ESSAYS ON HEALTH ECONOMICS:

Optimal Incentives for Physicians' Diagnostic and Treatment Choice and

Empirical Evidence on Physicians' Provision of Advanced Diagnostic Tests

UNIVERSITÄT HAMBURG

Fakultät für Wirtschafts- und Sozialwissenschaften

DISSERTATION

Zur Erlangung der Würde einer Doktorin der

Wirtschafts- und Sozialwissenschaften

„Dr. rer. pol."

(gemäß der PromO vom 18. Januar 2017)

vorgelegt von

Katharina Beenk

aus Hamburg

Hamburg, den 23. Januar 2022

Vorsitzender:	Prof. Dr. Jan Marcus
Erstgutachter:	Prof. Dr. Mathias Kifmann
Zweitgutachter:	Prof. Dr. Tom Stargardt
Datum der Disputation:	17. November 2021

Contents

Chapter 1

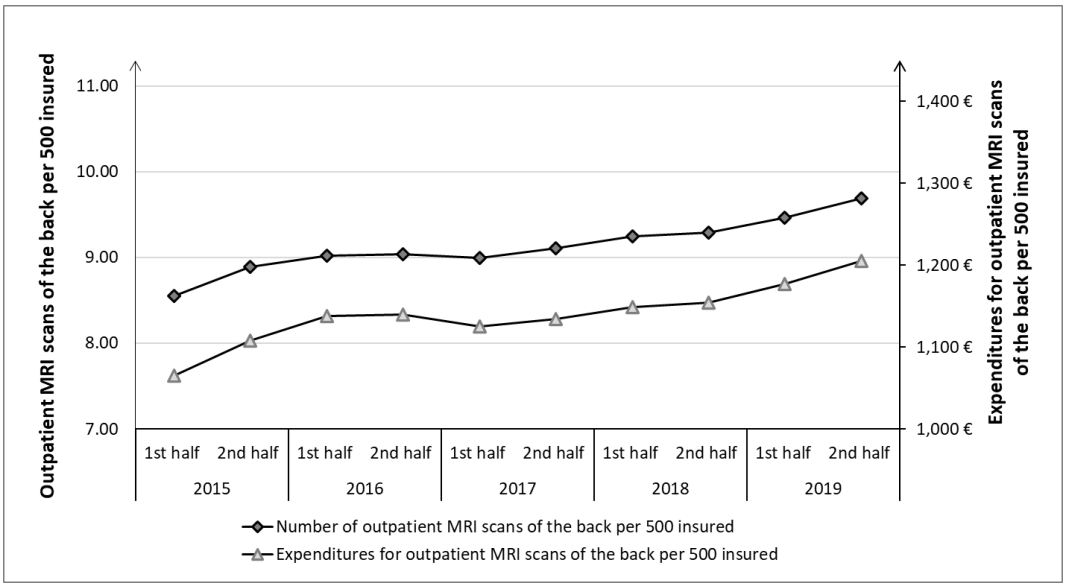
Introduction

Over the past decade, health expenditures per capita and the percentage of gross domestic product (GDP) spent on health have increased continuously in OECD countries (OECD 2021b). In Germany, for instance, expenditures per capita increased from 3,694 € in 2011 up to 4,944 € in 2019 (+33.8 %) and the percentage of GDP spent on health rose from 11.0 % to 11.9 % (+8.2 %) (Destatis 2021). On OECD average, both indicators are projected to increase even further until 2030 (OECD 2019). This places a growing burden on social systems, private households and employers because rising tax rates and contributions serve as sources for health care financing.

In the literature, three factors are identified as drivers of the growth in health expenditures: the continued aging of the population, rising incomes and technological innovations (OECD 2015). Firstly, the continued aging of the population steadily contributes to future health spending. Although the relation between growing life expectancy and health expectancy is controversially discussed in the literature, evidence suggests that there is a rising demand for health care services that improve life for the elderly or extend life (Breyer and Felder 2006). Secondly, rising incomes may contribute to the increase in expenditures as patients' expectations regarding the amount and the quality of care are found to grow in income (Chernew and Newhouse 2011). Thirdly, technological innovations are considered the main cost driver. Even though new technologies may also have a positive effect on expenditures as they potentially allow for cost savings, the negative effects are found to dominate. Essentially, technological innovations lead to a growing number of advanced diagnostic and treatment options. The extended range of physician services, the associated increase in actual utilization of these services as well as the increased prices for the innovative procedures will continue to drive up expenditures for physician services in the future (Chernew and Newhouse 2011; Willemé and Dumont 2015).

There is strong evidence that especially the utilization of (advanced) diagnostic tests has worryingly increased in recent years and, therefore, has contributed to the rise in health expenditures (Brownlee et al. 2017). In particular, magnetic resonance imaging (MRI) has become a widely employed diagnostic device in numerous OECD countries, whereby more than half of the MRI scans are undertaken in the outpatient sector (OECD 2021a). In Germany, for example, the number of MRI scans of the back is continuously rising.

