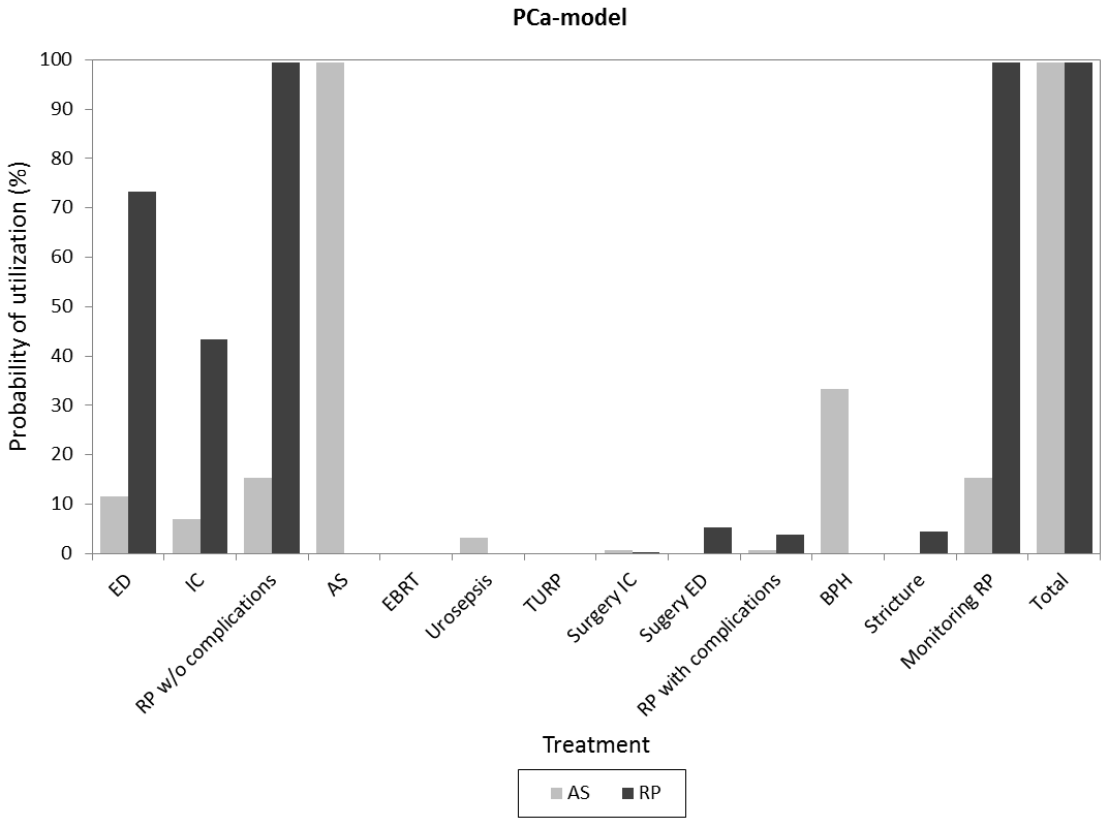


Figure 4-3: Probability of utilization (%), by treatment strategy and cost group - microsimulation (PCa-model)

Figure 4-3 shows the probability of utilization, categorized in cost groups of the PCa-model, for the RP- and the AS-group.

Highest utilization in the RP-group is found for the RP surgery without complications and monitoring after surgery by an urologist. Both treatments are utilized by 99% of individuals in the RP-group. It is assumed in the model that 100% of individuals in the RP-group receive these treatments; the utilization simulated during 10 cycles is less, because 2 individuals die before treatment (Table A-3 appendix). Health technologies for the treatment of ED and IC also have a high utilization in the RP-group (73% and 43%).

Individuals under AS predominantly utilize technologies that relate to the surveillance itself (99%). Again, AS is assumed to be utilized by all individuals in this group, but utilization is lower because of early mortality. About 15% of individuals under AS receive a RP surgery and are monitored after the surgery.

Treatment of ED and IC is considerably lower in the AS-group than in the RP-group with 12% vs. 73% and 7% vs. 43%, respectively. Pharmaceutical treatment of BPH is utilized by 33% of individuals in the AS-group, as compared to no utilization in the RP-group; by assumption BPH is cured in the RP-group by the initial surgery. In the remaining cost groups utilization is low (<5%) in the AS- and the RP-group. Due to the early death of 4 individuals (2 in each treatment group), total utilization of any health technology amounts to 99%.

In both treatment groups per capita costs, accumulated during follow-up, are highest for RP surgery (RP: mean €6,843, STD 543, median 6,886, interquartile range 0; AS: mean €1,051, STD 2,480, median 0, interquartile range 0) (Table A-3 appendix). In the RP-group €1,073 per person (STD 85, median 1,080, interquartile range 0) arise for monitoring after the surgery; €889 per person (STD 884, median 189, interquartile range 1,890) accrue due to the treatment of ED. For the surveillance of individuals under AS on average €416 (STD 111, median 463, interquartile range 0) are spent over the study duration of 2.5 years.

For cost groups with low utilization estimates of per capita costs are extremely uncertain, indicated by large absolute values of STD. This is, for example, the case for IC surgery; in both treatment groups mean costs are low (€31 and €63) while STDs are high (€562 and €794). Total per capita costs amount to €9,712 (STD 2,585, median 9,856, interquartile range 1,505) in the RP-group and €2,220 (STD 3,354, median 463, interquartile range 1,080) in the AS-group, respectively.

Cost group	Incremental costs (€) AS vs. RP	
	m	STD
ED	-818	652
IC	-214	379
RP w/o complications	-5,792	1,795
AS	416	79
EBRT	0	0
Urosepsis	96	378
TURP	0	0
Surgery IC	31	688
Surgery ED	-448	1,341
RP with complications	-298	1,390
BPH	317	336
Stricture	-88	291
Monitoring RP	-908	282
Total	-7,492	2,994

m=mean

Table 4-13: Disease-related, per capita incremental costs (€), by cost group – microsimulation (PCa-model)

The comparison of per capita costs shows that in the AS-group €5,792 (STD 1,795) less is spent on RP surgery without complications (Table 4-13). Treatment of ED (€-818, STD 652) as well as monitoring after RP surgery (€-908, STD 282) are also less costly in the AS-group. Costs of surveillance, on the other hand, are by €416 (STD 79) higher in the AS-group than in the RP-group.

Generally, individuals in the AS-group incur fewer costs than individuals in the RP-group in 7 out of 13 cost groups. Total per capita costs are by €7,492 (STD 2,994) lower in the AS-group than in the RP-group in the microsimulation of the decision model.

4.5.2.2 Monte Carlo simulation

By using Monte Carlo simulation, 95% CIs of per capita costs are estimated per cost group and treatment strategy; for incremental costs p-values are estimated.

Similar to results of microsimulation, highest per capita costs in both treatment strategies are incurred by RP surgery with (RP: €6,950, CI [5,392; 8,677]; AS: €1,028, CI [638; 1,534]) or without complications (RP: €6,826, CI [5,291; 8,551]; AS: €1,010, CI [627; 1,515]) (Table 4-14). The range of CIs shows, though, that considerable uncertainty is present in the cost estimates.

Not considering cost groups with null costs, lowest per capita costs incurs IC surgery in the RP-group (€19, CI [8; 36]); treatment of BPH by TURP (€4, CI [4; 4]) and surgery due to ED (€4, CI [2; 7]) incur lowest costs in the AS-group. Total per capita costs, as estimated by Monte Carlo simulation, amount to €9,627 (CI [8,009; 11,387]) in the RP-group and to €2,141 (CI [1,662; 2,738]) in the AS-group, respectively.

Cost group	Costs p.p. (€)			
	RP		AS	
	m	95% CI	m	95% CI
ED	837	608 to 1,095	69	42 to 106
IC	240	0 to 1,960	26	15 to 39
RP w/o complications	6,826	5,291 to 8,551	1,010	627 to 1,515
AS	0	0 to 0	417	404 to 428
EBRT	0	0 to 0	0	0 to 0
Urosepsis	0	0 to 0	95	37 to 183
TURP	0	0 to 0	4	4 to 4
Surgery IC	19	8 to 36	33	18 to 53
Surgery ED	415	286 to 589	4	2 to 7
RP with complications	6,950	5,392 to 8,677	1,028	638 to 1,534
BPH	0	0 to 0	297	227 to 376
Stricture	69	61 to 77	10	7 to 14
Monitoring RP	1,075	1,075 to 1,075	159	104 to 220
Total	9,627	8,009 to 11,387	2,141	1,662 to 2,738

m=mean

Table 4-14: Disease-related, per capita costs (€), by cost group – Monte Carlo simulation (PCa-model)

When comparing per capita costs estimated in Monte Carlo simulation, the highest cost difference between AS and RP is found for RP surgery with (€-5,922, CI [-7,480; -4,589], p<0.0001) and without complications (€-5,816, CI [-7,339; -4,493], p<0.0001) (Table 4-15); these differences are highly significant. The AS scheme incurs significantly higher costs in the AS-group than in the RP-group (€417, CI [404; 428], p<0.0001), which corresponds to the model assumption that AS is not performed after RP surgery. Pharmaceutical treatment of BPH is also significantly more costly in the AS-group than in the RP-group (€297, CI [227; 376], p<0.0001).

Overall, 7 out of 13 cost groups incur higher costs in the RP-group than in the AS-group and cost differences in all of these are highly significant (p<0.0001). Total costs of the AS-group are significantly lower by €7,486 (CI [-9,059; -6,093], p<0.0001) than total costs of the RP-group.

Cost group	Incremental costs (€) AS vs. RP		
	m	95% CI	p-value
ED	-769	-1,003 to -552	**< 0.0001**
IC	-236	-317 to -167	**< 0.0001**
RP w/o complications	-5,816	-7,339 to -4,493	**< 0.0001**
AS	417	404 to 428	**< 0.0001**
EBRT	0	0 to 0	-
Urosepsis	95	37 to 183	**< 0.0001**
TURP	4	4 to 4	**< 0.0001**
Surgery IC	13	-7 to 33	**< 0.0001**
Surgery ED	-411	-584 to -282	**< 0.0001**
RP with complications	-5,922	-7,480 to -4,589	**< 0.0001**
BPH	297	227 to 376	**< 0.0001**
Stricture	-59	-66 to -51	**< 0.0001**
Monitoring RP	-916	-971 to -856	**< 0.0001**
Total	-7,486	-9,059 to -6,093	**< 0.0001**

m=mean

Table 4-15: Disease-related, per capita incremental costs (€), by cost group – Monte Carlo simulation (PCa-model)

4.6 Comparison between model and claims data

Results of the comparison between outcomes of the PCa-model and AOK claims data are presented in the following; methods for outcome comparison are described in step 6 of the validation approach.

4.6.1 Input parameters: resource use and unit costs

The PCa-model, generally, assumes that the quantity of resource utilization is identical in the AS- and the RP-group (Table 4-12). Claims data analysis, however, shows that resource use differs between the treatment groups; individuals in the AS-group have higher resource use than individuals in the RP-group in outpatient treatment, pharmaceuticals, assistive technologies, and physiotherapy (Table 4-8). The exception is outpatient follow-up of RP surgery (RP: 8.62, AS: 3.50) and physiotherapy as a treatment of incontinence (RP: 22.52, AS: 17.00), where resource use is lower in the AS-group.

Inpatient procedures are utilized once, both in AOK dataset and model, with the exception of TURP in the AS-group (AOK) which is utilized 2 times by one of the treated individuals. The quantity of utilized outpatient procedures is higher in the PCa-model in case of follow-up of RP surgery (AOK: 8.62 and 3.50, model: 10.00) and surveillance of AS individuals (model: 9.00, AOK: 7.36). The model assumes lower resource use for all remaining outpatient procedures than estimated in claims data analysis for both treatment groups.

The number of pharmaceutical prescriptions varies widely between AOK estimates and model input parameters; claims data analysis estimates lower resource use throughout. Especially consumption of PDE5 inhibitors for the treatment of ED is, with 1 prescription, considerably lower than assumed in the model (22 prescriptions). In this case, the AOK estimate is probably too low, because PDE5 inhibitors are generally not covered by SHI; thus, prescriptions are not coded in AOK data. This is also the explanation for lower estimated utilization of assistive technologies for the treatment of ED in claims data (RP: 1.00, AS: 1.00) than in the model (10.00).

The PCa-model assumes equal unit costs of health technologies for the AS- and the RP-group. Unit costs estimated in claims data analysis are overall comparable between treatment groups, which supports the assumption of equal unit costs in the model.

Unit costs of inpatient procedures are comparable between the AOK data analysis and the PCa-model. Unit costs of inpatient treatment of stricture (AOK: €4,298; model: €2,010) and artificial urethral sphincter surgery (AOK: €11,732; model: €6,394), however, are about 2 times higher in claims data analysis than assumed in the decision model. Both claims data estimates are based on an extremely low number of observations, which limits validity considerably. Unit costs included in the model may be more valid in this case.

In outpatient treatment, unit costs of EBRT are considerably overestimated in the PCa-model with €4,742, in comparison to €1,380 estimated in claims data analysis. Here, claims data analysis is based on a number of observations sufficient to estimate valid unit costs of EBRT.

In case of pharmaceutical treatment, unit cost estimates vary widely between claims data analysis and decision model. Especially, unit costs of pharmaceutical treatment of BPH (AOK: €47; model: €108) and ED (AOK: €23; model: €114) included in the model are 2 and 5 times higher, respectively, than claims data estimates. As described previously, pharmaceutical treatment of ED is usually not covered by SHI and unit cost estimates may, thus, be limited in their validity.

4.6.2 Simulation outcome: probability of utilization and per capita costs

4.6.2.1 Unadjusted claims data analysis vs. microsimulation model

Figure 4-4 and Figure 4-5 show the comparison of utilization probability, as estimated in claims data analysis and microsimulation of the model, for the AS-group and the RP-group, respectively. Estimates of utilization probability of the initial treatment (AS or RP with/ without complications) and of total costs are equal in both data sources (AOK: 100%, model: 99%) (see also Table A-2 and Table A-3 appendix). However, claims data analysis estimates a considerably higher probability of RP with complications (19%) compared to the model (4%).

In microsimulation of the PCa-model treatment of ED is utilized by 12% in the AS-group compared to just 7% estimated in claims data analysis, respectively. In the RP-group the estimate of ED-treatment is also considerably lower in claims data analysis than in microsimulation (AOK: 15%, model: 73%). As described previously, claims data analysis is limited in this respect, because treatment of ED is covered by SHI in exceptional cases only.

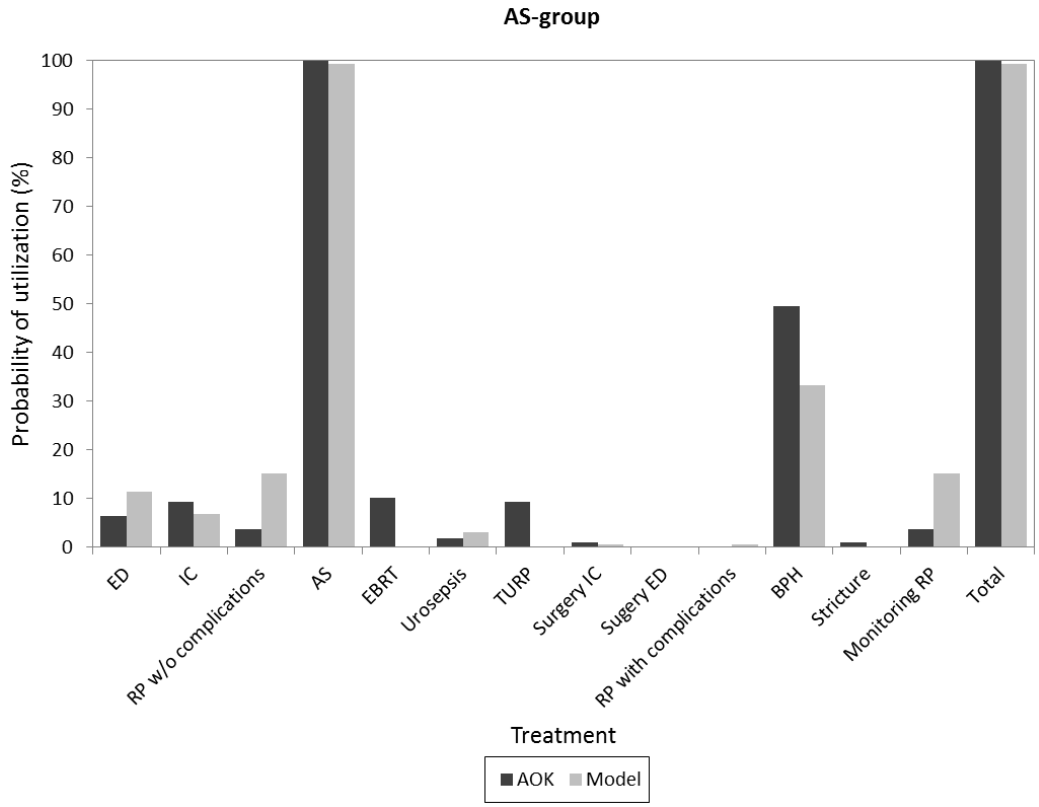


Figure 4-4: Probability of utilization (%) in AS-group, by data source

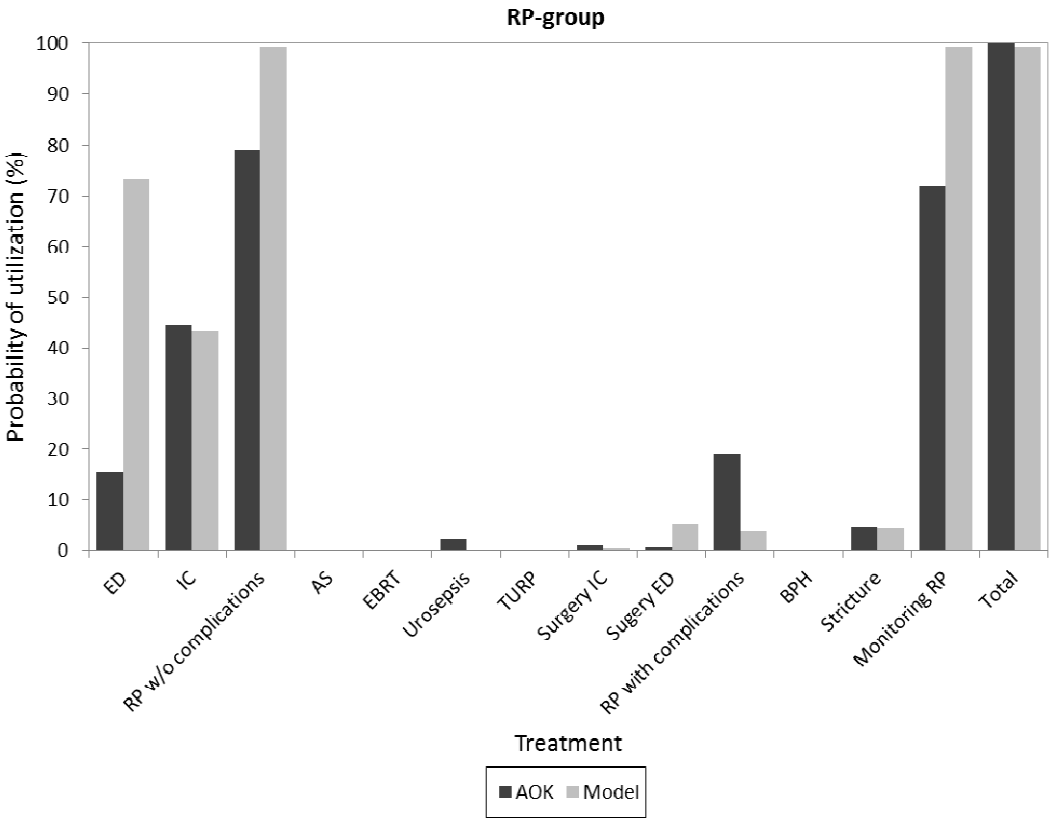


Figure 4-5: Probability of utilization (%) in RP-group, by data source

Treatment of IC is similar in AOK dataset and model, with 7% and 9% in the AS-group and 44% and 43% in the RP-group. The PCa-model assumes that in 15% of individuals in the AS-group a RP surgery is performed in follow-up, whereas this number is estimated to be considerably lower in claims data (4%).