Figure 1-1: Trends in the Provision of Outpatient MRI Scans of the Back in Germany



Note: The figure depicts the trends in the provision of outpatient MRI scans of the back for persons who are insured at Techniker Krankenkasse, the largest statutory health insurance in Germany. For each half-year, MRI scans were identified using doctor's fee schedule position "GOP 34411 – MRI scan of parts of the spinal column". Expenditures were calculated based on the respective price of the service in the quarter of service provision, as given by the doctor's fee schedule (Federal Association of Statutory Health Insurance Physicians 2021). Prices were further adjusted for inflation and expressed in 2015 Euros. The number of MRI scans and the expenditures for the MRI scans in each half-year are stated per 500 insured as of the preceding reporting date December 31st and June 30th, respectively. Data: Techniker Krankenkasse

Figure 1-1 depicts the trends in the number of outpatient MRI scans of the back provided per 500 insured and in the expenditures for these scans per 500 insured for the years 2015 through 2019. Inspecting the graphs reveals a clear upwards trend in both indicators: The number of MRI scans of the back increased by 13.3 % from 8.6 scans per 500 insured in the first half-year of 2015 up to 9.7 scans per 500 insured in the second half-year of 2019. Expenditures for the scans rose accordingly by 13.2 %. Over the same time period, expenditures for overall outpatient services only increased by 10 % per 500 insured. This lets the rise in the expenditures for MRI scans of the back stand out (Techniker Krankenkasse 2017, 2020).[1]

[1] It could also be argued that the increase in outpatient MRI scans of the back may be caused by (medically-justified) factors such as a simple shift of services from the inpatient to the outpatient sector, the aging of the insured population or changes in the recommendations from professional societies. However, analyzing the trend in the provision of inpatient MRI scans of the back over time refutes the first argument because the number of inpatient MRI scans of the back per 500 insured also increased by 10.5 %. Based on data from Techniker Krankenkasse, inpatient MRI scans of the back were identified using Operation and Procedure Codes (OPS) 3802 und 3823 (German Institute of Medical Documentation and Information 2020). Moreover, the insured population has not aged substantially during the observational period. In the first half-year of 2015 insured were on average 39.8 years old, while in the second half-year of 2019 they were on average 40.5 years old. Recommendations for the provision of MRI scans of the back from professional societies did not significantly change over time (AWMF 2017; Chonet and Becker 2017). Another, rather unlikely explanation might be an increase in average morbidity of the insured population, which does not depend on age, or a continuing reduction of excess demand.

Although MRI scans often provide superior diagnostics compared to alternative procedures, including conventional imaging, the provision is also much more expensive (Federal Association of Statutory Health Insurance Physicians 2021). Furthermore, it needs to be considered to what extent the utilization of expensive MRI scans actually contributes to improved treatment choice. Literature suggests that MRI scans, for instance, for the diagnosis of breast cancer or for management of lower back pain, may have controversial downstream effects on medical cost and patient outcomes. The scans may cause additional tests and referrals, result in costly invasive procedures of questionable value that tend to be performed more frequently in geographic areas with higher MRI scan rates, and, in the worst case, lead to adverse effects on patient survival (Baras and Baker 2009; Bedrosian et al. 2016; Chou et al. 2011; Lurie et al. 2003; Onega et al. 2018; Padia et al. 2016). Moreover, international evidence proposes that there are gaps between actual physician practice and evidence-based clinical practice guidelines, which recommend a prudent use of MRI scans (Chou et al. 2011; Ghamat et al. 2017). Therefore, regulators are concerned that the observed increase in MRI scans may result from inefficient overuse (Baxi et al. 2017; Foster et al. 2018). Overuse of medical services should generally be avoided as finite health care resources should not be wasted and diverted from the budget for beneficial services. Consequently, especially against the background of the continuously rising numbers of MRI scans across OECD countries (OECD 2021a), the cost-effective provision of expensive (advanced) diagnostic tests is of major importance for the future.

While there may be several reasons for the extensive use of (advanced) diagnostic tests, including physicians' concern about litigation and poor diagnostic ability, peer effects and patients' expectations, the remuneration scheme plays a special role (Baras and Baker 2009; Baicker et al. 2007; Chandra and Staiger 2007; Doyle et al. 2010; Little et al. 2004). Health care markets are characterized by various inefficiencies, which may lead to suboptimal market outcomes (Dranove and Satterthwaite 2000). Most notably, asymmetric information between third-party payers, providers and patients creates moral hazard problems: Depending on the payment contract they face, physicians, for instance, may not place enough weight on the patients' health-benefit but rather act on their own monetary self-interest (Arrow 1963). This may result in inefficient overprovision of services. In the principal-agent framework, the incentives that are placed on physicians by a payment system determine their actions (Ellis and McGuire 1986, 1990; McClellan 2011; McGuire 2000).