As assumed in the PCa-model, EBRT is not performed at all during follow-up after AS, whereas claims data analysis shows that EBRT is utilized by 10% of individuals in the AS-group. Equal results are found for the treatment of BPH by TURP, where the model assumes no utilization, whereas in claims data analysis 9% of AS-individuals are treated.

Treatment of urosepsis, surgery for IC and ED, and treatment of stricture have an equally low probability of utilization in both datasets and treatment strategies.

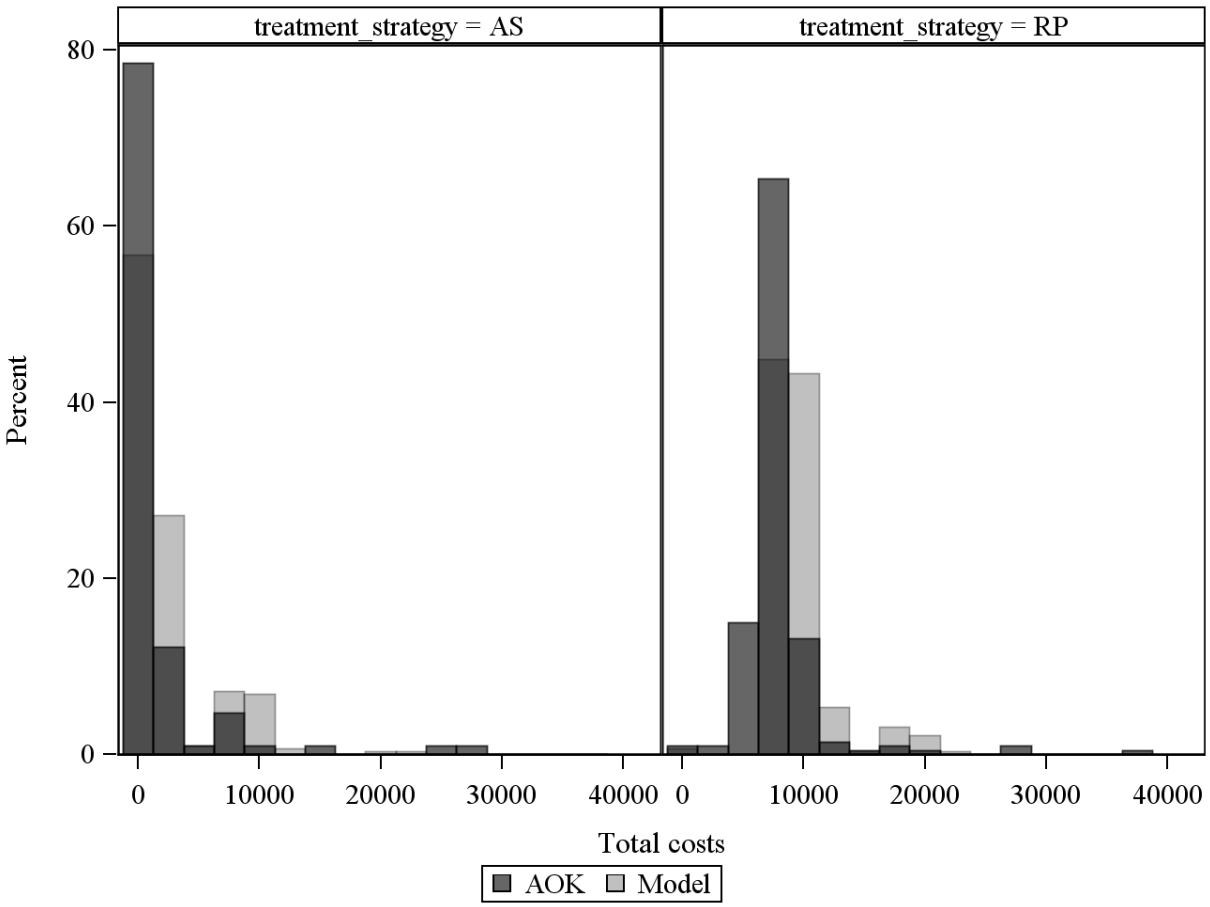


Figure 4-6: Total disease-related, per capita costs (€), by treatment strategy and data source

When comparing unadjusted incremental costs estimated in claims data analysis and incremental costs simulated in the model, the largest deviation is found for RP surgery with complications. Mean incremental costs in AOK dataset (€-1,943, STD 3,624) are about 6.5

times higher than simulated incremental costs in the model (€-298, STD 1,390). Both, however, suggest that costs in the AS-group are lower than in the RP-group.

Generally, in cost groups where the PCa-model simulates lower costs for the AS-group, this relationship is also found in claims data analysis. The exception is inpatient treatment of urosepsis, where claims data analysis estimates that the AS-group has slightly lower mean costs than the RP-group (€-10, STD 306), whereas the model simulates by 96€ (STD 378) higher costs per individual in the AS-group.

Total incremental costs estimated in unadjusted claims data analysis (€-6,054, STD 3,733) are €1,438 (19%) lower than total incremental costs simulated in the PCa-model (€-7,492, STD 2,994). Figure 4-6 pictures the distribution of total incremental costs estimated in claims data analysis and microsimulation of the model for the AS- and the RP-group separately (for comparison of single cost groups see figure A-1 in appendix).

4.6.2.2 Adjusted, bootstrapped claims data analysis vs. Monte Carlo simulation model

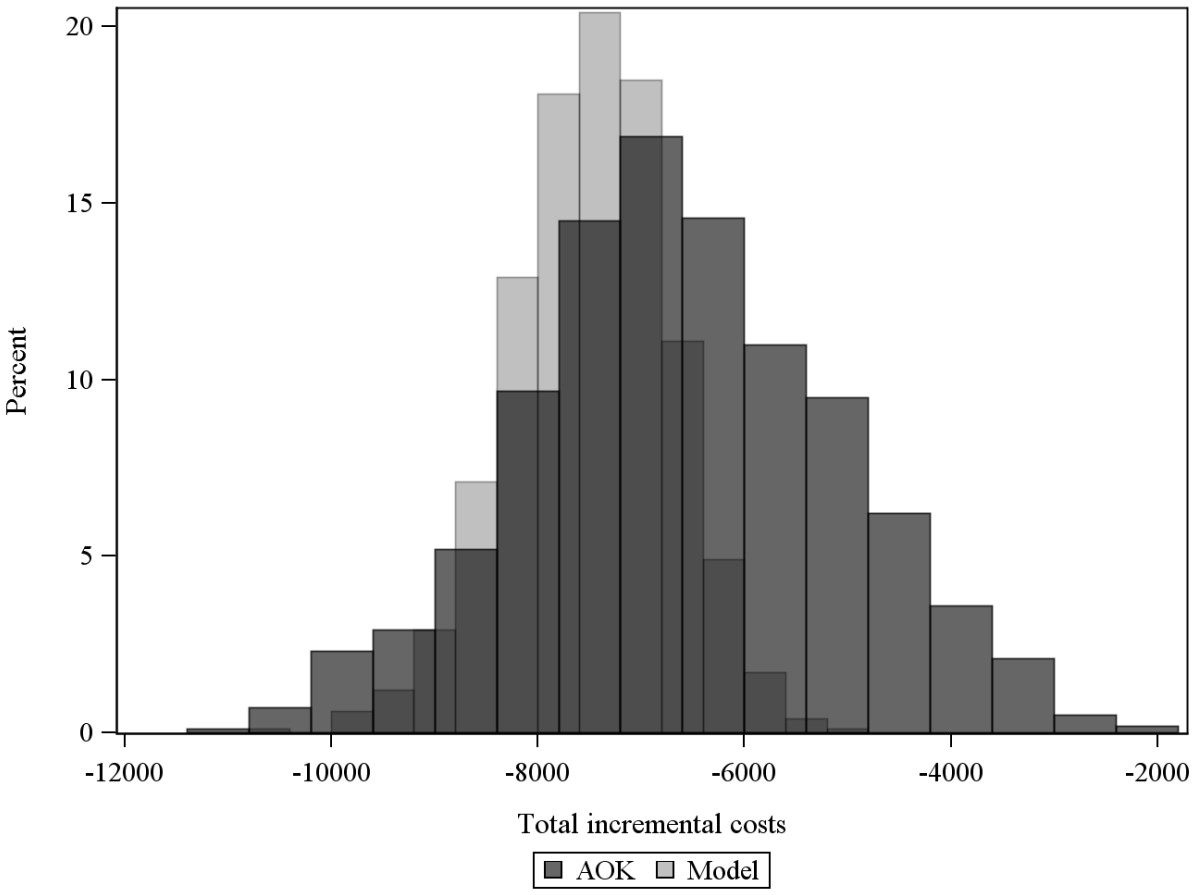


Figure 4-7: Histogram of total incremental costs estimated by excess cost analysis and Monte Carlo simulation (€), by data source

Excess costs of the AS-group estimated in adjusted, bootstrapped claims data analysis amount to €-6,611 (CI [-9,734; -3,547]), as described previously. Monte Carlo simulation of the PCa-model reports total incremental costs of €-7,486 (CI [-9,059; -6,093]). The difference between the claims data and model estimate amounts to €875 (CI [-2376; 4301]). A p-value of 0.605, estimated via bootstrap hypothesis testing, suggests a significant agreement between estimates on the 95% level, which is not due to random variation in both datasets. Figure 4-7 shows a graphical overlap of the distributions of mean excess costs generated by bootstrapping of claims data and Monte Carlo simulation, respectively.

For the comparison of incremental costs estimated in disease-related cost analysis with model outcomes, the difference in costs between AOK data and model is calculated for each cost group individually in addition to total costs (Table 4-16).

Cost group	Mean AOK	Mean model	Difference AOK-model	95% CI	p-value
ED	-19	-769	750	530 to 986	0.004
IC	-133	-236	103	10 to 199	0.038
RP w/o complications	-4,806	-5,816	1,011	-400 to 2,527	**0.186**
AS	88	417	-383	-342 to -313	<0.0001
EBRT	271	0	271	119 to 469	<0.0001
Urosepsis	-13	95	-108	-212 to -8	0.040
TURP	771	4	768	183 to 1,788	<0.0001
Surgery IC	60	13	46	-28 to 314	**0.936**
Surgery ED	-34	-411	377	216 to 556	<0.0001
RP with complications	-1,986	-5,922	3,936	2,406 to 5,602	<0.0001
BPH	140	297	-157	-244 to -69	<0.0001
Stricture	-181	-59	-122	-364 to 62	**0.214**
Monitoring RP	-131	-916	785	723 to 840	<0.0001
Total	-6,260	-7,486	1,226	-621 to 2,937	**0.180**

Table 4-16: Comparison of disease-related incremental costs in claims data and model, by cost group

Significant overlap – corresponding to a p-value larger than 0.05 – between claims data estimate and model estimate is found for RP with complications (€1,011, CI [-400; 2,527], p=0.186), surgery for IC (€46, CI [-28; 314], p=0.936), inpatient treatment of stricture (€-122, CI [-364; 62], p=0.214), and total costs (€1,226, CI [-621; 2,937], p=0.180). Figure 4-8 shows results of total incremental cost comparison graphically and strengthens the notion that confidence limits of both estimates significantly overlap (for comparison of single cost groups see Figure A-2 in appendix).

P-values of treatment of IC (p=0.038) and urosepsis (p=0.040) are just below the threshold for significant agreement, in all remaining cost groups highly significant difference in estimates is indicated (p<0.0001).

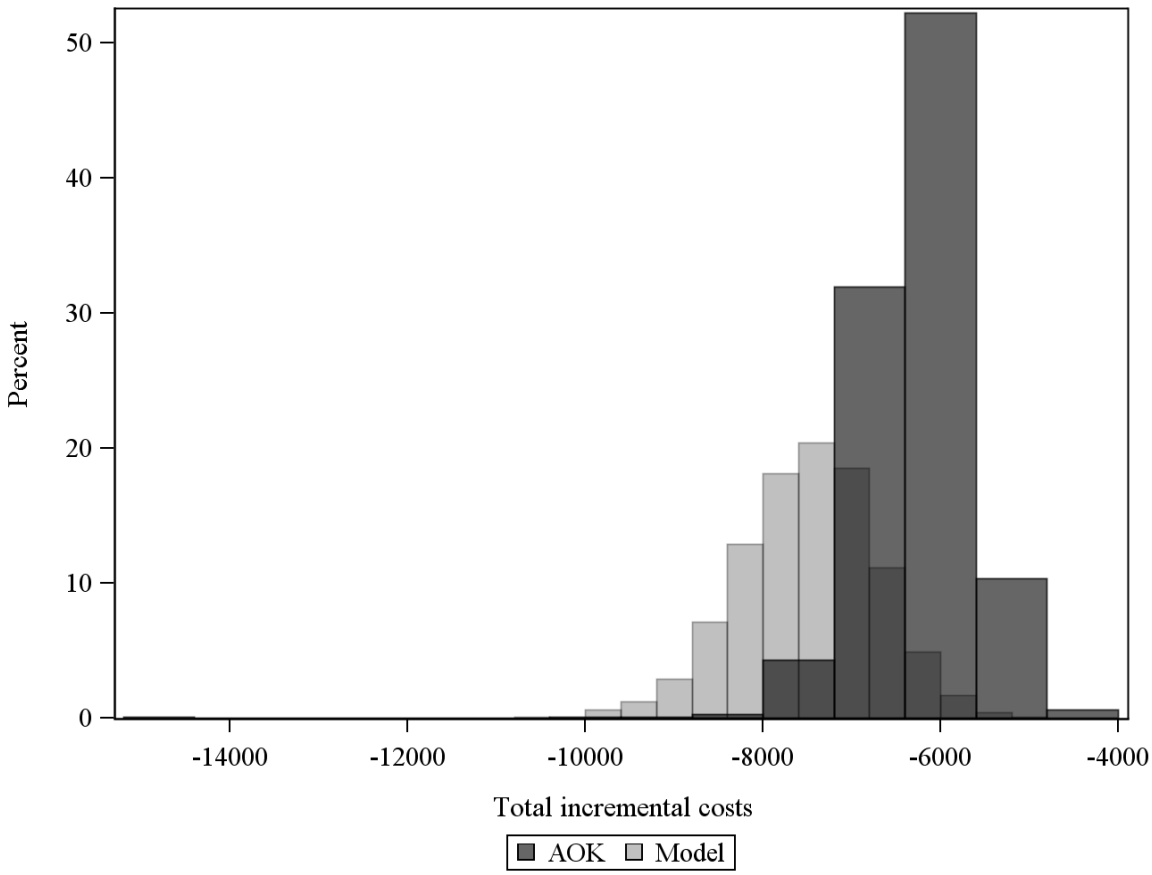


Figure 4-8: Histogram of disease-related total incremental costs estimated by adjusted claims data analysis and Monte Carlo simulation (€), by data source

4.7 Sensitivity analyses

Results of sensitivity analysis described in step 7 of the validation approach are presented in the following.

4.7.1 Incident PCa-cases in claims data analysis

In the first sensitivity analysis, including only incident PCa-cases, excess costs and total disease-related costs are analyzed for an age-matched cohort of 192 men (AS: 64, RP: 128). Mean age in the incident cohort is 72 years, 2 years older than the base case cohort.

Excess cost analysis reveals that mean co-morbidity adjusted costs of AS increase by €1,650 to €13,358 (CI [9,698; 17,308]) in total compared to base case. Per capita costs of the RP-group (€18,641, CI [16,216; 22,107]), on the other hand, do not change notably compared to the base case. AS is still significantly less costly than RP (€-5,283, CI [-9,585; -1,101], p=0.016), though. When comparing excess costs of claims data analysis with model simulation outcome (€-7486), a difference in incremental costs of €2,202 (CI [-2,132; 6,652]) is found. A p-value of 0.342 indicates that a significant agreement between estimates of excess cost analysis and model simulation is indicated.

The analysis of disease-related costs of the incident PCa-cohort reports only results of total costs, because estimation of valid costs of single health technologies is not possible due to very small numbers of treated patients. Bootstrapped and co-morbidity adjusted analysis estimates total costs of €7,793 (CI [7,185; 8,619]) in the RP-group and €1,976 (CI [885; 3,328]) in the AS-group, respectively. As in excess cost analysis, total costs of the RP-group are almost unchanged, whereas costs of the AS-group increase by €375 compared to base case analysis. Despite this increase, AS is still significantly less costly than RP (€-5,817, CI [-7,300; -4,213], p<0.0001) in the incident PCa-cohort. When comparing outcomes of disease-related cost analysis with total incremental costs simulated in the PCa-model (€-7,486) a mean difference of €1,670 (CI [-388; 3720], p=0.106) is found. In addition to this result, Figure 4-9 indicates that there is a significant overlap between estimates of incident claims data analysis and model simulation.

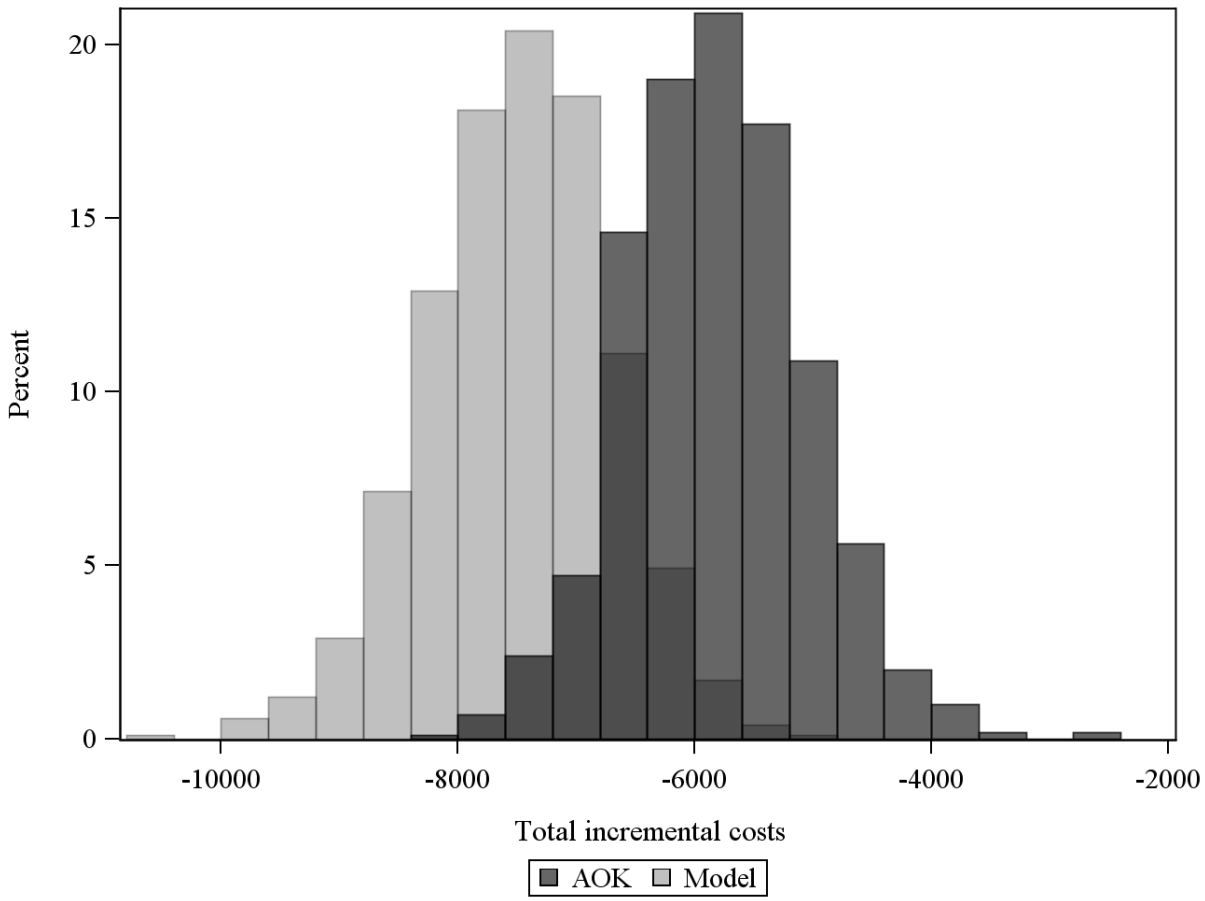


Figure 4-9: Histogram of disease-related total incremental costs (€) in incident PCa-cohort, by data source

4.7.2 Age distribution in the decision model

To test the influence of age at treatment initiation on total costs, the PCa-model is run twice; first with a mean age of 51 years at treatment initiation and corresponding age-dependent transition probabilities, secondly with a mean age of 79 years, respectively.