In conclusion, in health care systems with limited resources, regulators do not only have to determine the optimal amount of services to purchase but they also have to elaborate how to

incentivize physicians to provide the optimal amount (Jack 2005). Due to the increasing complexity of international health care markets, which are formed by inefficiencies, various actors (i.a. payer, providers and patients) with differing objective functions and an extending range of diagnostic and treatment options, continuous research is necessary to inform regulators on these matters.

This dissertation consists of three studies that examine the provision of (advanced) diagnostic tests by means of microeconomic models and microeconometric analyses. The focus lies on the relation between a public insurer (the payer) and outpatient physicians. Physicians represent the center of health care systems because their decisions essentially determine health care cost and quality of care (see McGuire 2000 for an overview). In particular, this dissertation has two main objectives: Firstly, it examines optimal financial incentives for physicians' diagnostic (testing) and treatment choice. Secondly, it presents empirical evidence on the actual provision of advanced diagnostic tests, in particular MRI scans of the breast, and the effect on direct medical cost and patient outcomes.

The first study is presented in Chapter 3 and examines optimal financial incentives for physician's sequential diagnostic testing and treatment choice. Based on a principal-agent model, the study analyzes the use of expensive advanced diagnostic tests, such as MRI scans of the back, to improve cost-effective treatment choice. A first-best analysis reveals that it can be optimal to use an advanced diagnostic test only conditional on the outcome of a less expensive, yet imperfect, pretest. Then, contracts, which implement optimal sequential diagnostic testing and treatment choice given a situation with double moral hazard from hidden action and hidden information, are determined. Results show that a mixed payment contract, consisting of a capitation payment, conditional fee-for-service (FFS) payments and a fee for the utilization of the advanced diagnostic test, provides optimal incentives for the physician. Yet, implementing the optimal diagnosis and treatment course can create information rents for the physicians, which may yield alternative diagnosis and treatment courses superior and eventually leads to a rent-efficiency trade-off. By contrast, a less restrictive contract, comprising only a capitation payment and unconditional FFS payments, as common in numerous OECD countries, is not incentive compatible. This important finding indicates that existing payment contracts do not seem to put sufficient responsibility for ordering expensive diagnostic tests on the attending physician.

Chapter 4 presents the second study. The study investigates the provision of MRI scans of the female breast in Germany. The Federal Joint Committee (G-BA) enacts under which conditions

outpatient MRI scans of the breast for breast cancer treatment are reimbursed by the social health insurance and how providers are paid for the service - which is FFS in this case (Federal Association of Statutory Health Insurance Physicians 2021; Federal Joint Committee 2020). However, considering the prevailing non-RCT standard evidence on the effect of these MRI scans and the observation that, i.a. due to asymmetric information, payment schemes may provide suboptimal incentives, providers may disregard the reimbursement conditions. In a first-step, using administrative data from Techniker Krankenkasse, the study investigates whether outpatient MRI scans of the female breast are provided non-adherent to the reimbursement conditions. In a second step, it examines the effect of the non-adherent provision on direct medical cost and patient outcomes. To allow for causal interpretation, propensity score matching and a difference-in-difference approach are applied. Results suggest that more than one third of all breast MRI scans that were performed in 2016 may be classified as non-adherent to reimbursement conditions. Women who received a non-adherent MRI scan on average cause significantly higher direct medical cost in the follow-up period than women in a control group, yet, they also benefit from increased overall survival. Hence, considering cost and outcomes, it is not straightforward whether the overprovision of MRI scans is indeed ineffective and should be targeted by regulators. In brief, the study concludes that further research, essentially based on data from complementary sources such as clinical data and cause of death statistics, should be conducted to support regulators in making informed decisions on the enforcement of the reimbursement conditions.

The third study is presented in Chapter 5 and examines payment contracts to promote optimal diagnostic and treatment choice among physicians with heterogeneous diagnostic ability. In the principal-agent framework, heterogeneities in diagnostic ability pose a special challenge to regulators. Considering a treatment choice model with endogenous diagnostic effort, the study examines how a public insurer should optimally contract with providers with heterogeneous ability. Hence, a situation with simultaneous double moral hazard from hidden action and hidden information as well as adverse selection from hidden knowledge is assumed. A first-best analysis reveals that it may turn out suboptimal to incentivize all ability-types to exert costly diagnostic effort. While single policy mixed payment contracts, consisting of a capitation payment and conditional FFS payments, are found to promote optimal incentives, they also entail costly information rent payments to the physicians. Therefore, in the second-best situation, particularly depending on the share of the high-ability types in the market, the parameter constellations for which all ability-types should be incentivized are further restricted. In contrast to standard principal-agent theory, a menu of policies that aims at inducing self-

selection among the physicians does not improve on this outcome. Although medical expert associations frequently campaign for the special importance of a correct diagnosis to effective health care and technological innovations allow for an increasing number of diagnostic tests, considering the social cost of care provision (given heterogeneous physician ability) may paint a more differentiated picture and yield the provision of these tests suboptimal.