In the model with mean age of 51 years at treatment initiation, the probability of utilization of any health technology is 100%. No deaths occur in the follow-up period, owing to the considerably lower mortality rates in this age group. Monte Carlo simulation reports per capita costs of €9,682 (CI [8,058; 11,450]) in the RP-group and €2,178 (CI [1,691; 2,786]) in the AS-group respectively, which are not notably different from base analysis. Incremental cost analysis shows that per capita costs in the AS-group are by €-7,505 (CI [-9,079; -6,106], $p<0.0001$) significantly lower than in the RP-group. Comparison of model incremental costs with incremental costs from disease-related claims data analysis reports a difference of

€1,245 (CI [-612; 2,966], p=0.172), which indicates a significant agreement between estimates (Figure 4-10).

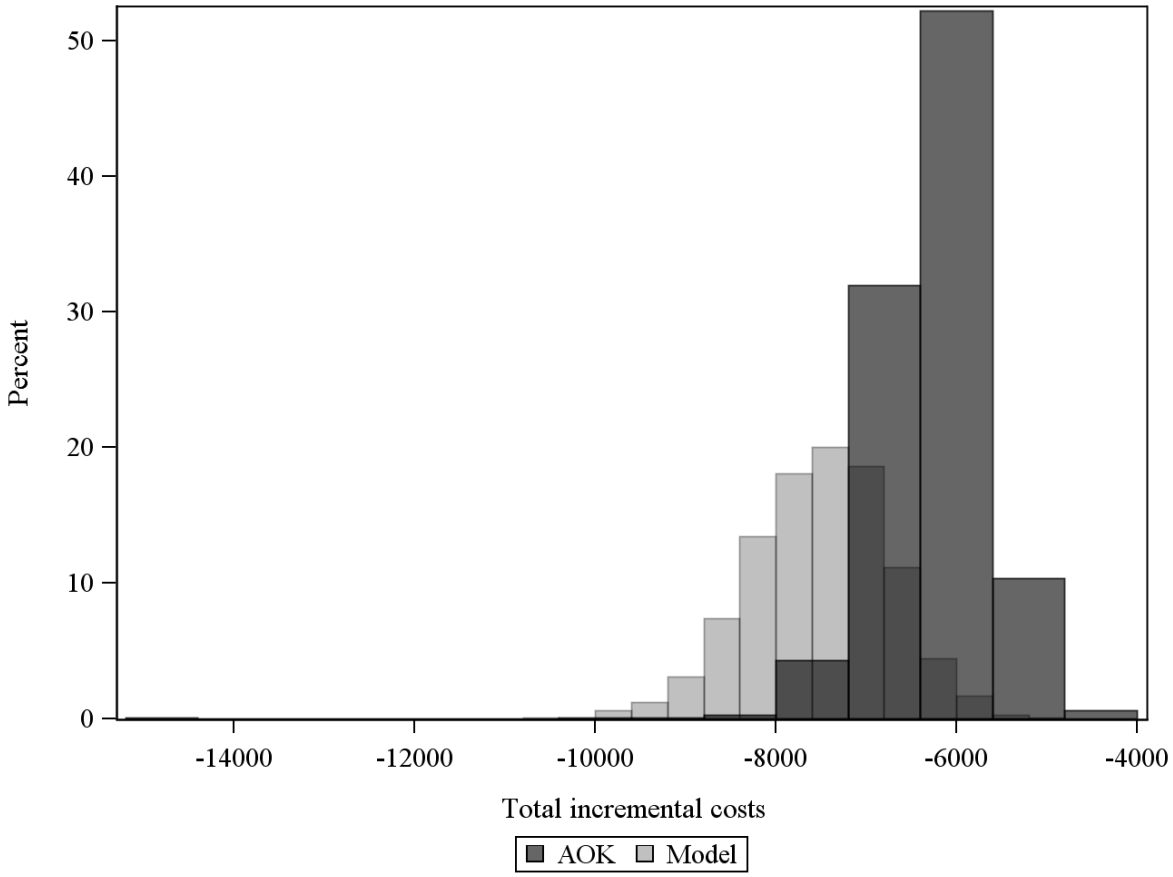


Figure 4-10: Histogram of disease-related total incremental costs (€) in PCa-model cohort with mean age of 51 years, by data source

When the PCa-model assumes an age of 79 years at treatment initiation, utilization of health technologies is reported for 97.8% of individuals in the cohort. Due to an increase in mortality in this age group, 7 individuals in each treatment group die early in the follow-up period and incur no costs. In Monte Carlo simulation per capita costs of €9,445 (CI [7,846; 11,181]) are estimated for individuals in the RP-group and €1,542 (CI [1,267; 1,852]) for individuals in the AS-group, respectively. These estimates of per capita costs are slightly lower than estimates of base case analysis, which is due to the higher number of observations with 0 costs. Furthermore, all individuals in the AS-group progressing to radical treatment are treated by EBRT and not RP, because of age-based assumptions in the model. This has the effect that total costs of the AS-strategy decrease compared to the base case, because EBRT is less costly than RP surgery. This is also reflected in incremental cost

analysis; total costs in the AS-group are significantly lower by €7,903 (CI [-9,628; -6,336], p<0.0001) than in the RP-group. In the base case analysis, in contrast, incremental costs amount to €-7,486 (CI [-9,059; -6,093]). In comparison to incremental costs of disease-related claims data analysis a difference of €1,644 (CI [-211; 3,530]) is found. A p-value of 0.08 indicates a significant agreement between estimates (Figure 4-11).

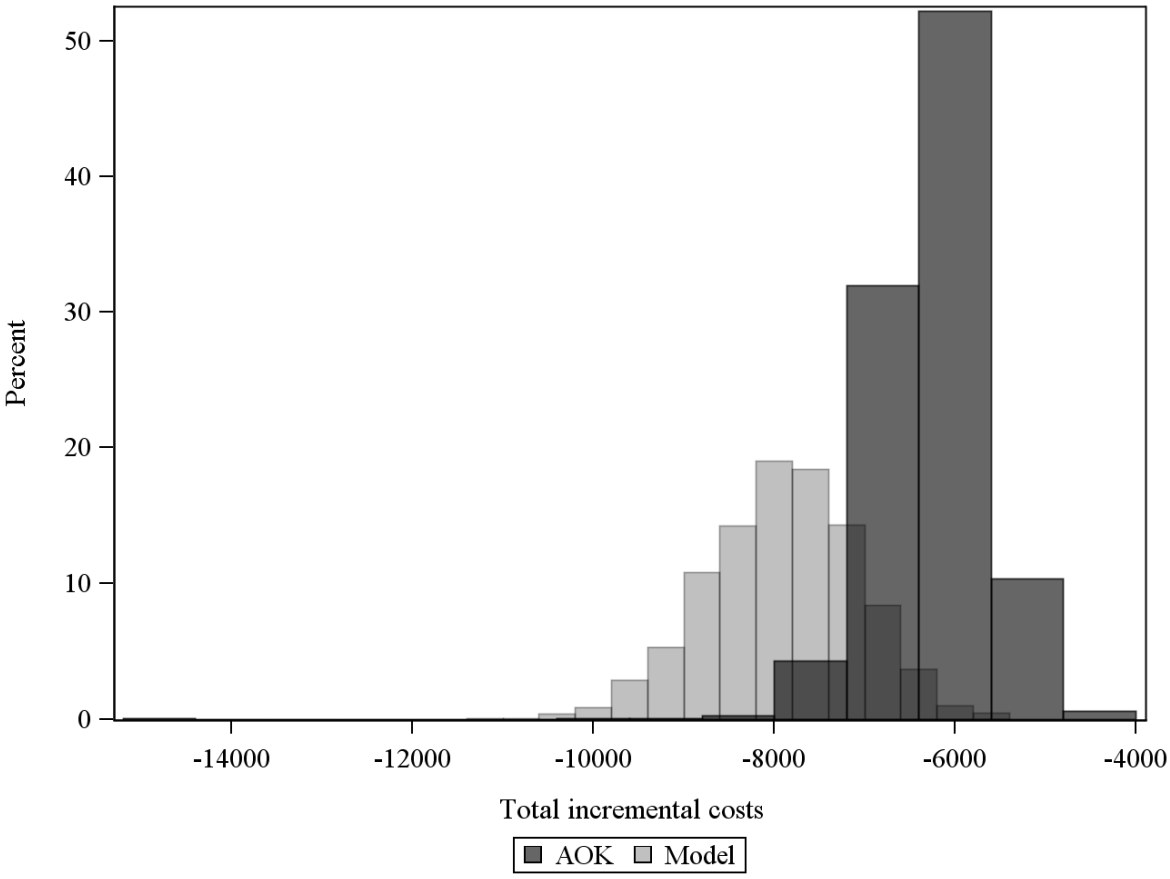


Figure 4-11: Histogram of disease-related total incremental costs (€) in PCa-model cohort with mean age of 79 years, by data source

5 Discussion

5.1 Interpretation of results

5.1.1 Patient cohort

The mean age in the claims data cohort is 70 years at treatment initiation which is replicated in the model cohort. Based on evidence from RCTs the PCa-model assumes that no co-morbidity is present in the cohort at treatment initiation; claims data analysis, however, shows that this is not reflected in treatment practice. Especially BPH is a common co-morbidity in the AS- and the RP-group with rates of 68% and 77%, respectively. ED is also present at baseline in both treatment groups with rates of 5% and 11%, respectively. The difference in baseline co-morbidity between treatment groups in claims data analysis is probably conditioned by the initial treatment decision; men with BPH and ED tend to radical therapy because their symptoms are cured as well. To account for this difference, costs are adjusted for co-morbidity in claims data analysis.

During cohort selection in the claims dataset, the majority of PCa-cases are lost in the process of assignment to treatment groups; 25,293 individuals are identified as valid PCa-cases, whereas only about 1,000 individuals are included each in the AS- and the RP-group, another 1,000 individuals are treated by EBRT (before limitation to localized PCa disease). This might be explained by comparatively higher prevalence rates than incidence rates in Germany; most PCa-cases in the AOK data set are not newly diagnosed and are, thus, not treated by AS, RP, or EBRT. The analysis of incident PCa-cases with no PCa-diagnosis before initial treatment supports this assumption; the number of individuals with a validated PCa-diagnosis decreases to 5,720. Individuals excluded from the claims data cohort are either in an advanced stage of PCa disease and are, thus, treated with other therapies than analyzed here. Alternatively, individuals are under WW because of advanced age. It is also possible that individuals with localized PCa were treated before beginning of the study period and PCa-diagnosis is still coded in the claims dataset, due to the coding practice in Germany.

Mortality rates in the claims data cohort and the model cohort differ significantly. Compared with the annual mortality rate of the male German population for the age group of 70 years (2.3%), mortality in the decision model (0.62%) is notably lower; mortality in the claims data cohort is notably higher in the AS-group (7.5%), whereas mortality rates in the RP-group

(2.8%) accord with rates of the German population. One reason for the notable difference in mortality rates between treatment groups in claims data analysis might be that follow-up of AS-patients begins on average 12 months later than follow-up of RP-patients; AS-patients are, thus, on average 12 months older than RP-patients. Mortality rates do, however, not increase so notably in the general German population in the course of 12 months.

As overall and PCa-specific mortality cannot be differentiated in claims data no inference on effectiveness of treatment strategies based on mortality rates is possible. The extreme difference in mortality rates between the treatment strategies in the claims data analysis and between claims data and decision model influences, however, analysis of resource use and costs. The comparatively higher mortality rate in the AS-group may, on the one hand, lead to an underestimation of health technology utilization, because due to the premature death fewer resources are utilized by these individuals than by individuals surviving in the study duration. On the other hand, evidence suggests that health care utilization and costs are significantly higher in end-of-life care, 6 months before death, than in the phases of initial treatment and follow-up, which might lead to an overestimation of costs of the AS-strategy compared to the RP-strategy (137).

5.1.2 Claims data and model analysis

Claims data analysis shows that individuals with RP surgery experience significantly more complications in follow-up (49%) compared to individuals under AS (12%). These complication rates should, however, be interpreted in the light of the limitations of analysis of effectiveness in the AOK dataset described in chapter 5.3.

The PCa-model does not report complication rates; however, it is shown that the RP-strategy generates less QALYs than the AS-strategy, which predominantly originates in the quality of life loss due to complications of ED and IC after RP-surgery.

Excess cost analysis and disease-related cost analysis of AOK claims data both show that overall AS is the less costly strategy compared to RP, with total incremental costs of €-6,611 and €-6,260, respectively.

Considering excess costs only, AS is significantly less costly than RP in inpatient care which originates mainly in the high costs of the hospital stay of the initial RP surgery (€-5,845). Disease-related cost analysis confirms this; the AS-group incurs €4,805 less for RP surgery

without complications and €1,986 less for RP surgery with complications. The difference in costs of RP surgery predominantly accounts for the total difference in costs between treatment groups, in excess cost analysis as well as in disease-related cost analysis. PCa-model simulation reports this as well; the difference in costs of RP surgery (€-5,816) constitutes the predominant part of total incremental costs of €-7,486.

Excess cost analysis reports that, compared to RP, outpatient care is less costly under AS (€-961) in the period of 2.5 years follow-up, even though the main costs of AS arise in outpatient care. In disease-related analysis, outpatient treatment is further categorized in treatment with single health technologies and results show a more differentiated picture of outpatient care; in the AS-group costs of the surveillance scheme itself (€88) and EBRT (€271) incur higher costs than in the RP-group. Outpatient treatment of ED (€-19) and IC (€-133) as well as monitoring after RP surgery by an urologist (€-131) is less costly in the AS-group. In the decision model cost differences are larger, but point to the same conclusion that outpatient surveillance is more costly in the AS-group (€417), while outpatient treatment of ED (€-769), IC (€-236), and monitoring after RP surgery (€-916) are significantly less costly than in the RP-strategy. The AS- and the RP-strategy, generally, show different cost patterns; the RP-strategy incurs high initial outpatient costs (e.g. for monitoring after the surgery) while costs of AS are more equally distributed over time (102, 104, 105, 138).

Pharmaceutical treatment of BPH is more costly in the AS-group, as estimated by disease-related claims data analysis (€140) and Markov model analysis (€297). Excess cost analysis does not explicitly analyze BPH treatment; however, overall costs of pharmaceutical therapy are by €587 higher in the AS-group than in the RP-group.

Furthermore, excess cost analysis of claims data shows that costs of assistive technologies - mainly incontinence aids - are significantly higher in the RP-group by €141, attributable to the higher IC rate in the RP-group. When complication is included as a variable in the regression analysis of claims data, a significant increase in costs for assistive technologies is found for individuals with complications. In disease-related claims data analysis and Markov model simulation utilization of assistive technologies is included in the cost groups ED and IC which both incur higher costs in the RP-group than in the AS-group, as described previously.

5.1.3 Comparison between model and claims data

No significant differences between the claims data estimate and model estimate of incremental costs are found for RP surgery without complications, surgery in case of IC, and surgical treatment of a stricture. The incremental cost estimate of RP surgery is based on a large number of observations and is afflicted by relatively little variability in the parameter estimates of the two treatment groups; the estimate of claims data analysis is very likely representative for costs incurred by AOK. The agreement between AOK and model estimate can be interpreted as sufficient to assume that the PCa-model simulates costs of RP surgery validly. Cost estimates of IC surgery and treatment of stricture, on the other hand, are based on a small number of observations in claims data analysis; the same limitation is found in PCa-model analysis. Both estimates of incremental costs are afflicted by considerable uncertainty, which is represented in wide 95% CIs. Even though an agreement between both estimates is found, it remains unclear whether the model predicts costs of IC and stricture surgery in treatment practice validly. However, in the interest of the decision maker validating the PCa-model total cost difference is most important. As described previously, the cost difference in RP surgery is the best indicator of the total cost difference and this parameter is predicted validly by the decision model.

No agreement between claims data and model estimate could be found in the remaining cost groups; this finding stems from different sources. In case of costs of the AS scheme for example, the model assumes higher resource use and unit costs than observed in claims data analysis, which in turn leads to about 4.5 times higher incremental costs of the AS-strategy in the decision model than in claims data. In case of EBRT treatment, on the other hand, no patients in both treatment groups receive this treatment in the decision model, whereas individuals in the AS-group in claims data analysis are treated by EBRT; consequently, no agreement of incremental cost estimates of claims data and model is found for EBRT treatment.

Overall, an agreement between total incremental cost estimates of excess cost analysis as well as disease-related cost analysis of claims data and PCa-model exists. P-values of 0.61 and 0.18, respectively, show a significant agreement of total incremental costs estimated by claims data analysis and PCa-model simulation. Results are robust to both changes to the

claims data cohort (incident PCa-cases; p=0.342) and changes to the model cohort (age 51 years: p=0.172, age 79 years: p=0.08).

The question arises whether inferring lifetime costs of AS compared to RP based on the validation of short-term resource use and costs of the PCa-model is possible: Incremental total costs of AS compared to RP in the short-term decision model amount to €-7,486, which show significant agreement with the observed incremental costs of €-6,260 in claims analysis; incremental costs are almost exclusively due to the high costs of the initial RP-surgery, both in model and claims data analysis. In the published PCa-model by Koerber et al., which evaluates a patient cohort of men aged 65 years in a lifetime perspective, incremental costs of AS compared to RP amount to €-6,883, which are also due to the costs of the initial RP-surgery. As the simulation of short-term costs is valid, it is very likely that long-term costs simulated by the decision model are valid as well. Inferring the validity of effectiveness outcome and, following this, cost-effectiveness outcome of the PCa-model is not possible based on this validation, though, because quality of life outcome cannot be assessed validly in AOK claims data.

5.2 Comparison with literature

5.2.1 Literature on PCa treatment

Rates of short- and long-term ED after RP (15% and 12%) estimated in claims data analysis are considerably lower than rates reported in clinical trials. A recent meta-analysis by Ficarra et al. (2012) reports ED rates of about 35%-75%, 12 months after surgery (75). In the PCa-model short-term ED occurs in 39% of individuals in the RP-group and long-term ED in 35%, respectively, based on data from a meta-analysis by Ollendorf et al. (2009) (114).