The remainder of this dissertation is organized as follows: Chapter 2 provides a brief literature review. Considering the special features of the health care market, the current state of evidence on how to purchase diagnostic and treatment services and on whether expensive advanced diagnostic tests, in particular MRI scans, are provided in adherence to guideline-based reimbursement conditions as well as its impact on cost and outcomes is discussed. Moreover, it is outlined how the studies in this dissertation contribute to the existing literature. Afterwards, Chapter 3 to Chapter 5 present the single studies.

References

Arrow, Kenneth J. (1963). Uncertainty and the Welfare Economics of Medical Care. The American Economic Review 53 (5), pp. 941–973.

AWMF (2017). Nationale VersorgungsLeitlinie Nicht-spezifischer Kreuzschmerz. 2. Auflage. Available online at https://www.leitlinien.de/mdb/downloads/nvl/kreuzschmerz/kreuzschmerz-2aufl-vers1-kurz.pdf, checked on 4/10/2021.

Baicker, Katherine; Fisher, Elliott S.; Chandra, Amitabh (2007). Malpractice Liability Costs and The Practice of Medicine in the Medicare Program. Health Affairs (Millwood) 26 (3), pp. 841–852. DOI: 10.1377/hlthaff.26.3.841.

Baras, Jacqueline D.; Baker, Laurence C. (2009). Magnetic resonance imaging and low back pain care for Medicare patients. Health Affairs (Project Hope) 28 (6), w1133-40. DOI: 10.1377/hlthaff.28.6.w1133.

Baxi, Shrujal S.; Kale, Minal; Keyhani, Salomeh; Roman, Benjamin R.; Yang, Annie; Derosa, Antonio P.; Korenstein, Deborah (2017). Overuse of Health Care Services in the Management of Cancer: A Systematic Review. Medical Care 55 (7), pp. 723–733. DOI: 10.1097/MLR.0000000000000734.

Bedrosian, Isabelle; Xing, Yan; Abdel Rahman, Shereen; Allen, Lisa; Le-Petross, Huong; Whitman, Gary J. et al. (2016). A Cost Analysis of Preoperative Breast MRI Use for Patients with Invasive Lobular Cancer. Annals of Surgical Oncology 23 (1), pp. 23–29. DOI: 10.1245/s10434-015-4710-1.

Breyer, Friedrich; Felder, Stefan (2006). Life expectancy and health care expenditures: a new calculation for Germany using the costs of dying. Health policy (Amsterdam, Netherlands) 75 (2), pp. 178–186. DOI: 10.1016/j.healthpol.2005.03.011.

Brownlee, Shannon; Chalkidou, Kalipso; Doust, Jenny; Elshaug, Adam G.; Glasziou, Paul; Heath, Iona et al. (2017). Evidence for overuse of medical services around the world. The Lancet 390 (10090), pp. 156–168. DOI: 10.1016/S0140-6736(16)32585-5.

Chandra, Amitabh; Staiger, Douglas O. (2007). Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. Journal of Political Economy 115, pp. 103–140. DOI: 10.1086/512249.

Chernew, Michael E.; Newhouse, Joseph P. (2011). Health Care Spending Growth. In : Handbook of Health Economics, Volume 2: Elsevier, pp. 1–43. DOI: 10.1016/B978-0-444-53592-4.00001-3.

Chonet, Jean-Francois; Becker, Annette (2017). Update der Nationalen Versorgungsleitlinie Kreuzschmerz. Zeitschrift für Allgemeinmedizin 93 (6). DOI: 10.3238/zfa.2017.0250-0254.

Chou, Roger; Qaseem, Amir; Owens, Douglas K.; Shekelle, Paul (2011). Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. Annals of internal medicine 154 (3), pp. 181–189. DOI: 10.7326/0003-4819-154-3-201102010-00008.

Destatis (2021). Gesundheitsausgaben. Available online at https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Gesundheitsausgaben/_inhalt.html#sprg229148, checked on 4/10/2021.

Doyle, Joseph J.; Ewer, Steven M.; Wagner, Todd H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. Journal of Health Economics 29, pp. 866–882. DOI: 10.1016/j.jhealeco.2010.08.004.

Dranove, David; Satterthwaite, Mark A. (2000). The industrial organization of health care markets. In : Handbook of Health Economics, Volume 1, Part B: Elsevier, pp. 1039–1139. DOI: 10.1016/S1574-0064(00)80033-5.

Ellis, Randall P.; McGuire, Thomas G. (1986). Provider Behaviour under Prospective Reimbursement: Cost-sharing and Supply. Journal of Health Economics 5 (2), pp. 129–151. DOI: 10.1016/0167-6296(86)90002-0.

Ellis, Randall P.; McGuire, Thomas G. (1990). Optimal Payment Systems for Health Services. Journal of Health Economics 9 (4), pp. 375–396. DOI: 10.1016/0167-6296(90)90001-J.