According to claims data analysis, 6% of individuals in the AS-group develop ED in the follow-up period. In comparison, a study by Braun et al. (2014) reports an ED rate of roughly 35% in a cohort of men under AS at baseline; even without possible decrease of erectile function due to repeat prostate biopsies, ED rates are considerably higher than in the AOK cohort. ED rates are generally underestimated in the AOK dataset, because treatment of ED is usually not covered by SHI and, thus, diagnoses are not validly coded.

Claims data analysis reports rates of IC after RP of 44% (short-term) and 33% (long-term). These rates are slightly higher than the IC rate reported in a meta-analysis by Ficarra et al. (2012) of about 20%, 12 months after surgery (74). IC rates included in the model (short-term 9%, long-term 2.5%) are, however, considerably lower than rates of claims data analysis (114). In case of IC, treatment is covered by SHI and diagnoses are validly coded in AOK dataset. It is, however, not possible to distinguish age-related and surgery induced decline of urinary function in SHI data. This may be a reason for the higher IC rates reported in AOK data analysis compared to published evidence.

Validity of ED and IC rates also influences validity of cost estimates of ED and IC treatment. In case of ED treatment in the AS-group, for example, estimates of unit costs may not be valid due to a small number of observations.

Anderson et al. (2011) compare costs of RP and WW in a study based on data from the SPCG-4 trial (96); the authors calculate unit costs of penis prosthesis surgery of €7,010, similar to the unit cost estimate of claims data analysis (€7,586) presented in this study. Anderson et al. (2011) estimate per capita costs of penis prosthesis surgery in the RP-group of €131 over a study duration of 12 years. AOK claims analysis, however, estimates lower costs of about €36 per individual in the RP-group, which is on the one hand due to the

considerably shorter follow-up period; on the other hand, per capita costs are underestimated in SHI claims because ED-cases are not comprehensively included in the analysis.

Unit costs of IC surgery (€9,072) based on the SPCG-4 trial are similar to unit cost estimated in claims data analysis (€8,799). Even though the number of individuals with IC surgery is very low in the AOK data set, the estimate of unit costs is comparable to published costs. However, per capita costs in AOK data analysis for IC surgery (RP: €55, AS: €110) are considerably lower than in the SPCG-4 trial (RP: €1,017, WW: €86), due to the shorter follow-up period.

Anderson et al. (2011) estimate unit costs of €6,031 for RP surgery, similar to the unit costs presented in this study (€6,356). Per capita costs of RP surgery accumulate to €7,732 in the RP-group and €746 in the AS-group, as estimated by the SPCG-4 trial; in AOK claims analysis unit costs amount to €5,051 and €246, respectively. AOK estimates are lower than SPCG-4 trial estimates; however, in both studies per capita costs of RP surgery are significantly higher in the RP-group than in the AS-group with a p-value lower than 0.0001. In the SPCG-4 trial, costs attributable to the WW scheme (physician contacts and laboratory tests) do not differ between treatment strategies, because WW does not have such a strict surveillance scheme as AS.

Over a study duration of 12 years, total incremental costs of €6,123 between the WW- and the RP-strategy are estimated based on SPCG-4 data, corresponding to results of claims data and model analysis presented here. Despite differing length of study duration in claims data analysis and PCa-model simulation compared to the study by Anderson et al. (2011), the cost difference is driven almost exclusively by the costs of RP surgery in all 3 studies.

As described in chapter 2.2, several US-based modeling and observational studies with comparable patient cohorts and time horizons show that the AS-strategy or the CM-strategy (including AS and WW) are least costly over the whole study duration (78, 97, 99, 102, 103). Only the study by Perlroth et al. (2012) reports that from year 2 of the study on costs of CM become equal to RP (100). Perlroth et al. do not state unit costs of the surveillance scheme; however, other US-based studies show that the unit costs of prostate biopsy ($605-$1,102) alone are considerably higher in the US health-care context than unit costs of the whole surveillance scheme (PSA testing and biopsy) in Germany (€44) (78, 102, 103).

Results presented here suggest that in an European health care context - contrary to US health care - cost differences between AS and RP arise in the first years after treatment and do not converge in a lifetime perspective because of the relatively low costs of the AS scheme (101).

5.2.2 Literature on model validation

In a qualitative literature search several studies are found which validate decision models with external data sources (20-25). Both studies by Palmer et al. (2004) and McEwan et al. (2014) use published RCT data to externally validate complication and mortality rates of the CORE diabetes model (22, 24). Similarly, Eddy and Schlessinger (2003) conduct an external validation of the Archimedes diabetes model with RCT data, which are not incorporated in the decision model (23). The study by Ishida et al. (2008) employs data from medical records to externally validate survival rates predicted by a Markov model which compares treatment for hepatitis C virus-related hepatocellular carcinoma (20). Van Staa et al. (2013) validate several RCT-based decision models comparing costs and effects of nonsteroidal anti-inflammatory drugs and selective cox-2 inhibitors with outcomes from a claims data based simulation model (21). Stollenwerk et al. (2009) validate a Markov model for coronary artery disease risk screening in Germany with empirical data from an observational study (25).
In addition, 2 studies are found which do not validate a decision model, but give indications on the methodological approach to validation with external data sources. For one, Bratzler et al. (2011) validate a claims data based regression model which predicts 30-day hospital mortality by mortality rates reported in medical records (32). Secondly, a study by Janson et al. (2005) conducts an economic evaluation of open versus laparoscopic surgery for colon cancer based on self-reported patient data from a clinical trial; self-reported data of a subset of study participants are validated with data from medical records and social security claims (31).

This study compares outcomes independently observed in the trial and predicted in the decision model; evidence from the external trial is not incorporated in the decision model as in the studies by Ishida et al. (2008) and Stollenwerk et al. (2009) (20, 25).
In contrast to studies changing characteristics of the decision model (21-24), the aim of the study presented here is to validate the decision model as it was originally designed and avoid

changes to the model as far as possible; instead, the cohort of the external data source – the AOK dataset – is adapted to the model's cohort. Likewise, the study design assumed in the model, including length of follow-up and treatment strategies, is replicated in the external data source in this study. However, adaptation of the external data source is limited to available parameters; demographic information, for example, is not sufficient in the AOK dataset to replicate all cohort characteristics of the PCa-model, as it is possible in the study by Ishida et al. (2008) which uses detailed information from medical records (20).
This replication of model characteristics in the external dataset is possible, because AOK data are originally analyzed for the purpose of model validation, contrary to studies where external evidence stems from published studies (22-25).

In this study, as proposed in the literature, the output of the external data source and the decision model are plotted graphically in form of histograms such that the horizontal axis denotes costs and the vertical axis denotes the observed and simulated values, respectively (29, 30).
As statistical tests are recommended in addition to graphical analysis to obtain quantitative information about the validity of the decision model, in this study a hypothesis test based on bootstrapping is employed. It is tested whether the point estimate of the external data source is included in the model's CI and the other way round (22, 25, 30). By using bootstrap hypothesis testing the problems of other test statistics, such as the Student t-test, described in chapter 1.2.3, are avoided.
Furthermore, linear regression analysis is proposed in the literature to compare observed and simulated outcomes (22, 24, 32). A prerequisite for regression analysis is that pairs of patients (paired observations) are present in model simulation and observed data. Observations are not paired in this study, though, and consequently linear regression analysis would estimate correlation for random pairs of observations. No conclusion on the validity of the decision model can be drawn from linear regression analysis in this study.
Studies comparing outcomes of different decision models apply goodness of fit tests to assess validity of the validated model (22, 33). In this study, outcomes of the decision model are compared with independently analyzed outcomes in claims data, so tests for goodness of model fit are not relevant to assess the validity of the PCa-model.

5.3 Strengths and limitations

In the following, the proposed approach for the validation of decision models based on claims data is discussed by highlighting strengths and limitations of claims data in the individual validation steps.

5.3.1 Validation level

One of the strengths of this study is that in step 1 of the validation approach, both input parameters and simulation outcome of the decision model are validated. Differences in input parameters between AOK data and the PCa-model help to explain differences in simulation outcome.
Another strength is that outcomes of claims data and model are estimated independently by adapting the study design of the claims data analysis to model assumptions and not the other way round, as it is often seen in published validation studies. That way the decision model is validated in the form it was originally constructed.

A limitation, on the other hand is that quality of life indicators are not representable in claims data analysis; consequently, the effectiveness outcome of the decision model cannot be validated in this study. Only the cost outcome of the decision model is validated, which is of greater importance anyway, because input parameters of unit costs and assumptions on resource use are afflicted by a higher degree of uncertainty than effectiveness parameters incorporated from high quality clinical trials.
This study is also limited in respect to its explanatory power of the model's predictive validity. In the external validation presented here, the model's structure is not validated; errors in the model assumptions are, thus, replicated in claims data analysis. During claims data analysis it became apparent that observed treatment pathways differ from the model's definition occasionally; for example, utilization of laboratory services in line with the surveillance scheme are found in the AOK data set, but these are not considered in the model and are, thus, not considered as disease-related costs in claims data analysis. Validation of the model structure or predictive power is not the focus of this study. The study does, however, indicate that validation of the models structure, especially regarding clinical care pathways, is possible based on claims data.

5.3.2 Claims data set, study design, and patient cohort

The major strength of the analysis of the AOK data set in step 2 of the validation approach is that actual treatment practice and costs of early-stage PCa incurred by the SHI scheme insured population in Germany is pictured in SHI claims data. A real-world cohort of patients who are treated with RP or AS is followed for complications and costs. In AOK data, analysis is based on exact and detailed utilization and cost information on different health care sectors. In contrast to the decision model, assumptions on resource use, reimbursement practice, and prices are not required. Overall, utilization and cost information from SHI claims data is reliable because actual spending on a broad range of services and technologies incurred by SHI is reported (52). For the validation of the model it is of importance to picture real-world costs in the external data source, because these are the relevant outcome measures in the decision model. Claims data of AOK are, thus, well suited to report these outcomes for validation of the PCa-model.

In addition, the AOK dataset provides a large number of cases with PCa to select treatment groups from. This permits the detection of rare events, such as treatment of ED and IC with surgery. In a smaller cohort these events are likely to be overlooked.

The claims data analysis replicates the study design simulated in model analysis, which is a major strength of the AOK data. Patients are initially treated with either AS or RP and are followed-up to assess complications and corresponding costs of the initial treatment. A long-enough follow-up is chosen to replicate the first part of the decision model which represents treatment of early-stage PCa, not considering treatment and costs of advanced disease. Additionally, a unique identification number in the AOK dataset allows merging of claims over the study duration and health care service categories on an individual insurant level without loss of information due to aggregation of data.

A limitation of SHI data in general is that no detailed clinical information is included. This is especially limiting in this study because no information on tumor stage or Gleason score – which might allow clinical classification of PCa – is available in the AOK dataset. To overcome this limitation, early-stage PCa is defined as absence of diagnoses of metastases. However, by this definition only a distinction between metastatic and non-metastatic PCa is possible. A sub-division of non-metastatic PCa in localized and locally advanced PCa and further

distinction of risk groups, as described in chapter 2.1.2, is not possible. To take this limitation into account, individuals receiving treatment associated with recurrence or advanced tumor progression are also excluded from the cohort.

In addition to this limitation, a study by Stausberg et al. (2008) shows that reliability of diagnoses coding with ICD-10 GM is only fair to moderate with agreement rates between coders of around 50% (139). Regarding the coding quality of diagnostic codes, inpatient data are assumed to be more comprehensive than outpatient data (140, 141). To take the lower validity of outpatient diagnostic codes into account in this study, outpatient diagnoses are only considered when 2 consecutive codes are found in the dataset.

Generally, only medical care and diagnoses that are relevant for SHI reimbursement are included in claims data; coding is biased by the underlying billing purpose of claims data. This effect should be equally distributed across treatment groups and not bias comparative outcomes. However, when complication rates or costs of single treatments are supposed to be representative for the general population, this limitation biases results considerably. Diagnosis and treatment of ED is an example in the AOK data analysis, where comparison with published complication rates shows that ED rates are considerably underestimated in claims data analysis. The reason for this is that treatment of ED, surgical or pharmaceutical, is generally not covered by SHI. As treatment of ED is not relevant for SHI reimbursement, physicians either have no incentive to code ED comprehensively despite established diagnosis or patients do not consult a physician in the first place. In turn, the underestimation of complication rates influences the validity of resource use and cost estimates. Unit costs of ED treatment estimated in claims data analysis are, thus, not valid for the validation of unit costs incorporated in the decision model. Validation of incremental costs, on the other hand, is not afflicted by this limitation of the claims data analysis.

Furthermore, while date of death is coded for all insured persons in claims data, no information on cause of death is available. This infringes validity of claims data as death cannot be causally linked to a certain diagnosis or health technology utilization. Especially in analysis of cancer diseases, cancer-specific mortality is an outcome of interest which cannot be validly reported in claims data. In this study on PCa, mortality is not of primary interest because due to the nature of the tumor progression mortality rates are low in a study period of 2.5 years. Still, it would be interesting to have information on cause of death to assess

whether the high difference in mortality rates between AS and RP found in claims data analysis is due to the initial treatment decision.

Concerning the study design several limitations are present in this study. Claims data are available for one insurance fund with a regional focus on Baden-Württemberg only. AOK insurance funds tend to insure a proportionally larger population of individuals with lower educational status and low-skilled professions than found in the general German population, as described in chapter 3.3.1. This may influence outcomes of claims data analysis, because patients with lower socioeconomic status tend to have a higher PCa tumor grade and more advanced stage of disease at the time of diagnosis than patients with higher socioeconomic status (142). Evidence suggests that even when considering only patients with early-stage PCa, men with lower socioeconomic status are more likely to be treated with WW and less likely to receive treatment with curative intent (142). However, socioeconomic information is not included in such detail in the AOK claims data set to adjust for differences in socioeconomic status between treatment groups. Individuals in the AS-group are, thus, more likely to have a lower socioeconomic status than individuals in the RP-group; out-of-pocket payment for ED-treatment might, for example, be affected by this. However, as costs relevant for the SHI insured community are considered here and not out-of-pocket payments, socioeconomic status is likely to play a minor role in this study.

And while by including only one insurance fund conclusions on complication rates and per capita costs in both treatment strategies might not be representable for the general German population, in this study incremental costs are relevant for the model validation, not absolute costs of treatment strategies (143).

A further limitation is that the study period of 4 years allows a follow-up time of only 2.5 years, which is too short to assess long-term complication rates and costs for individuals with PCa-diagnosis. Especially in case of AS, long-term costs are of interest because published studies indicate, as described previously, that continuously accumulating costs of the surveillance scheme may exceed one-time costs of the RP surgery over time. In addition, treatment with EBRT in case of tumor progression under AS causes common complications, such as bowel problems; these complications usually develop after a longer period than analyzed here and are hence not represented in this analysis. Costs of the AS-strategy are likely underestimated when only the first years after treatment initiation are examined. In

addition, the relatively short study duration permits a pre-observation period of only 6 months; baseline co-morbidity may not be determined validly in this short period. In this study a trade-off between longer pre-observation period and longer follow-up period is necessary, and the longer follow-up is chosen.

Concerning cohort selection, another limitation of the AOK claims data is that randomization of individuals into treatment groups is not possible, as in any other observational study. Hence, estimated differences between groups might be attributed to unequal distribution of confounding variables. This bias is reduced by matching of individuals in treatment groups based on age and regression analysis adjusting for co-morbidity. This bias is, however, not fully eliminated in this study, because the number of variables available for confounder adjustment is limited in the AOK dataset. Inference of effectiveness outcome of AS and RP (e.g. number of complications) is, thus, afflicted by a high degree of uncertainty and validation of the effectiveness outcome of the decision model is not undertaken in this study.

Cohort selection is additionally limited by the fact that AS cannot be identified by a specific procedure code in the AOK claims dataset. To overcome this limitation AS is defined by a combination of procedures and corresponding codes, following the treatment guideline of DGU. However, not all of these procedures are specific for the AS scheme, such as PSA testing. Especially a distinction between AS and WW is difficult based on generic procedure codes. Analysis of coded prostate biopsies shows that only about 7% of men under AS receive a biopsy during follow-up. This suggests that only a minority of individuals in the AS-group actually is under AS according to treatment guidelines in the follow-up period, while the remaining men may be under WW. The cohort's life expectancy, however, is with a mean age of 70 years at baseline greater than 10 years and men are recommended to be treated with AS – according to treatment guideline; patients might actually be under AS, but are not surveyed by regular biopsies according to guideline. One reason for this might be that AS is a relatively novel treatment strategy; it may not have been performed regularly in Germany during the study period (2008-2011). Another reason might be that current studies report adverse effects of serial biopsies on erectile function and infectious complications and urologists may, thus, deviate from the recommended treatment protocol (82, 83). In consequence, the AS treatment group identified in this study is probably rather a mixture of

AS- and WW-patients; it might be more appropriate to refer to ‘conservative management’ than AS. The main purpose of this study is the validation of the PCa-model, though, and in the model transition probabilities for AS are extracted from a study comparing WW with RP, so that a mixture of AS and WW treatment effects is expected in the PCa-model, as well. Consequently, AOK data analysis is suitable to estimate outcomes comparable to the PCa-model, even though it is limited in reporting outcomes of AS.

5.3.3 Relevant health technologies and costs

A strength of this study is that the definition of health technologies and corresponding costs in claims data analysis (step 3) is set closely to the definition of the decision model. The same diagnostic and procedure codes reported in the decision model are used in claims data analysis. The use of these codes allows a distinct identification of PCa-relevant health technologies and corresponding costs. Thereby, all health states and treatments considered in the first part of the PCa-model – simulating treatment of early-stage PCa – are identified in claims data analysis.
A further strength of this study is that 2 different approaches are followed for the model validation. For one, all health technologies utilized during follow-up are considered (excess approach). Additionally, only PCa-relevant treatment utilization (disease-related approach) is analyzed. Comparison of outcomes of both approaches with model outcomes indicates whether the less complex excess approach is sufficient to make inferences on the decision model’s validity.

Several limitations are found in the definition of PCa-relevant health technologies. For one, utilization of antibiotics after biopsy is not considered in the RP-group, because it is not considered in the decision model either. In the AS-group antibiotics are considered; the decision model, however, does not report a specific pharmaceutical code (PZN) in this case. In claims data analysis, thus, antibiotics are considered that are reported in the literature to be predominantly used for antibiotic prophylaxis with prostate biopsies (65, 144). Secondly, in outpatient treatment, costs of single EBM codes are considered as defined in the decision model. In claims data analysis, utilization of materials is not included in these costs; however, this is a relatively small portion of the total outpatient treatment costs. Furthermore, the treatment lump sums for GP and specialist visits are considered as disease-

related costs if a treatment of interest (e.g. PSA or biopsy) is coded in the same quarter, as assumed in the PCa-model. Bias is possible in claims data analysis, though, because the outpatient visit could be motivated by another cause than PCa-treatment; in claims data no attribution of diagnoses to procedures is possible.

Finally, analysis of PCa-specific health technologies is prone to bias in cases where resource use for PCa treatment cannot be specifically circumscribed (e.g. GP visits). As it is the aim of this study to validate model outcomes, definition of PCa-specific costs is adapted to the model's definition. Disease-related costs may not represent all costs incurred by the SHI fund for the treatment of PCa, but they reflect the costs assumed in the decision model, which is of greater importance for this study.

5.3.4 Statistical methods for claims data analysis

Another strength of this study, concerning statistical methods of claims data analysis described in step 4 of the validation approach, is that matching of individuals by age makes treatment groups more comparable and thereby replicates a characteristic of the decision model's cohort, where individuals in the AS- and the RP-group are assumed to be of same age. Standard statistical methods, such as regression analysis, are used to adjust for co-morbidity and thereby further ensure comparability of claims data and model outcomes. Additionally, recycled predictions method is employed to estimate absolute values of complication rates and costs per treatment strategy; bootstrapping is used to calculate variability of estimates via 95% CIs.