Federal Association of Statutory Health Insurance Physicians (2021). Online Version of the EBM 2021. Available online at https://www.kbv.de/html/online-ebm.php, checked on 4/10/2021.

Federal Joint Committee (2020). Richtlinie des Gemeinsamen Bundesausschusses zu Untersuchungs- und Behandlungsmethoden der vertragsärztlichen Versorgung. Berlin.

Available online at https://www.g-ba.de/downloads/62-492-2211/MVV-RL_2020-06-18_iK-2020-07-22.pdf, checked on 9/24/2020.

Foster, Nadine E.; Anema, Johannes R.; Cherkin, Dan; Chou, Roger; Cohen, Steven P.; Gross, Douglas P. et al. (2018). Prevention and treatment of low back pain: evidence, challenges, and promising directions. The Lancet 391 (10137), pp. 2368–2383. DOI: 10.1016/S0140-6736(18)30489-6.

German Institute of Medical Documentation and Information (2020). OPS Version 2020. Cologne, checked on 4/10/2021.

Ghamat, Salar; Zaric, Gregory S.; Pun, Hubert (2017). Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment. Production and Operations Management 63 (5), p. 1606. DOI: 10.1111/poms.12780.

Jack, William (2005). Purchasing health care services from providers with unknown altruism. Journal of Health Economics 24 (1), pp. 73–93. DOI: 10.1016/j.jhealeco.2004.06.001.

Little, Paul; Dorward, Martina; Warner, Greg; Stephens, Katharine; Senior, Jane; Moore, Michael (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. BMJ (Clinical research ed.) 328 (7437), p. 444. DOI: 10.1136/bmj.38013.644086.7C.

Lurie, Jon D.; Birkmeyer, Nancy J.; Weinstein, James N. (2003). Rates of advanced spinal imaging and spine surgery. Spine 28 (6), pp. 616–620. DOI: 10.1097/01.BRS.0000049927.37696.DC.

McClellan, Mark (2011). Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality? Journal of Economic Perspectives 25 (2), pp. 69–92. DOI: 10.1257/jep.25.2.69.

McGuire, Thomas G. (2000). Physician Agency. In : Handbook of Health Economics, Volume 1, Part A: Elsevier, pp. 461–536. DOI: 10.1016/S1574-0064(00)80168-7.

OECD (2015). Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives. Paris: OECD Publishing. DOI: 10.1787/9789264233386-en.

OECD (2019). Health at a Glance 2019: OECD Indicators. Paris: OECD Publishing. DOI: 10.1787/4dd50c09-en.

OECD (2021a). Magnetic resonance imaging (MRI) exams. Available online at https://www.oecd-ilibrary.org/social-issues-migration-health/magnetic-resonance-imaging-mri-exams/indicator/english_1d89353f-en, checked on 4/24/2021.

OECD (2021b). OECD Health Statistics. Health expenditure and financing. Available online at https://stats.oecd.org/Index.aspx?DataSetCode=SHA, checked on 4/10/2021.

Onega, Tracy L.; Zhu, W.; Weiss, J. E.; Goodrich, M.; Tosteson, A. N. A.; DeMartini, W. et al. (2018). Preoperative breast MRI and mortality in older women with breast cancer. Breast Cancer Research and Treatment 170 (1), pp. 149–157. DOI: 10.1007/s10549-018-4732-z.

Padia, Shilpa A.; Freyvogel, Mary; Dietz, Jill; Valente, Stephanie; O'Rourke, Colin; Grobmyer, Stephen R. (2016). False-positive Extra-Mammary Findings in Breast MRI: Another Cause for Concern. The Breast Journal 22 (1), pp. 90–95. DOI: 10.1111/tbj.12524.

Techniker Krankenkasse (2017). Geschäftsbericht 2016. Available online at https://docplayer.org/52501979-2-geschaeftsbericht-2016.html, checked on 4/11/2021.

Techniker Krankenkasse (2020). Geschäftsbericht 2019. Available online at https://gb2019.tk.de/app/uploads/2020/06/TK_Gescha%CC%88ftsbericht_2019.pdf, checked on 4/11/1021.

Willemé, Peter; Dumont, Michel (2015). Machines that Go 'Ping': Medical Technology and Health Expenditures in OECD Countries. Health economics 24 (8), pp. 1027–1041. DOI: 10.1002/hec.3089.

Chapter 2

Literature Review and Contribution

In view of the continuously rising number of MRI scans that are provided across OECD countries (OECD 2021) and the associated increase in expenditures, this dissertation generates new insights on the provision of expensive (advanced) diagnostic tests and their contribution to improved treatment choice. With the help of microeconomic models, the dissertation examines optimal financial incentives for physicians' diagnostic (testing) and treatment choice and highlights the challenges, which regulators face in creating incentive-compatible payment schemes. Furthermore, it presents empirical evidence on whether MRI scans of the female breast for breast cancer treatment are provided in adherence to guideline-based reimbursement conditions in Germany and analyzes the effects on direct medical cost and patient outcomes. Therefore, the dissertation addresses two fundamental questions that arise for regulators in health care systems with limited resources (Jack 2005):

1. *How much health care to purchase?* Given individual patient characteristics and/or given a certain medical indication, regulators need to evaluate, which health care services should be covered and reimbursed by compulsory insurance schemes or national health services.
2. *How to purchase health care services?* Given an optimal amount of health care services, regulators need to examine how to incentivize health care providers to supply this particular amount of services.