Regarding analysis of claims data, a limitation is that matching of treatment groups by age results in a considerable loss of individuals in both treatment groups. Especially in the AS-group, where 17 patients (about 14%) are lost due to matching, representativeness of this treatment group is questionable. However, to ensure comparability between treatment groups it is necessary to disregard patients with a differing, usually higher, age in the AS-group than in the RP-group; according to treatment guideline, patients with a life expectancy smaller than 10 years are not eligible for AS treatment. Still, a greater patient heterogeneity is found in the claims data cohort than in the model, especially regarding co-morbidity, because the variables available for adjustment and matching are very limited in the AOK dataset.

Additionally, use of CCS for adjustment is limited; when no diagnosis is coded in the pre-observation period it cannot be differentiated whether the individual is free of disease or an existing illness is just not coded in the SHI scheme. Furthermore, in excess costs analysis CCS is not able to adjust for co-morbidities with high outpatient and pharmaceutical costs which are not related to PCa-treatment, because the Charlson index intends to assess inpatient mortality.

Furthermore, dependency of data due to matching is not accounted for by using conditional regression analysis. Dependency of data is very low, though, as matching is only performed for age; also no longitudinal analysis of single individuals is intended.

Outcomes of excess cost analysis are limited in their informative value because absolute costs do not reflect PCa treatment-specific costs. By adjusting for co-morbidity, differences in costs can be attributed to the initial treatment strategy, though. Disease-related cost analysis overcomes this limitation, despite its own pitfalls. The most severe limitation of disease-related cost analysis is the small number of observations which limits the validity of single cost estimates, especially in cases where technologies are utilized by only 1 or 2 individuals, such as surgical procedures for ED and IC. As a result, calculation of costs for single procedures may not be representative for certain health technologies in this study. To mitigate this limitation health technologies are combined in cost groups so that observation numbers increase.

Another limitation, which is common in the analysis of health care utilization, is that resource utilization is skewed; a small number of individuals utilizes health technologies and incurs substantial costs, whereas the remaining cohort has no utilization. In this case, unadjusted means and STDs have no informative value. To account for this problem, two-part regression models are employed in this study.

5.3.5 Changes to the decision model

A further strength of this study is that access to the originally published Markov model in Treeage is possible to conduct changes described in step 5. Thus, information is available for the model validation which is not reported in the publication of the PCa-model, for example codes defining PCa-related treatment. By the use of this information, outcomes estimated in claims data analysis are more comparable to outcomes of model simulation than without this additional information. The additional information also allows adapting the claims data

cohort as closely as possible – given available variables – to the model's cohort characteristics.

Access to the decision model also makes it possible to conduct additional analyses in the model assessing resource use and costs of single treatments, which are not reported in the original publication of Koerber et al. (2014); cost differences in specific treatments explain overall cost difference between treatment strategies. The total cost difference between AS and RP, for example, is almost exclusively determined by the cost difference found in RP surgery. The structure of the model itself is not changed in this analysis, which allows a validation of the originally published model.

Generally, a thorough understanding of model structure and assumptions facilitates the explanation of differences between AOK and model estimates, which is only possible when the original model can be retraced.

Furthermore, it is a strength of this study that in Treeage simulated model data are analyzed in SAS software. This way it is possible to calculate measures of uncertainty, such as CIs for incremental costs, which are not available in Treeage. Graphical presentation of model outcomes in form of bar charts and histograms is possible, too. Import of model data in SAS also allows formatting of data in a layout which makes calculation of differences between AOK and PCa-model outcomes as well as statistical hypothesis testing possible.

A limitation in the model analysis concerns the age distribution assumed in the model. As described previously the age distribution in the decision model needs to be adapted to the claims data cohort for a valid comparison of outcomes. Thus, it is assumed that all men in the model start treatment at the age of 70 years which is the mean age of the AOK cohort. This is, however, a simplification of the age distribution found in claims data; to be more precise the exact age structure of the AOK cohort should be represented in the model. The model by Koerber et al. does not allow incorporation of an age distribution at treatment initiation, though, and to change this assumption the model structure would have to be changed fundamentally. To estimate the impact of this limitation on outcomes, sensitivity analyses varying age at treatment initiation are conducted, as described previously. Results of the sensitivity analyses show that the overall agreement between AOK and model estimates is not affected by age at treatment initiation.

Another limitation in the model analysis is that the microsimulation cannot simulate different observation numbers for treatment strategies as in the claims data analysis where the RP-group includes twice as many individuals as the AS-group. The microsimulation is consequently simulated with 321 individuals in each treatment group; the AS-group in the microsimulation is thus 3 times larger than the AS-group in the claims dataset and the RP-group 1.5 times larger, respectively. In addition, the number of observations in microsimulation is so low that not every pathway through the model can be simulated and results between cost groups may vary because of differences in pathways and not differences in costs. To take this limitation into account Monte Carlo simulation is performed which evaluates the magnitude of the parameter uncertainty.

5.3.6 Comparison between model and claims data

One of the strengths in the comparison of claims data and model, described in step 6 of the validation approach in this study, is that graphical presentations as well as statistical methods are used to quantify the agreement between AOK and model outcomes. Simulated (bootstrapped) AOK samples and PCa-model samples are merged to calculate cost differences. Incremental costs of treatment strategies are compared between AOK data and model which is of most relevance for a decision-maker.

Furthermore, comparison of cost estimates is based on 2 different approaches; disease-related as well as excess costs are estimated in claims data analysis. Disease-related cost analysis allows the validation of costs of single treatments simulated in the model with outcomes of claims data analysis. In disease-related analysis, resource use of treated individuals and unit costs of PCa-relevant health technologies estimated in claims data analysis are compared with input parameters incorporated in the model. This validation of input parameters helps explain differences in per capita costs.

Additionally, sensitivity analyses are employed to assess the impact of claims data cohort's characteristics (newly diagnosed PCa-cases) and model cohort's assumptions (age at treatment initiation) on agreement between AOK and PCa-model outcomes.

A limitation, on the other hand, is that the explanatory power of the comparison between AOK data and PCa-model might be limited by greater patient heterogeneity in the claims data cohort than in the model, especially regarding co-morbidity. Additionally, the

reproduction of treatment strategies as defined in the model is limited in claims data which also infringes the informative value of the comparison.

Outcomes of the comparison are also affected by the way mortality is accounted for. In claims data analysis all resource use and costs accumulated until death are considered. In model simulation, on the other hand, costs of individuals dying during follow-up are not added to costs of the cohort. As mortality rates are small, this limitation minimally affects outcomes of the comparison between claims data and decision model, though.

Another limitation, regarding the sensitivity analysis where inclusion criteria for the claims data cohort are changed, becomes apparent: incidence of PCa-diagnosis is established in the pre-observation period; this period might be too short to ensure that PCa has not been diagnosed before study initiation, though. As overall agreement between AOK data and model outcome does not change, this limitation seems to have no notable influence on results.

Furthermore, quantitative comparison is limited to per capita costs in this study. Difference in utilization probability between model and claims data is not tested statistically because the simulation of the data needed for the quantitative comparison is extremely complex in the model by Koerber et al. (2014).

5.4 Practical implications

5.4.1 Generalizability of validation approach

The approach to claims data based validation described in chapter 3.1 can be applied for a variety of model validations. The strengths and limitations of claims data based validation discussed in chapter 5.3 may, however, not apply in every case. Generalizability of the approach to model validation based on claims data is assessed exemplarily for 2 cases: first, the validation of models comparing treatment for indications other than PCa and, second, validation based on administrative data sources other than German SHI data.

Examples of an adaptation of the validation approach to a variety of indications are highlighted in the following.

Concerning the decision on the validation level, effectiveness measures may be validated as well if outcome is replicable in claims data – contrary to the validation of the PCa-model where effectiveness outcome is measured in QALYs. Outcome measures concerning the utilization of health technologies can generally be reproduced validly in claims data. Health care utilization is, for example, an important measure of effectiveness in the treatment of cartilage defects in the knee, where the aim of novel treatment options, such as autologous chondrocyte implantation, is the prevention of joint replacement (145). Effectiveness measured as the number of replacement surgeries is, thus, an outcome which can be validly analyzed in SHI claims data.

Concerning cohort selection and study design, validation based on claims data is only useful if the indication of interest is validly coded in the dataset. The example of ED diagnosis shows that diseases whose treatment is not or only partially covered by SHI may not be coded validly in SHI data. Furthermore, severity of disease is not coded in SHI data and may only be deduced from diagnoses of co-morbidity – as in this study where advanced PCa is defined by diagnosis of metastases – or coding of relevant procedures. Diabetes is, for example, an indication where severity of disease can only be inferred from the intake of insulin and the co-morbidity status (146). The same applies to parameters describing patient characteristics which are essential to replicate the model's cohort; an example is smoking status which is not coded in SHI data, but is of paramount importance in studies on lung diseases (147). Furthermore, when indications are evaluated in a decision model where the

cost and effectiveness outcome can only be determined in a lifetime perspective, SHI claims data usually cannot provide valid estimates; due to technical and data protection issues it is generally not possible to analyze individual patient data with a lifetime perspective. An example where a long time of follow-up is necessary to assess differences between treatment groups, is the treatment of cartilage defects in the knee avoiding joint replacement mentioned above; replacement surgery usually becomes necessary 10 to 20 years after the initial therapy (145).

Regarding the definition of disease-relevant health states and technologies, adaptations of the validation approach presented for the PCa-model to the indication of interest might also be necessary. On the one hand, the validity of the claims data estimate of technology utilization depends on how detailed the definition of disease-related health technologies is reported in the decision model. If no information on codes for health technologies is provided in the model, it is difficult to accurately replicate the analysis of resource use and costs in claims data analysis. On the other hand, treatment of the indication of interest might not have any specific codes in the SHI scheme or it is not covered and thus not coded by SHI at all; again the example of ED treatment is applicable here.

Statistical methods for analysis of SHI claims data must be adapted to the indication of interest. Matching techniques might, as shown in this study, result in a loss of cases; when indications with a low number of observations, for example rare diseases, are studied, adjusting for confounding in regression analysis might be the better option (28).

Changes to the model structure and additional analysis conducted with the decision model might be necessary, depending on the cohort characteristics and the study design chosen for the validation. Apart from adaptation of age, as in this study, this might include change of gender distribution in the decision model or change of the model's perspective from a societal to a SHI insured population perspective.

Adaptation of the validation approach to administrative data sources other than the German SHI scheme is highlighted in the following, using the examples of the US Medicare/ Medicaid claims database and the UK General Practitioners Research Database. The UK GP Research Database includes medical information from inpatient and outpatient care as well as pharmaceutical prescriptions which converge at the GP in the UK national health service system (148). US Medicare covers inpatient, outpatient, and pharmaceutical claims of

individuals 65 years of age and older, whereas Medicaid covers disabled individuals; dual eligibility is possible (149).

The extent of clinical information included in the claims dataset has a particularly strong influence on the practicality of claims data for the validation of effectiveness measures. In the GP research database, for example, laboratory values are available and access to original medical records is possible (148), which would allow classifying PCa tumor stage based on PSA value and Gleason score, in contrast to SHI claims data.

Regarding the selection of a patient cohort reflecting the model's cohort, the UK GP research database includes a population-based dataset which is representative of the UK population and contains a large number of individuals (about 3 million) (148); based on the UK GP database, selection of a cohort validly reflecting model assumptions is very likely. US Medicare/ Medicaid, on the other hand, covers a population which is not representative of the general US population due to eligibility criteria; it might, thus, be more difficult to select a cohort equivalent to model assumptions (150). Information on socioeconomic status is not included in the UK GP research database, which biases cohort selection according to model assumptions.

To identify relevant health states and health technologies in the administrative data source, it is necessary that these can be uniquely identified. This is possible in both the UK GP research database and US Medicare/ Medicaid database, as diseases and procedures are identified with unique codes, similar to German SHI data (148, 150). In the UK GP research database, however, no information on health technology utilization outside the GPs' responsibility is included in the dataset, which leads to similar bias as in this study, where not all health care utilization is covered and, thus, not coded by SHI.

Other aspects of claims data discussed previously, for example confounding and data protection issues as well as disease-specific difference in validity of claims data, are applicable to all insurance systems.

5.4.2 Implications for model validation

For the validation of decision models this study generally implicates that access to the validated decision model is necessary, especially, if model assumptions, such as age distribution and run time, have to be adapted to the external data source to generate

comparable outcomes. Often additional analyses of the decision model are necessary to produce data in a format comparable to data of the claims data source.
Furthermore, it would facilitate validation of decision models based on claims data, if resource utilization and costs relevant for the SHI or SHI scheme insured community perspective are distinctly displayed in decision models, so that a calculation of the SHI perspective, in addition to the societal perspective, is simplified.

It is shown in this study that the relatively less time-consuming excess approach is able to validate total incremental costs of the decision model by itself; the results are comparable to the more complex disease-related approach where health technologies are defined according to the model's definition of disease-related technology utilization. If in a validation based on claims data the concordance of claims data cohort and model cohort can be ensured and statistical techniques are employed to make treatment strategies comparable (matching, regression analysis), an excess cost approach might be sufficient to validate total costs. This study, however, shows that even if agreement between estimates of total costs is sufficient to assume that the model simulates real-world costs, this might not be the case for costs of individual treatments considered in the model. The disease-related approach is useful in this case to validate incremental costs of single treatments. If in a model validation resource utilization and costs of individual health technologies are not relevant, the less complex excess approach may be suitable.

Concerning the results of the validation, it is apparent that results are subject to interpretation. Graphical and statistical presentation can only indicate an agreement between outcome simulated in the model and outcome observed in the external data source. In addition, the agreement between model and claims data has to be interpreted in light of the uncertainty afflicting the individual parameters.
Whether this agreement is sufficient to infer that the model simulates real-world outcomes validly depends greatly on the limitations of the external data source, most of all limitations of the claims data analysis in replicating the model's structure and cohort assumptions. SHI data may be the gold standard for real-world resource use and costs, if however the study design (e.g. length of follow-up) and cohort (e.g. patient characteristics and treatment groups) of the decision model cannot be replicated in the claims data, the estimates of

resource use and costs, in turn, are not comparable to the model outcomes and, thus, not suitable for validation.

5.4.3 Implications for industry

A potential application of the validation of decision models based on claims data is described exemplarily in the context of the integration of novel health technologies in the SHI system's benefit catalogue.

The clinical effectiveness and costs – key components of a new medical technology's value – are usually uncertain at the point of coverage decision-making by a health care payer. This is in part due to the fact that market-approval agencies ask for evidence regarding safety and efficacy, while coverage decision makers are predominantly interested in effectiveness and real-world costs. Decision models may be employed, in this context, to incorporate available evidence and predict real-world effectiveness and costs; these predictions of effectiveness and costs are, however, still subject to high variability due to the uncertainty in the underlying evidence (151).

Under such uncertainty the decision maker may reject coverage and thus deny patients access to potentially beneficial technologies. A positive coverage decision, on the other hand, may lead to reimbursement of potentially clinically or cost ineffective medical technologies and strain already tight budgets (152). To address this uncertainty, market entry can be accompanied by arrangements where further evidence regarding the performance and utilization of the technology is collected alongside use; new evidence can either be incorporated in the decision model or used to validate the predictions of the original model. That way the final coverage decision can be delayed until sufficient evidence is available, while enabling early patient access to novel technologies (45). These schemes are usually referred to as "managed entry agreements" (MEA) (Figure 5-1) (153). In Germany, since the introduction of the Healthcare Provision Act (GKV-Versorgungsstrukturgesetz) in 2011, novel examination and treatment methods can be covered by SHI on the condition that clinical data are gathered alongside use in clinical practice (51).

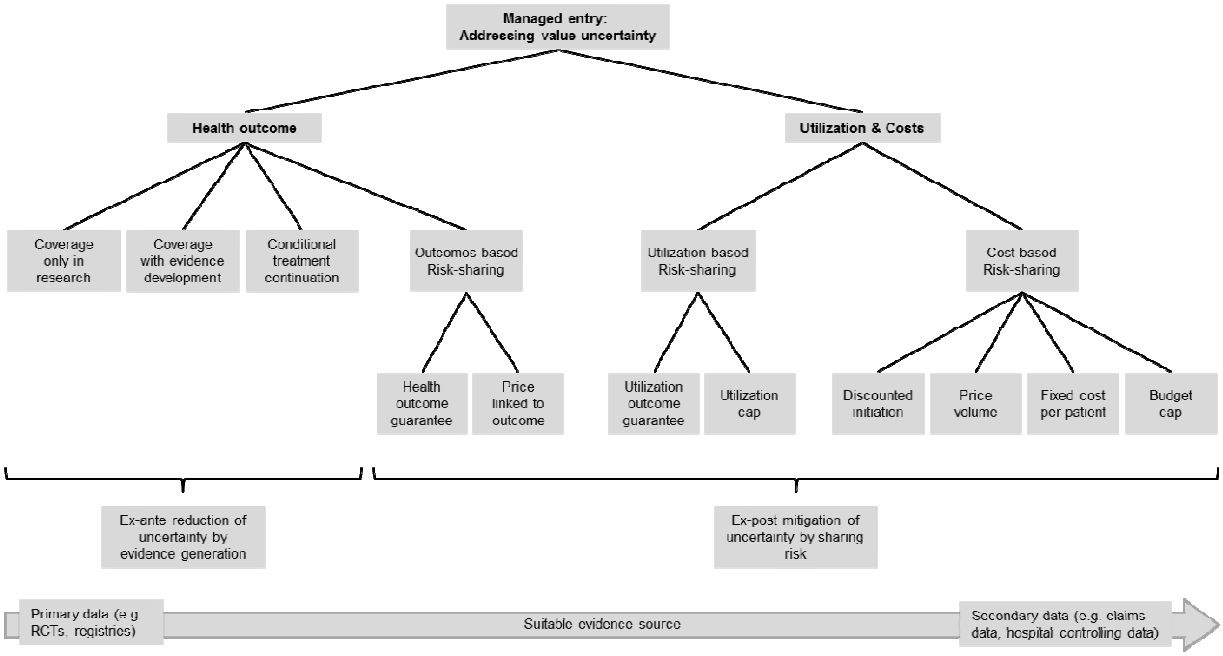


Figure 5-1: Typology of managed entry agreements

As MEA rely on the collection of information, the question arises where to find suitable evidence to validate the predicted outcomes of the model. Original data collected within randomized controlled trials are seen as the gold standard for evidence development. However, they inhibit limitations regarding evidence on effectiveness and cost-effectiveness in clinical practice, which is of primary interest for decision makers (38, 39). Original data collected within registries may overcome some of these problems, but have their own distinct limitations (42, 154). In both cases, data collection is likely to incur high costs to manufacturers or health care payers. Given the already high costs of clinical research in the development of medical innovations (155), there is a risk that the costs of evidence-collection exceed the potential value associated with the additional evidence. A valid option for the validation of decision models in line with a MEA are, consequently, claims data (156). Especially, when MEA are formed between a manufacturer of a novel technology and a SHI fund, it is of interest for the SHI fund to validate the models predictions with a patient cohort reflecting the funds insured community; the SHI funds own claims are obviously the most valid data source to achieve this. Use of claims data is useful for the validation of decision models conducted in line with MEAs which concern not only one insurance fund but the whole SHI insurance system, too. The relevant patient cohort, in this case, is the SHI scheme insured community.

The suitability of claims data is especially high for agreements that concern utilization or cost outcomes of a novel technology. In agreements where the level of reimbursement is dependent on utilization and cost outcomes – so called risk-sharing agreements – validation of these outcomes based on claims data is useful.

6 Conclusion

To answer the research question '*Are claims data of health insurance funds suitable to externally validate decision-analytic models?*' this study proposes a step-wise validation approach. Applicability of the validation approach is assessed on the exemplary validation of a Markov model comparing treatment of localized PCa based on claims data of a large German SHI fund. Strengths and limitations of claims data based validation is discussed for each validation step. Generalizability and implications of claims data based model validation are presented.

Concerning the medical point of view, the analysis indicates that in the first years after treatment initiation costs of AS are significantly lower than the costs of radical therapy with RP for early-stage PCa – predominantly due to the high initial costs of the RP surgery. Treatment of complications following initial therapy has a very small impact on costs, albeit ED and IC have a substantial impact on patients' quality of life. These results are consistent in the 2 different cost analysis approaches presented in claims data analysis – excess cost and disease-related cost analysis – and are predicted by the decision model, as well.
Concerning the validation of the decision model, a degree of an overall agreement between the AOK data and PCa-model outcome is found which is sufficient to assume that the model simulates short-term real-world resource use and costs of AS compared to RP validly. The outcomes of excess cost analysis alone are able to validate total incremental costs in this study, which would make the more extensive disease-related analysis unnecessary; disease-related analysis is, however, useful to validate incremental costs of individual treatments of PCa. Validation of individual treatments shows that resource use and costs simulated in the decision model are most valid for inpatient care, whereas outpatient care differs significantly from observed outcomes in claims data.

The exemplary model validation reveals strengths and limitations of claims data for model validation, which are characteristic for claims data based model validation in general.
Claims data are able to provide evidence on real-world resource utilization and, with limitations regarding clinical information, effectiveness of a wide range of indications and treatments in a large patient cohort. Validation based on claims data is especially suitable

when the decision maker, interested in the validity of the model in question, is the insurance fund providing access to the claims data; the claims data cohort is representative of the insured population of the insurance fund. Claims data may be regarded as the gold standard for real-world evidence on resource utilization and costs in this case. Furthermore, use of claims data for the validation of decision models is less costly and time consuming than collection of primary evidence in RCTs.

Suitability of claims data based validation is, however, limited concerning the replication of the decision model's structure and cohort assumptions. For one, the identification of distinct health states is limited, because clinical information, such as laboratory values or tumor stage, is not included in SHI claims data. Also, due to non-randomization and a restricted number of variables available to adjust for confounding, comparability of treatment groups is limited in SHI claims data analysis; claims data analysis may not reflect model analysis which is usually based on randomized trials. Furthermore, distinct identification of health technology utilization and corresponding costs is not possible, if the technology of interest is not specifically coded. Claims data are, generally, collected for billing purposes; diagnoses and technology utilization are only coded if they are relevant for reimbursement, which biases outcomes of model validation in cases where treatment is not covered by the insurance fund.

The suitability of claims data for the validation of the decision model of interest eventually depends on the ability of the claims data source to reflect the model's patient cohort and outcome measures. If study design and cohort assumptions of the decision model cannot be replicated in the claims data, the estimates of resource use and costs, in turn, are not comparable to the model outcomes and, thus, not suitable for validation.

Weighing up of strengths against limitations of claims data based validation has to be made for each case independently. The proposed validation approach indicates critical aspects in the validation based on claims data, which may support researchers and decision makers in their decision on the suitability of claims data.

Further research is necessary to assess the applicability of the validation approach in models with indications other than PCa and validation based on other external data sources, for example registry data.

Methodological aspects of the comparison of claims data and model outcomes, such as statistical techniques that test difference between estimates, need to be refined. As seen in this study, the validity of statistical methods, for example, depends on the presence of paired observations in the two data sources. A systematic assessment of various statistical methods for different types of external model validation would be a useful guidance for model validation.

It is also interesting to assess the practical implication of claims data based validation for industry. For example, the practicality of claims data based model validation for the management of novel health technologies in the SHI benefit catalogue should be assessed in further research.

Finally, further research is needed to assess whether claims data are suitable for model validation apart from external parameter validation. This study indicates that validation of model structure is possible based on claims data. Studies on the suitability of claims data to assess predictive validity of decision models might be interesting, as well.

References

1. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press, 2005.
2. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press, 2006.
3. Annemans L, Geneste B, Jolain B. Early modelling for assessing health and economic outcomes of drug therapy. Value Health. 2000; 3: 427-34.
4. Hunink MG, Weinstein MC, Wittenberg E, et al. Decision making in health and medicine: integrating evidence and values. 2nd ed. Cambridge: Cambridge University Press, 2014.
5. Paisley S. Classification of evidence in decision-analytic models of cost-effectiveness: a content analysis of published reports. Int J Technol Assess Health Care. 2010; 26: 458-62.
6. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. Pharmacoeconomics. 2000; 17: 461-77.
7. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health. 2003; 6: 9-17.
8. Bilcke J, Beutels P, Brisson M, et al. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Med Decis Making. 2011; 31: 675-92.
9. Caro JJ, Moller J. Decision-analytic models: current methodological challenges. Pharmacoeconomics. 2014; 32: 943-50.
10. Elliott H, Popay J. How are policy makers using evidence? Models of research utilisation and local NHS policy making. J Epidemiol Community Health. 2000; 54: 461-8.
11. Kopec J, Fines P, Manuel D, et al. Validation of population-based disease simulation models: a review of concepts and methods. BMC Public Health. 2010; 10: 710.
12. Sendi PP, Craig BA, Pfluger D, et al. Systematic validation of disease models for pharmacoeconomic evaluations. J Eval Clin Pract. 1999; 5: 283-95.

13. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. Pharmacoeconomics. 2000; 17: 501-13.
14. Hammerschmidt T, Goertz A, Wagenpfeil S, et al. Validation of health economic models: the example of EVITA. Value Health. 2003; 6: 551-59.
15. Becker C, Leidl R, Stollenwerk B. Entscheidungsanalytische Modellierung in der ökonomischen Evaluation. Gesundh ökon Qual manag. 2010; 15: 260,64.
16. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation. Med Decis Making. 2012; 32: 733-43.
17. Philips Z, Bojke L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics. 2006; 24: 355-71.
18. Sonnenberg FA, Roberts MS, Tsevat J, et al. Toward a peer review process for medical decision analysis models. Med Care. 1994; 32: JS52-64.
19. Weinstein M, Toy E, Sandberg E, et al. Modeling for health care and other policy decisions: uses, roles, and validity. Value Health. 2001; 4: 348-61.
20. Ishida H, Wong JB, Hino K, et al. Validating a Markov model of treatment for hepatitis C virus-related hepatocellular carcinoma. Methods Inf Med. 2008; 47: 529-40.
21. van Staa TP, Leufkens HG, Zhang B, et al. A comparison of cost effectiveness using data from randomized trials or actual clinical practice: selective cox-2 inhibitors as an example. PLoS Med. 2009; 6: e1000194.
22. McEwan P, Foos V, Palmer JL, et al. Validation of the IMS CORE Diabetes Model. Value Health. 2014; 17: 714-24.
23. Eddy DM, Schlessinger L. Validation of the Archimedes Diabetes Model. Diabetes Care. 2003; 26: 3102-10.
24. Palmera AJ, Rozea S, Valentinea WJ, et al. Validation of the CORE diabetes model against epidemiological and clinical studies. Curr Med Res Opin. 2004; 20: S27-S40.
25. Stollenwerk B, Gerber A, Lauterbach KW, et al. The German Coronary Artery Disease Risk Screening Model: development, validation, and application of a decision-analytic model for coronary artery disease prevention with statins. Med Decis Making. 2009; 29: 619-33.

26. Hay J, Jackson J. Panel 2: methodological issues in conducting pharmacoeconomic evaluations--modeling studies. Value Health. 1999; 2: 78-81.
27. Soto J. Health economic evaluations using decision analytic modeling. Principles and practices--utilization of a checklist to their development and appraisal. Int J Technol Assess Health Care. 2002; 18: 94-111.
28. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991.
29. Kleijnen JPC. Verification and validation of simulation models. Eur J Oper Res. 1995; 82: 145-62.
30. Sargent RG. Verification and validation of simulation models. In: Kuhl M, Steiger N, Armstrong F, et al., eds., 2005 Winter Simulation Conference, 2005.
31. Janson M, Carlsson P, Haglind E, et al. Data validation in an economic evaluation of surgery for colon cancer. Int J Technol Assess Health Care. 2005; 21: 246-52.
32. Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS ONE. 2011; 6: e17401.
33. Kaur G, Arora AS, Jain VK. Prediction of BP reactivity to talking using hybrid soft computing approaches. Comput Math Methods Med. 2014; 2014: 13.
34. Hulley SB, Cummings SR, Browner WS, et al. Designing clinical research. 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2013.
35. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996; 25: 435-42.
36. Fünftes Buch Sozialgesetzbuch - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das durch Artikel 3 des Gesetzes vom 22. Dezember 2011 (BGBl. I S. 3057) geändert worden ist. §§ 294-303, 1988.
37. Garrison LP, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007; 10: 326-35.
38. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000; 342: 1887-92.
39. Avorn J. In defense of pharmacoepidemiology — embracing the yin and yang of drug research. N Engl J Med. 2007; 357: 2219-21.

40. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58: 323-37.
41. Miller FG, Pearson SD. Coverage with evidence development: ethical issues and policy implications. Med Care. 2008; 46: 746-51.
42. Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research. Value Health. 2009; 12: 1053-61.
43. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. J Clin Epidemiol. 2004; 57: 131-41.
44. Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. Value Health. 2009; 12: 1062-73.
45. McCabe CJ, Stafinski T, Edlin R, et al. Access with evidence development schemes. Pharmacoeconomics. 2010; 28: 143-52.
46. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. Med Care. 2002; 40: IV-26-35.
47. Mitra N, Indurkhya A. A propensity score approach to estimating the cost–effectiveness of medical therapies from observational data. Health Econ. 2005; 14: 805-15.
48. Zhou X-H, Melfi CA, Hui SL. Methods for comparison of cost data. Ann Intern Med. 1997; 127: 752-56.
49. Schubert I, Köster I, Küpper-Nybelen J, et al. Versorgungsforschung mit GKV-Routinedaten. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2008; 51: 1095-105.
50. Busse R, Riesberg A. Health care systems in transition: Germany. Copenhagen: WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2004.

51. Fünftes Buch Sozialgesetzbuch - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das durch Artikel 3 des Gesetzes vom 22. Dezember 2011 (BGBl. I S. 3057) geändert worden ist. § 137e, 1988.
52. Schwarzkopf L, Menn P, Leidl R, et al. Excess costs of dementia disorders and the role of age and gender - an analysis of German health and long-term care insurance claims data. BMC Health Serv Res. 2012; 12: 165.
53. Deutsche Krankenhausgesellschaft e.V., Spitzenverband der gesetzlichen Krankenversicherungen. Datenübermittlung nach § 301 Abs. 3 SGB V, Stand: 1.8.2012. Düsseldorf, 2012.
54. Fünftes Buch Sozialgesetzbuch - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das durch Artikel 3 des Gesetzes vom 22. Dezember 2011 (BGBl. I S. 3057) geändert worden ist. § 301, 1988.
55. Fünftes Buch Sozialgesetzbuch - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das durch Artikel 3 des Gesetzes vom 22. Dezember 2011 (BGBl. I S. 3057) geändert worden ist. §295 Abs 2, 1988.
56. Willison DJ. Health services research and personal health information: privacy concerns, new legislation and beyond. Can Med Assoc J. 1998; 159: 1378-80.
57. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer, 2013.
58. Kraywinkel K, Bertz J, Laudi A, et al. Epidemiologie und Früherkennung häufiger Krebserkrankungen in Deutschland. Berlin: Robert Koch-Institut, 2012.
59. Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Krebs in Deutschland 2009/ 2010. 9th ed. Berlin, 2013.
60. Bray F, Lortet-Tieulent J, Ferlay J, et al. Prostate cancer incidence and mortality trends in 37 European countries: an overview. Eur J Cancer. 2010; 46: 3040-52.
61. De Angelis R, Sant M, Coleman MP, et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. Lancet Oncol. 2014; 15: 23-34.

62. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet. 2014; 384: 2027-35.
63. Luengo-Fernandez R, Leal J, Gray A, et al. Economic burden of cancer across the European Union: a population-based cost analysis. Lancet Oncol. 2013; 14: 1165-74.
64. Statistisches Bundesamt. Bevölkerung Deutschlands bis 2060 - 12. koordinierte Bevölkerungsvorausberechnung. 2009.
65. Mottet N, Bastian PJ, Bellmunt J, et al. Guidelines on prostate cancer. European Association of Urology, 2014.
66. Sobin L, Gospodarowicz M, Wittekind C. TNM classification of malignant tumours. In: Cancer IUA, ed. 7th ed. Oxford: Wiley-Blackwell, 2009.
67. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol. 2013; 64: 876-92.
68. Epstein JI. An update of the Gleason Grading System. J Urol. 2010; 183: 433-40.
69. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms Langversion 3.1 AWMF Registernummer: 034/022OL. 2014.
70. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998; 280: 969-74.
71. Popiolek M, Rider JR, Andrén O, et al. Natural history of early, localized prostate cancer: a final report from three decades of follow-up. Eur Urol. 2013; 63: 428-35.
72. Ilic D, Neuberger Molly M, Djulbegovic M, et al. Screening for prostate cancer. Cochrane Database Syst Rev: John Wiley & Sons, Ltd, 2013.
73. Bianco FJ, Jr., Scardino PT, Eastham JA. Radical prostatectomy: long-term cancer control and recovery of sexual and urinary function ("trifecta"). Urology. 2005; 66: 83-94.
74. Ficarra V, Novara G, Rosen RC, et al. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. Eur Urol. 2012; 62: 405-17.

75. Ficarra V, Novara G, Ahlering TE, et al. Systematic review and meta-analysis of studies reporting potency rates after robot-assisted radical prostatectomy. Eur Urol. 2012; 62: 418-30.
76. Pinkawa M, Gagel B, Piroth MD, et al. Erectile dysfunction after external beam radiotherapy for prostate cancer. Eur Urol. 2009; 55: 227-36.
77. Robinson JW, Moritz S, Fung T. Meta-analysis of rates of erectile function after treatment of localized prostate carcinoma. Int J Radiat Oncol Biol Phys. 2002; 54: 1063-8.
78. Hayes JH, Ollendorf DA, Pearson SD, et al. Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis. Ann Intern Med. 2013; 158: 853-60.
79. Godtman RA, Holmberg E, Khatami A, et al. Outcome following active surveillance of men with screen-detected prostate cancer. Results from the Goteborg randomised population-based prostate cancer screening trial. Eur Urol. 2013; 63: 101-7.
80. Bastian PJ, Carter BH, Bjartell A, et al. Insignificant prostate cancer and active surveillance: from definition to clinical implications. Eur Urol. 2009; 55: 1321-30.
81. Klotz L, Zhang L, Lam A, et al. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. J Clin Oncol. 2010; 28: 126-31.
82. Braun K, Ahallal Y, Sjoberg DD, et al. Effect of repeated prostate biopsies on erectile function in men on active surveillance for prostate cancer. J Urol. 2014; 191: 744-49.
83. Ehdaie B, Vertosick E, Spaliviero M, et al. The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. J Urol. 2014; 191: 660-4.
84. Klotz L. Active surveillance, auality of life, and cancer-related anxiety. Eur Urol. 2013; 64: 37-39.
85. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med. 2014; 370: 932-42.
86. Johansson E, Steineck G, Holmberg L, et al. Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. Lancet Oncol. 2011; 12: 891-9.

87. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med. 2012; 367: 203-13.
88. Lane JA, Donovan JL, Davis M, et al. Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. Lancet Oncol. 2014; 15: 1109-18.
89. Wiegel T, Stockle M, Bartkowiak D. PREFEREnce-based randomized evaluation of treatment modalities in low or early intermediate-risk prostate cancer. Eur Urol. 2015; 67: 1-2.
90. Aizer AA, Chen MH, Hattangadi J, et al. Initial management of prostate-specific antigen-detected, low-risk prostate cancer and the risk of death from prostate cancer. BJU Int. 2014; 113: 43-50.
91. Herden J, Ernstmann N, Schnell D, et al. The HAROW study: an example of outcomes research : A prospective, non-interventional study comparing treatment options in localized prostate cancer. Urologe. 2014; 53: 1743-52.
92. Punnen S, Cowan JE, Chan JM, et al. Long-term health-related quality of life after primary treatment for localized prostate cancer: results from the CaPSURE registry. Eur Urol. 2014.
93. Resnick MJ, Koyama T, Fan K-H, et al. Long-term functional outcomes after treatment for localized prostate cancer. N Engl J Med. 2013; 368: 436-45.
94. Grimm P, Billiet I, Bostwick D, et al. Comparative analysis of prostate-specific antigen free survival outcomes for patients with low, intermediate and high risk prostate cancer treatment by radical therapy. Results from the Prostate Cancer Results Study Group. BJU Int. 2012; 109 Suppl 1: 22-9.
95. Bul M, Zhu X, Valdagni R, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. Eur Urol. 2013; 63: 597-603.
96. Andersson SO, Andren O, Lyth J, et al. Managing localized prostate cancer by radical prostatectomy or watchful waiting: Cost analysis of a randomized trial (SPCG-4). Scand J Urol Nephrol. 2011; 45: 177-83.
97. Wilson LS, Tesoro R, Elkin EP, et al. Cumulative cost pattern comparison of prostate cancer treatments. Cancer. 2007; 109: 518-27.

98. Molinier L, Castelli C, Bauvin E, et al. Cost study of the clinical management of prostate cancer in France: results on the basis of population-based data. Eur J Health Econ. 2011; 12: 363-71.
99. Snyder CF, Frick KD, Blackford AL, et al. How does initial treatment choice affect short-term and long-term costs for clinically localized prostate cancer? Cancer. 2010; 116: 5391-9.
100. Perlroth DJ, Bhattacharya J, Goldman DP, et al. An economic analysis of conservative management versus active treatment for men with localized prostate cancer. J Natl Cancer Inst Monogr. 2012; 2012: 250-7.
101. Manoharan M, Eldefrawy A, Katkoori D, et al. Comparison of urologist reimbursement for managing patients with low-risk prostate cancer by active surveillance versus total prostatectomy. Prostate Cancer Prostatic Dis. 2010; 13: 307-10.
102. Keegan KA, Dall'Era MA, Durbin-Johnson B, et al. Active surveillance for prostate cancer compared with immediate treatment: an economic analysis. Cancer. 2012; 118: 3512-8.
103. Eldefrawy A, Katkoori D, Abramowitz M, et al. Active surveillance vs. treatment for low-risk prostate cancer: a cost comparison. Urol Oncol. 2013; 31: 576-80.
104. Corcoran AT, Peele PB, Benoit RM. Cost comparison between watchful waiting with active surveillance and active treatment of clinically localized prostate cancer. Urology. 2010; 76: 703-7.
105. Koerber F, Waidelich R, Stollenwerk B, et al. The cost-utility of open prostatectomy compared with active surveillance in early localised prostate cancer. BMC Health Serv Res. 2014; 14: 163.
106. Lyth J, Andersson SO, Andren O, et al. A decision support model for cost-effectiveness of radical prostatectomy in localized prostate cancer. Scand J Urol Nephrol. 2012; 46: 19-25.
107. Cooperberg MR, Ramakrishna NR, Duff SB, et al. Primary treatments for clinically localised prostate cancer: a comprehensive lifetime cost-utility analysis. BJU Int. 2013; 111: 437-50.
108. Roehrborn CG, Black LK. The economic burden of prostate cancer. BJU Int. 2011; 108: 806-13.