The first question is ideally answered by means of cost-benefit-analyses (CBA) or cost-effectiveness analyses (CEA), whereby the latter are more popular in health economics. The analyses are based on patient outcomes from clinical trials and on (administrative) cost data for, inter alia, in- and outpatient services or pharmaceuticals. However, evidence from clinical trials may sometimes be suboptimal due to limited observation periods, ethical concerns or weaknesses in the study design such as small sample sizes or non-randomized controlled trial (non-RCT) standard. Moreover, the analyses always rely on value judgements. While CBA presuppose an explicit monetary valuation of health benefits, for CEA regulators (and society in general) still need to disclose their willingness to pay for patient outcomes in certain therapeutical areas by defining cut-off values (see Garber 2000 for an overview). This may not be politically intended in all health care systems. Therefore, coverage decisions also rely on

political prioritization (Niessen et al. 2000). Finally, recommendations from clinical practice guidelines, which are published by professional societies, serve as complementary decision support for regulators (Alonso-Coello et al. 2016).

In the last three decades, researchers increasingly focused on the second fundamental question, which can be answered with the help of microeconomic models. As already outlined in Chapter 1, health care markets are shaped by various inefficiencies that pose a challenge to researchers and regulators alike in creating incentive-compatible payment contracts. In health economics and in economics in general, principal-agent models allow for studying incentive mechanisms, which deal with inefficiencies caused by information asymmetries (Holmström 2017). In the following, the focus lies on the contracting between a public insurer (the principal) and physicians (the agents).[1] Physicians have the unique qualification to care for patients and, therefore, have an informational advantage over the public insurer (and the patients). In case physicians are better off by taking advantage of their expert role and act on their own monetary self-interest instead of pursuing the insurer's overarching objective of optimal care provision, a moral hazard problem arises. In particular, a moral hazard problem from hidden action arises because the insurer cannot observe whether physicians, for instance, undertook optimal cost-reducing effort when caring for patients or offered the optimal amount of quality (Garcia Mariñoso and Jelovac 2003; McGuire 2000). Therefore, regulators need to create incentive schemes that align the agents' objectives with the principal's objectives: In the health care market, both cost containment and high quality care constitute the acknowledged objectives (Jelovac 2001).[2]

The health economics literature distinguishes between two categories of payment policies, which differ in the degree of supply-side cost-sharing: retrospective, cost-based policies and prospective, fixed-price policies (Ellis and McGuire 1993; Ma 1994). Fully retrospective payment policies reimburse physicians for the actual cost they incurred. Due to the obvious lack of incentives for cost containment, these policies have become rare in current practice. Fee-for-service (FFS) policies, however, under which physicians are paid a fixed fee for each diagnostic test and treatment procedure they provide, are prominent in numerous health care systems (see Mossialos et al. 2017 for an international survey). While incentives for cost containment are

[1] Generally, two strategies for controlling health expenditures are frequently discussed in the health economic literature: supply-side cost sharing and demand-side cost sharing. While the former aims at altering the incentives for health care providers, the latter focuses on the incentives that patients face, such as co-payments or deductibles (Ellis and McGuire 1993). Because the focus of this dissertation lies on the relation between a public insurer and physicians as health care providers, demand-side cost sharing will not be further discussed.

[2] Note that "high quality care" is a comprehensive term that refers to the two dimensions of care: both the optimal quantity and quality of care.

still comparatively weak, physicians are perceived to offer high quality care. By contrast, under prospective payment systems physicians receive a lump-sum payment for an entire episode of care. Capitation payments in outpatient care or diagnosis-related group (DRG)-based payments in the inpatient sector serve as common examples for prospective payment policies (Street et al. 2011). Prospective payments policies induce supply side cost-sharing because the physician must bear the marginal cost of care provision and, therefore, shift some risk to the providers. While incentives for cost containment are clearly strong, incentives for high-quality care are lacking (Ellis and McGuire 1993). There is concern that physicians may lower the quality of care to a suboptimal level to minimize their cost and, therefore, to maximize net revenue. Asymmetric information between physicians and patients regarding the quality of care provision intensifies the problem (Chalkley and Malcomson 1998). Moreover, physicians may be incentivized to screen patients for disease severity and dump perceivably high cost patients in order to further maximize net revenue (Ma 1994). In summary, both kinds of payment policies, pure retrospective and pure prospective, lead to extreme outcomes, which is usually not intended. In general, the problem is solved by implementing mixed payment schemes, which transfer part of the risk to the supply-side and trade-off the objectives of cost containment and high-quality care (Ellis and McGuire 1986). In particular, the regulator faces a multitasking principal-agent problem, in which the agents must be induced to fulfill more than one task. Thus, more than one payment instrument is needed to provide optimal incentives (Holmström and Milgrom 1991). Up until today, mixed payment schemes are preferably used by regulators. In addition, new payment schemes such as performance-based payments and bundled payments, which especially target the coordination of care across providers, have been proposed during the last decade and have become increasingly popular (Fainman and Kucukyazici 2020; McClellan 2011; Shih et al. 2015).

As outlined above, health economic theory proposes that the incentives that are placed on physicians by a payment system essentially determine their behavior. This finding is reinforced by empirical and experimental papers that study the influence of different payment schemes on physicians' supply of medical services. Empirical evidence suggests that, in the outpatient sector, FFS payments are generally associated with a higher quantity of medical services than capitation payments. In particular, the number of consultations in a given period and the number of diagnostic and curative services provided is found to be higher under FFS payments than under capitation payments (Epstein et al. 1986; Gosden et al. 2001 for a systematic review). While the increased provision of services may indeed be beneficial for patients, it may as well result in pure waste or even be harmful. A change in renumeration from salaries to FFS

payments, for instance, led dentists to provide an increased number of potentially harmful dental x-rays as diagnostic procedures (Chalkley and Listl 2018). Hence, an increase in the quantity of services under FFS payments should not be equated with a general increase in quality of care. Experimental evidence also shows that physicians provide a significantly higher quantity of services under FFS payments than under capitation payments, whereby patients tend to be overserved under the former scheme and underserved under the latter scheme (Brosig-Koch et al. 2016; Hennig-Schmidt et al. 2011). Finally, experimental and empirical studies confirm the theoretical finding that mixed payment schemes are superior to pure payment schemes with the actual market outcome being closer the optimal outcome (Brosig-Koch et al. 2017; Kralj and Kantarevic 2013).

In the following subchapters, further literature specific to the two main objectives of this dissertation is discussed and the contributions of the single studies are highlighted.

2.1 Optimal Incentives for Physicians' Diagnostic and Treatment Choice

As stated in the seminal paper by Arrow (1963), the "*special characteristics"* of the health care market have to be accounted for when studying market outcomes. Health care markets are characterized by various inefficiencies, which may lead to suboptimal market outcomes (Dranove and Satterthwaite 2000). Most notably for this dissertation, information asymmetries between third-party payers, physicians and patients pose a challenge to creating optimal financial incentives for physician's diagnostic (testing) and treatment choice.

To begin with, physicians take on a dual expert role as they diagnose the patient and subsequently provide treatment services. Due to asymmetric information, the patient cannot verify ex-post whether the quality of the treatment was appropriate. Therefore, physician services are considered credence goods (Darby and Karni 1973). Comparable to other expert markets, like car repairs, the joint provision of diagnostic and treatment services gives the expert the opportunity to defraud his customer by choosing the most profitable treatment option instead of doing what is in the best interest of the customer. Theoretical evidence on the functioning of credence goods markets in the healthcare shows that patients are served efficiently, only if several crucial assumptions hold: i.a. physicians are homogeneous, diagnostic ability is perfect, patients cannot obtain a second opinion and treatments are verifiable. Furthermore, it is assumed that physicians can freely set their prices (Dulleck and Kerschbamer 2006 for a general model). However, especially the assumption of homogeneous physicians with perfect diagnostic ability is a strong one. Theory shows that if the latter

assumption did not hold as physicians are heterogeneous in ability and ability is unobservable, inefficient market outcomes could result from collusion between experts. Low-ability physicians rely on their diagnosis too often or high-ability physicians ignore their performed diagnosis for treatment choice, leading to inefficient over- or undertreatment (Liu et al. 2020). Evidence from various field and lab experiments complements and confirms these theoretical findings (Kerschbamer and Sutter 2017 for a survey). For instance, dentists were found to significantly overtreat patients in a field experiment (Gottschalk et al. 2020). Even though the credence goods literature mainly focuses on the relation between expert physicians and patients, optimal financial incentives are not derived, physicians are assumed to set their own prices and, strictly speaking, focuses on direct monetary incentives from the different treatment options instead of incentives from moral hazard, it becomes evident that asymmetric information between physicians and patients may result in inefficient market outcomes. These inefficiencies can partly be resolved by the provision of optimal incentives for physician's diagnostic and treatment choice because patient outcomes enter the public insurer's objective function.

The focus of Chapter 3 and Chapter 5 of this dissertation particularly lies on the physician's (the agent's) response to moral hazard, which results from the informational advantage over the public insurer (the principal), while the patient is assumed to be passive. When considering physician's diagnostic and treatment choice, two types of moral hazard, together also referred to as double moral hazard, are frequently identified in the literature (see McGuire 2000 for an overview): Moral hazard from hidden action and moral hazard from hidden information. Moral hazard from hidden action refers to the physician's private information about the diagnostic (test) choice, whereby the choice may be binary (Was the test used? Yes/no) or continuous (How thoroughly did the physician perform the test?). Moral hazard from hidden information results, as consequence, from the physician's private information regarding the outcome of a diagnostic (test) procedure (Which treatment is optimal given the test's results?).