109. Deutsche Gesellschaft für Urologie, Deutsche Krebshilfe e.V. Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms. 2011.
110. Bremner KE, Chong CA, Tomlinson G, et al. A review and meta-analysis of prostate cancer utilities. Med Decis Making. 2007; 27: 288-98.
111. Stewart ST, Lenert L, Bhatnagar V, et al. Utilities for prostate cancer health states in men aged 60 and older. Med Care. 2005; 43: 347-55.
112. Liu D, Lehmann HP, Frick KD, et al. Active surveillance versus surgery for low risk prostate cancer: a clinical decision analysis. J Urol. 2012; 187: 1241-6.
113. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med. 2011; 364: 1708-17.
114. Ollendorf DA, Hayes JH, McMahon PM, et al. Active surveillance & radical prostatectomy for the management of low-risk, clinically-localized prostate cancer. In: Institute for Clinical and Economic Review, ed., 2009.
115. Kassenärztliche Bundesvereinigung.
116. Institut für das Entgeltsystem im Krankenhaus.
117. Research Institute of the AOK.
118. Hayes JH, Ollendorf DA, Pearson SD, et al. Active surveillance compared with initial treatment for men with low-risk prostate cancer: a decision analysis. JAMA. 2010; 304: 2373-80.
119. AOK Baden-Württemberg. Geschäftsbericht 2013/ 2014. 2014.
120. Schnee M. Sozioökonomische Strukturen und Morbidität in den gesetzlichen Krankenkassen. In: Böcken J, Braun B, Amhof R, eds., Gesundheitsmonitor 2008 Gesundheitsversorgung und Gestaltungsoptionen aus der Perspektive der Bevölkerung. Gütersloh: Verlag Bertelsmann Stiftung, 2008.
121. Rothman KJ, Greenland S, Lash T. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
122. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol. 2000; 29: 891-98.

123. Grobe TG. Stationäre Versorgung - Krankenhausbehandlung. In: Swart E, Ihle P, eds., Routinedaten im Gesundheitswesen - Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Verlag Hans Huber, Hogrefe AG, 2005.
124. Kerek-Bodden H, Heuer J, Brenner G, et al. Morbiditäts- und Inanspruchnahmeanalysen mit personenbezogenen Abrechnungsdaten aus Arztpraxen. In: Swart E, Ihle P, eds., Routinedaten im Gesundheitswesen - Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern: Verlag Hans Huber, Hogrefe AG, 2005.
125. Travison TG, Sand MS, Rosen RC, et al. The natural progression and regression of erectile dysfunction: follow-up results from the MMAS and MALES Studies. J Sex Med. 2011; 8: 1917-24.
126. Graubard BI, Korn EL. Predictive margins with survey data. Biometrics. 1999; 55: 652-59.
127. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med. 2000; 19: 1141-64.
128. Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. J Health Serv Res Policy. 1998; 3: 233-45.
129. Institute for Quality and Efficiency in Health Care (IQWiG). Working Paper Cost Estimation. 2009.
130. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. J Health Serv Res Policy. 2004; 9: 197-204.
131. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. J Health Econ. 2004; 23: 525-42.
132. Afifi AA, Kotlerman JB, Ettner SL, et al. Methods for improving regression analysis for skewed continuous or counted responses. Annu Rev Public Health. 2007; 28: 95-111.
133. Salonia A, Castagna G, Saccà A, et al. Is Erectile Dysfunction a Reliable Proxy of General Male Health Status? The Case for the International Index of Erectile Function-Erectile Function Domain. J Sex Med. 2012; 9: 2708-15.
134. Hall WH, Ramachandran R, Narayan S, et al. An electronic application for rapidly calculating Charlson comorbidity score. BMC Cancer. 2004; 4: 94.

135. Briggs AH. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics. 2000; 17: 479-500.
136. MacKinnon JG. Bootstrap hypothesis testing. In: Belsley DA, Kontoghiorghes EJ, eds., Handbook of computational econometrics. Chichester: John Wiley & Sons Ltd, 2009.
137. Krahn MD, Zagorski B, Laporte A, et al. Healthcare costs associated with prostate cancer: estimates from a population-based study. BJU Int. 2010; 105: 338-46.
138. Hayes JH, Ollendorf DA, Pearson SD, et al. Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis. Ann Intern Med. 2013; 158: 853-60.
139. Stausberg J, Lehmann N, Kaczmarek D, et al. Reliability of diagnoses coding with ICD-10. Int J Med Inform. 2008; 77: 50-7.
140. Erler A, Beyer M, Muth C, et al. Garbage in - garbage out? Validity of coded diagnoses from GP claims records. Gesundheitswesen. 2009; 71: 823-31.
141. Klaus B, Ritter A, Grosse Hulsewiesche H, et al. Study of the quality of codification of diagnoses and procedures under DRG conditions. Gesundheitswesen. 2005; 67: 9-19.
142. Rapiti E, Fioretta G, Schaffar R, et al. Impact of socioeconomic status on prostate cancer diagnosis, treatment, and prognosis. Cancer. 2009; 115: 5556-65.
143. Jaunzeme J, Eberhard S, Geyer S. How "representative" are SHI (statutory health insurance) data? Demographic and social differences and similarities between an SHI-insured population, the population of Lower Saxony, and that of the Federal Republic of Germany using the example of the AOK in Lower Saxony. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2013; 56: 447-54.
144. Bootsma AMJ, Laguna Pes MP, Geerlings SE, et al. Antibiotic prophylaxis in urologic procedures: a systematic review. Eur Urol. 2008; 54: 1270-86.
145. Koerber F, Rolauffs B, Rogowski W. Early evaluation and value-based pricing of regenerative medicine technologies. Regen Med. 2013; 8: 747-58.
146. Chang HY, Weiner JP, Richards TM, et al. Predicting costs with diabetes complications severity index in claims data. Am J Manag Care. 2012; 18: 213-9.
147. Cao Z, Zou KH, Baker CL, et al. Respiratory-related medical expenditure and inpatient utilisation among COPD patients receiving long-acting bronchodilator therapy. J Med Econ. 2011; 14: 147-58.

148. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. Br J Clin Pharmacol. 1998; 45: 419-25.
149. Hennessy S, Leonard CE, Palumbo CM, et al. Quality of Medicaid and Medicare data obtained through Centers for Medicare and Medicaid Services (CMS). Med Care. 2007; 45: 1216-20.
150. Hennessy S, Bilker WB, Weber A, et al. Descriptive analyses of the integrity of a US Medicaid claims database. Pharmacoepidemiol Drug Saf. 2003; 12: 103-11.
151. Rogowski W, Hartz S, John J. Clearing up the hazy road from bench to bedside: A framework for integrating the fourth hurdle into translational medicine. BMC Health Serv Res. 2008; 8: 194.
152. Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. Int J Technol Assess Health Care. 2007; 23: 425-32.
153. Brandes A, Schwarzkopf L, Rogowski WH. Using claims data for evidence generation in managed entry agreements. Int J Technol Assess Health Care. 2016; 15: 1-9
154. Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force. Value Health. 2009; 12: 1044-52.
155. Morgan S, Grootendorst P, Lexchin J, et al. The cost of drug development: a systematic review. Health Policy. 2011; 100: 4-17.
156. O'Malley SP, Selby WS, Jordan E. A successful practical application of coverage with evidence development in Australia: Medical Services Advisory Committee interim funding and the PillCam Capsule Endoscopy Register. Int J Technol Assess Health Care. 2009; 25: 290-6.

Appendix

A Additional figures and tables

Treatment	P utilization		Resource		Unit costs		Costs per treated individual			
	RP	AS	RP	AS	RP	AS	RP		AS	
							Mean	STD	Mean	STD
Inpatient										
RP w/ o complications	169 (0.79)	4 (0.04)	1.00	1.00	6,416.59	5,294.86	6,416.59	903.95	6,294.86	794.92
RP with complications	41 (0.19)	0 (0.00)	1.00	0.00	10,140.88	0.00	10,140.88	4,417.47	0.00	-
Treatment stricture (RP)	10 (0.05)	1 (0.01)	1.10	1.00	4,297.91	4,823.91	4,727.70	5,240.59	4,823.91	-
Penis prosthesis (ED)	1 (0.01)	0 (0.00)	1.00	0.00	7,586.13	0.00	7,586.13	-	0.00	-
Artificia urethral sphincter (IC)	0 (0.00)	1 (0.01)	0.00	1.00	0.00	11,732.16	0.00	-	11,732.16	-
Sling surgery (IC)	1 (0.01)	0 (0.00)	1.00	0.00	5,865.31	0.00	5,865.31	3,546.96	0.00	-
Treatment urosepsis (AS)	5 (0.02)	2 (0.02)	1.00	1.00	2,015.35	2,002.46	2,015.35	634.94	2,002.46	727.85
TURP(RPH)	0 (0.00)	10 (0.09)	0.00	1.10	0.00	4,715.19	0.00	0.00	7,333.65	8,135.67
Outpatient										
Follow-up RP PSA test (urologist)	154 (0.72)	4 (0.04)	8.62	3.50	24.86	23.30	184.62	68.26	53.81	29.83
Treatment ED (urologist)	21 (0.10)	4 (0.04)	6.00	7.75	19.40	19.48	126.54	69.17	150.93	26.13
Treatment ED (GP)	20 (0.09)	5 (0.05)	4.75	6.20	34.86	35.70	181.27	135.48	221.51	124.88
Treatment IC (urologist)	58 (0.27)	7 (0.07)	6.24	9.14	19.97	19.32	133.59	68.91	176.66	26.13
Treatment IC (GP)	56 (0.26)	5 (0.05)	5.54	5.80	35.49	35.70	212.33	147.60	207.21	139.36
EBRT after AS (radiotherapist)	0 (0.00)	11 (0.10)	0.00	1.82	0.00	1,380.33	0.00	0.00	2,509.70	1,025.99
AS: PSA test & biopsies (urologist)	0 (0.00)	107 (1.00)	0.00	7.36	0.00	43.60	0.00	0.00	82.44	39.22

Table A-1: Probability of utilization, resource use, unit costs (€), and per capita costs of treated individuals (€), by treatment – unadjusted (AOK)

Treatment	P utilization		Resource use		Unit costs		Costs per treated individual			
	RP	AS	RP	AS	RP	AS	RP		AS	
							Mean	STD	Mean	STD
Pharmaceuticals										
Antibiotics before prostate biopsy	0 (0.00)	33 (0.31)	0.00	1.12	0.00	16.20	0.00	0.00	18.17	7.30
α-blockers, 5α-reductase inhibitors (BPH)	0 (0.00)	53 (0.50)	0.00	5.93	0.00	46.95	0.00	0.00	278.13	268.02
PDE5 inhibitors (ED)	2 (0.01)	0 (0.00)	1.00	0.00	23.21	0.00	23.21	32.82	0.00	0.00
Assistive technologies										
Vacuum pump, rings (ED)	10 (0.05)	2 (0.02)	1.00	1.00	205.37	197.89	205.37	26.22	197.89	217.95
Incontinence aids (IC)	44 (0.21)	4 (0.04)	6.11	8.25	38.90	29.40	237.87	563.59	242.53	338.36
Physiotherapy										
Physiotherapy pelvic floor (IC)	21 (0.10)	4 (0.04)	22.52	17.00	13.90	14.40	313.14	510.77	244.80	326.47

Table A-1: Probability of utilization, resource use, unit costs (€), and per capita costs of treated individuals (€), by treatment – unadjusted (AOK) (continued)

	P utilizatior		Costs p.p. (€)							
	RP	AS	RP				AS			
			Mean	STD	Median	Interquartile range	Mean	STD	Median	Interquartile range
ED	33 (0.15)	7 (0.07)	39.17	105.43	0.00	0.00	19.69	86.53	0.00	0.00
IC	95 (0.44)	10 (0.09)	171.41	380.29	0.00	229.10	39.46	212.91	0.00	0.00
RP w/o complications	169 (0.79)	4 (0.04)	5,067.31	2,741.14	6,335.41	1,424.59	235.32	1,207.18	0.00	0.00
AS	0 (0.00)	107 (1.00)	0.00	0.00	0.00	0.00	88.04	40.28	81.82	26.40
EBRT	0 (0.00)	11 (0.10)	0.00	0.00	0.00	0.00	258.01	828.09	0.00	0.00
Urosepsis	5 (0.02)	2 (0.02)	47.09	317.31	0.00	0.00	37.43	281.50	0.00	0.00
TURP	0 (0.00)	10 (0.09)	0.00	0.00	0.00	0.00	685.39	3,196.79	0.00	0.00
Surgery IC	2 (0.01)	1 (0.01)	54.82	615.69	0.00	0.00	109.65	1,134.19	0.00	0.00
Surgery ED	1 (0.01)	0 (0.00)	35.45	518.58	0.00	0.00	0.00	0.00	0.00	0.00
RP with complications	41 (0.19)	0 (0.00)	1,942.88	4,434.76	0.00	0.00	0.00	0.00	0.00	0.00
BPH	0 (0.00)	53 (0.50)	0.00	0.00	0.00	0.00	137.77	234.01	0.00	196.96
Stricture	10 (0.05)	1 (0.01)	220.92	1,469.95	0.00	0.00	45.08	466.35	0.00	0.00
Monitoring RP	154 (0.72)	4 (0.04)	132.86	101.27	153.08	224.47	2.01	11.42	0.00	0.00
Total	214 (1.00)	107 (1.00)	7,711.90	3,511.60	6,989.97	1,770.91	1,657.85	4,141.56	250.25	702.76

Table A-2: Probability of utilization and per capita costs (€), by cost group – unadjusted (AOK)

Cost group	P utilization		Costs p.p. (€)							
	RP	AS	RP				AS			
			Mean	STD	Median	Interquartile range	Mean	STD	Median	Interquartile range
ED	235 (0.73)	37 (0.12)	889	884	189	1,890	71	263	0	0
IC	139 (0.43)	22 (0.07)	240	517	0	196	26	141	0	0
RP w/o complications	319 (0.99)	49 (0.15)	6,843	543	6,886	0	1,051	2,480	0	0
AS	0 (0.00)	319 (0.99)	0	0	0	0	416	111	463	0
EBRT	0 (0.00)	0 (0.00)	0	0	0	0	0	0	0	0
Urosepsis	0 (0.00)	10 (0.03)	0	0	0	0	96	535	0	0
TURP	0 (0.00)	0 (0.00)	0	0	0	0	0	0	0	0
Surgery IC	1 (0.01)	2 (0.01)	31	562	0	0	63	794	0	0
Surgery ED	17 (0.05)	0 (0.00)	448	1,896	0	0	0	0	0	0
RP with complications	12 (0.04)	2 (0.01)	357	1,816	0	0	60	753	0	0
BPH	0 (0.00)	107 (0.33)	0	0	0	0	316	475	0	1,080
Stricture	14 (0.04)	0 (0.00)	88	411	0	0	0	0	0	0
Monitoring RP	319 (0.99)	49 (0.15)	1,073	85	1,080	0	165	389	0	0
Total	319 (0.99)	319 (0.99)	9,712	2,585	9,856	1,505	2,220	3,354	463	1,080

Table A-3: Probability of utilization and per capita costs (€), by cost group – microsimulation (PCa-model)

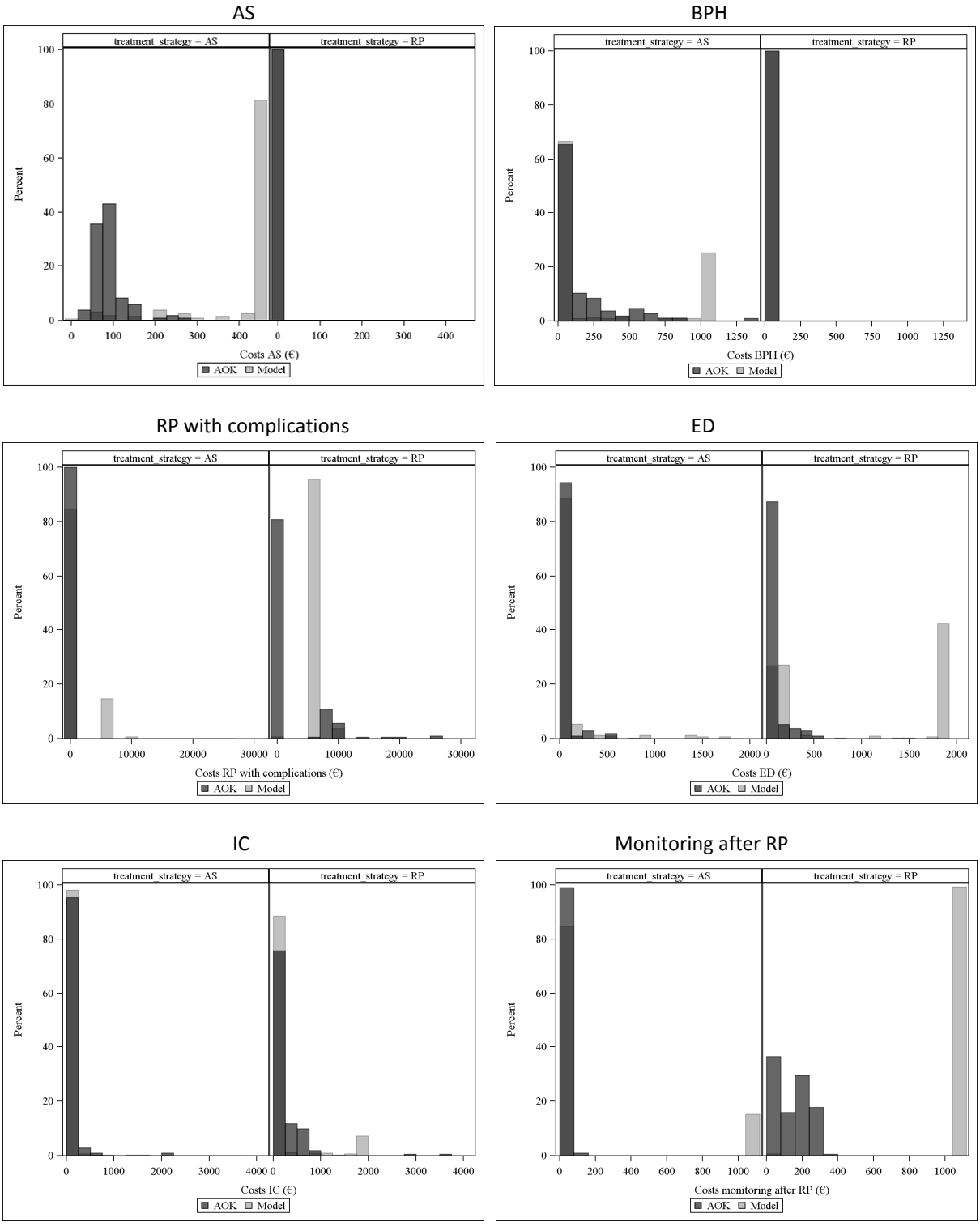
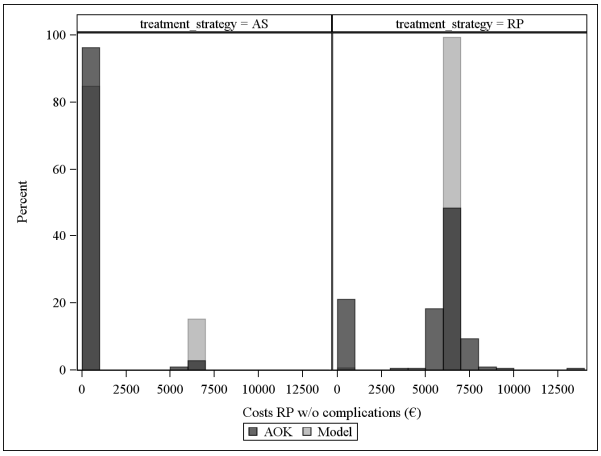
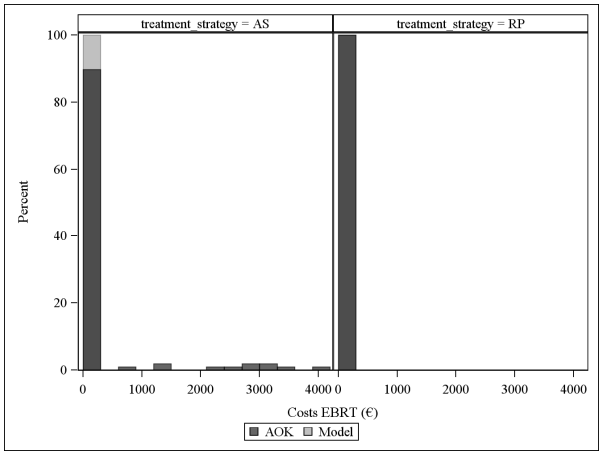


Figure A-1: Distribution of disease-related costs in cost groups (€), by treatment strategy and data source

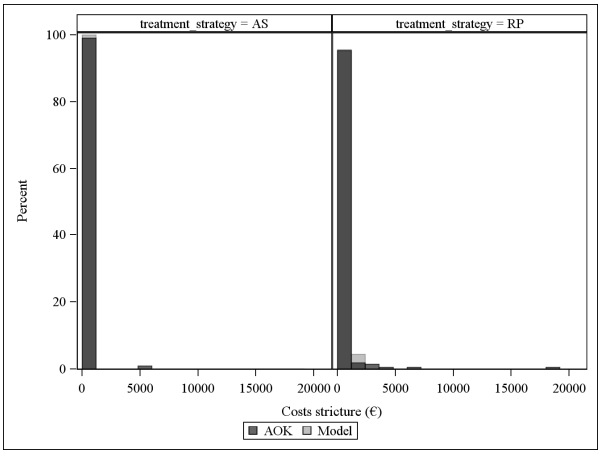
RP w/o complications



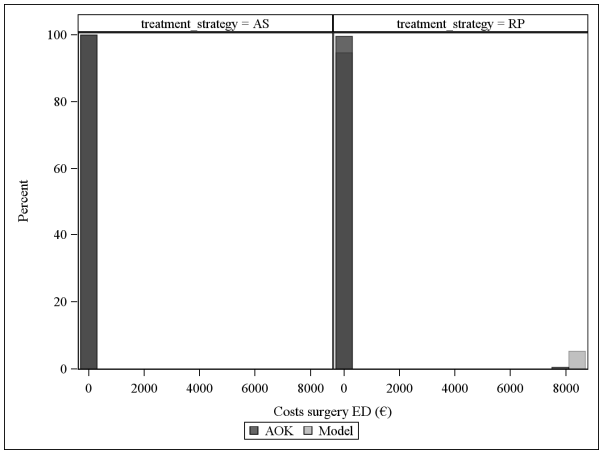
EBRT



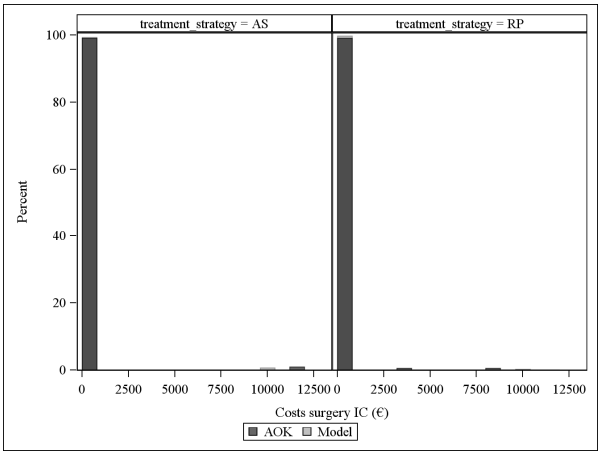
Stricture



Surgery for ED



Surgery for IC



Urosepsis

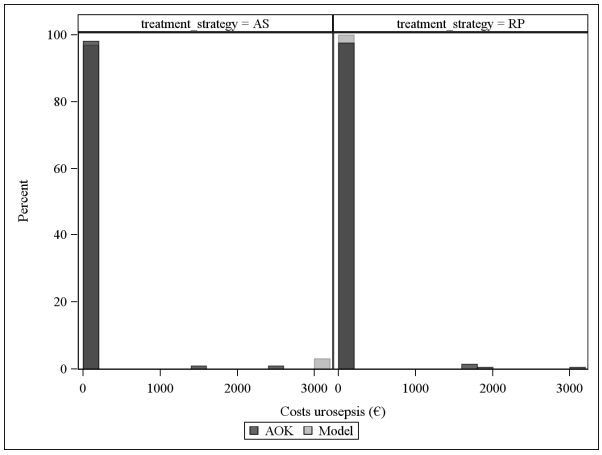


Figure A-1: Distribution of disease-related costs in cost groups (€), by treatment strategy and data source (continued)

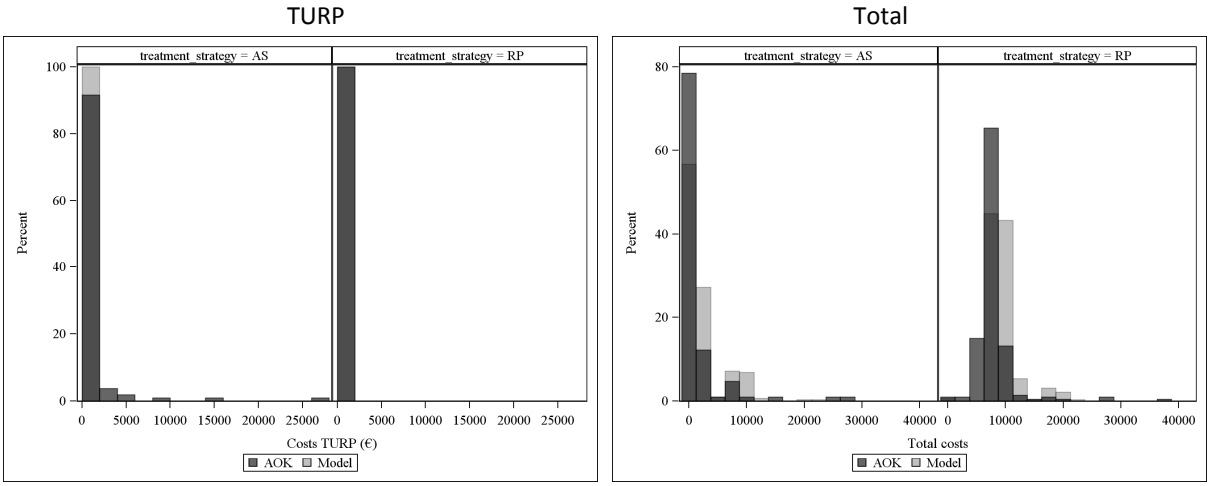


Figure A-1: Distribution of disease-related costs in cost groups (€), by treatment strategy and data source (continued)

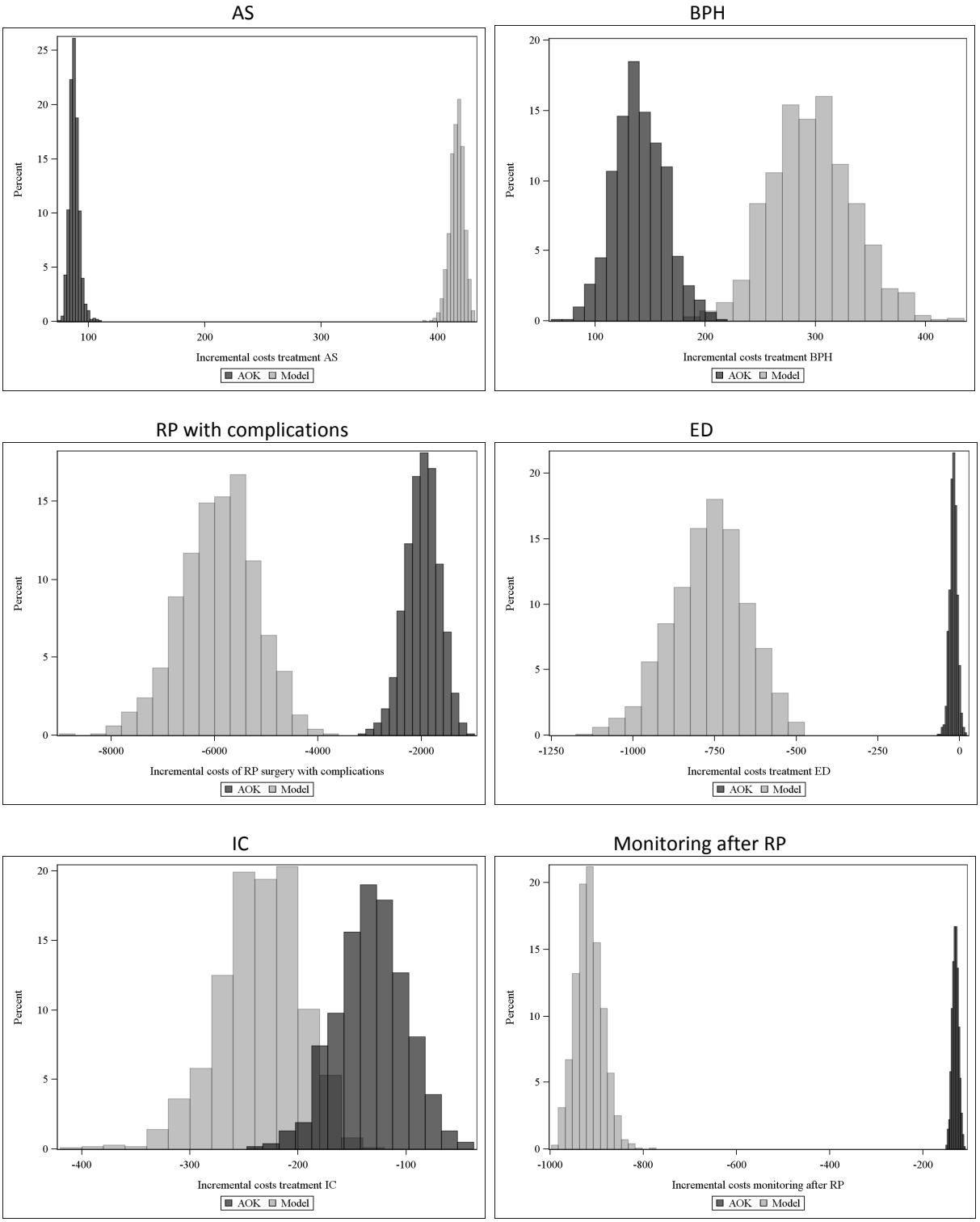


Figure A-2: Histogram of simulated, disease-related incremental costs (€), by cost group and data source

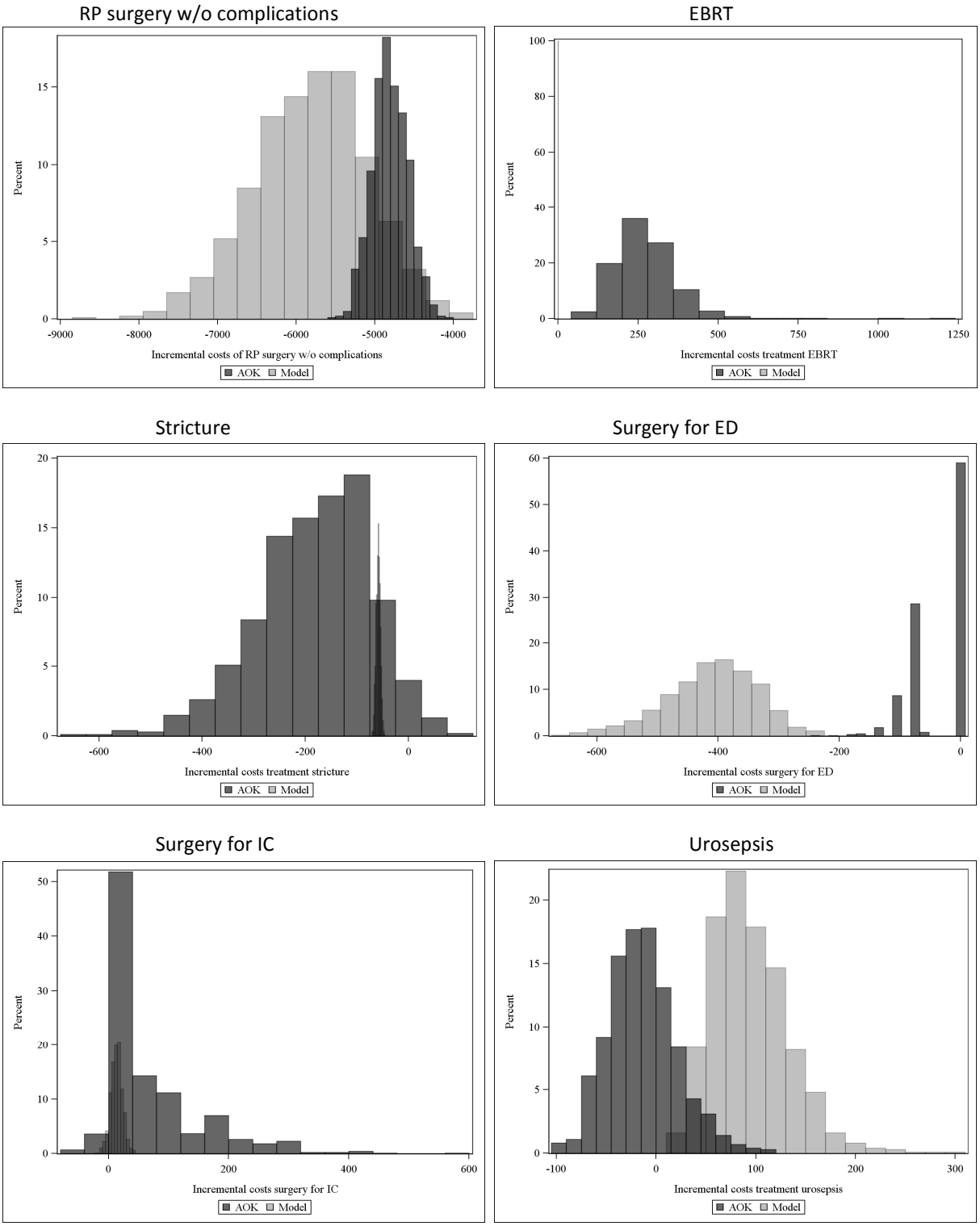


Figure A-2: Histogram of simulated, disease-related incremental costs (€), by data source (continued)

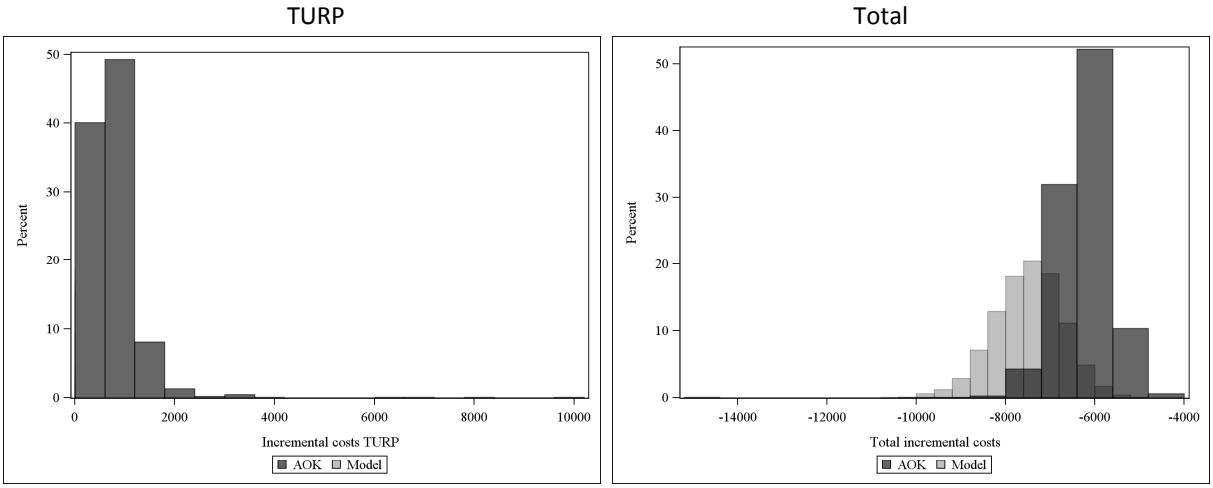


Figure A-2: Histogram of simulated, disease-related incremental costs (€), by data source (continued)

B Ethics committee statement

Brandes, Alina

Von: Fricke-Mathias, S. <S.Fricke-Mathias@blaek.de>
Gesendet: Freitag, 22. November 2013 09:23
An: Brandes, Alina
Cc: Herrmann, R.
Betreff: AW: Anfrage Ethikvotum GKV-Daten/ EK-Nr. 2013-108

Sehr geehrte Frau Brandes,

m.E. sind somit die Daten für Sie als auswertende Stelle anonymisiert und es besteht keine Beratungspflicht gemäß
§ 15 der Berufsordnung für die Ärzte Bayerns.

Freundliche Grüße

Sanja Fricke-Mathias
Apothekerin in der Geschäftsführung

Ethik-Kommission

Bayerische Landesärztekammer
Mühlbaurstraße 16
81677 München
Telefon: + 49 89 4147-165
Fax: + 49 89 4147-334
E-Mail: s.fricke-mathias@blaek.de
Internet: http://ethikkommission.blaek.de/

Von: Brandes, Alina [mailto:alina.brandes@helmholtz-muenchen.de]
Gesendet: Donnerstag, 21. November 2013 12:46
An: Ethikkommission
Betreff: Anfrage Ethikvotum GKV-Daten

Sehr geehrte Damen und Herren,

wir planen am Helmholtz Zentrum München eine Studie, welche die Auswertung von Daten einer gesetzlichen Krankenkasse beinhaltet. In den Daten werden **keine** Variablen vorhanden sein, die die Identifizierung eines bestimmten Menschen zulassen (Name, Adresse inkl. PLZ, Geburtsdatum, Institutionenkennzeichen der Leistungserbringer werden nicht übermittelt). Des Weiteren verbleiben Schlüssel zur Identifikation eines Versicherten beim Dateneigner und es besteht keine Möglichkeit für uns, diese Schlüssel zu erhalten.

Ist in diesem Fall eine Prüfung des Forschungsvorhabens durch die Ethikkommission nötig?

Freundliche Grüße,
Alina Brandes

Alina Brandes, MPH
Wissenschaftliche Mitarbeiterin

Acknowledgement

I wish to thank PD Dr. Wolf Rogowski for giving me the opportunity to participate in the project which was the basis for the work presented in this thesis. Furthermore, I am grateful to him for providing the inspiring work environment at IGM that made the completion of this thesis possible in the first place. Dr. Rogowski always supported my research endeavors and enriched the work on this thesis with numerous ideas.
I also would like to give my thanks to Prof. Rolf Holle for advice and motivation during the preparation of the thesis.

I am grateful for the methodological input that Dr. Larissa Schwarzkopf, Dr. Matthias Hunger, and PD Dr. Björn Stollenwerk provided. The clinical expertise of PD Dr. Raphaela Waidelich was very valuable for the study design and interpretation of the results. My thanks also go to Dr. Olaf Schneider for supporting me in the selection and understanding of the claims data set provided by AOK Baden-Württemberg.

I especially thank Franziska Severin, Margarethe Wacker, and Florian Koerber for their emotional support and friendship.

Eidesstattliche Versicherung

Brandes, Alina Christa Annemarie

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema
'External validation of decision-analytic models based on claims data of health insurance funds'
selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift Doktorandin/Doktorand