Apart from the moral hazard problems, the utilization of diagnostic tests may generally be helpful in choosing between treatment alternatives under diagnostic risk. Test accuracy is traditionally described by using sensitivity and specificity (Attia 2003). Given the prevalence of an illness, Bayesian updating enables physicians to form more accurate beliefs about patients' state of health, allowing for improved treatment choice. Since most tests are imperfect though, false positive and false negative test results may still lead to inappropriate treatment choice with potentially negative impacts on patient outcomes. Hence, diagnostic tests are only considered valuable, if they improve patient outcomes (Schünemann et al. 2008; Felder and Mayrhofer 2017). Moreover, they also impose costs on the payer. Thus, the utilization of

diagnostic tests does not necessarily improve overall welfare. In particular, there may be situations, in which it is optimal to forgo diagnostic testing and to make a blind treatment choice instead (Felder and Mayrhofer 2017). Depending on the structure of the payment system in place, additional payment instruments are needed to incentivize physicians to fulfill the additional task of optimal diagnostic (test) choice.

So far, the theoretical literature on the interaction of both diagnostic (testing) and treatment choice and the associated derivation of welfare maximizing payment contracts is relatively scarce. To begin with, Jelovac (2001) considers a treatment choice model with endogenous diagnostic effort. Physicians may exert costly effort to get a private signal on the patients' illness in a situation with double moral hazard from hidden action and hidden information. The higher the continuous effort level, the more precise is the signal on the patient's illness. Patients may suffer from a minor or a severe illness. For minorly ill patients an inexpensive treatment is adequate. With some probability, these patients may also be cured with an expensive treatment, however, they suffer from overtreatment then. Severely ill patients can only be cured with an expensive treatment. The author finds that, unless the induced effort is too small (in which case blind treatment choices are superior), in the first-best situation, physicians should exert effort and make an informed treatment choice. A payment scheme consisting of reward payments for successful treatment at first try is found to provide optimal incentives. Most notably, the scheme represents cost-sharing and has a pay-for-performance character because it is assumed that patients, who are not cured at first try, return to the physician to receive adequate treatment, while the physician receives zero payments for these patients. Pflum (2015) explicitly considers the test and treatment choice of a physician who has to compete for patients via treatment selection. Patients may be cured with two mutually exclusive treatments, whereby the optimal treatment varies with the patients' state of illness. Initially, the physician observes a private signal on the patient's state of illness. Based on the signal, the physician decides on the use of an optional, fully informative diagnostic test that reveals the patient's true state of illness and, thus, enables a profit-maximizing treatment choice. The model shows that only if the physician's private cost from utilizing the test is sufficiently high, a mixed contract, consisting of a capitation payment and fees depending on test and treatment choice, provides optimal incentives. Otherwise, the physician profits from inefficient overtesting and additional instruments such as utilization review are needed to implement the insurer's objective. Ghamat et al. (2017) also model the test and treatment decision of a physician. Some patients may only be cured with advanced treatment, while others may be cured with either basic or advanced treatment. The physician has private information about a patient's state of illness and decides

on the use of an optional diagnostic lab test, which determines whether advanced treatment is appropriate. The test's accuracy is not explicitly modelled though. Test results are private information to the physician, while the use of the test itself is verifiable and directly billed to the payer. Hence, a situation with moral hazard from hidden information is assumed. The authors find that, in the first-best case, it is not always socially optimal to make a diagnostic test compulsory, even if such policy can be implemented at no cost, because the test costs are too expensive. In the second-best case, a contract consisting of a fixed bundled payment for advanced treatment as well as reward payments for stable health after basic and advanced treatment, respectively, provides optimal incentives. Focusing on the inpatient sector, Brandt and Cassou (2018) study how a profit-maximizing hospital can be incentivized to make welfare-maximizing diagnostic test and treatment decisions in the presence of private information on the patient's primary symptoms of illness. In their model, hospitals can use an imperfect diagnostic test to update the belief about the patients' illness before choosing a treatment. The test's provision, the test's results and the subsequent treatment choice are verifiable by the regulator. The optimal diagnostic and treatment strategy is identified based on the value of information of the test such that, at the optimum, only selected patients should be tested. Flat-rate payments that are conditioned on the entire episode of care are found to be optimal. Yet, whenever some patients should be tested, the implementation of the first-best diagnostic and treatment strategies can cause information rents. Therefore, the second-best implementation may call for inefficient undertesting or even call for overtesting to reduce the implementation costs.

In addition, Garcia Mariñoso and Jelovac (2003) study the referral decision of a general practitioner (GP) in a gatekeeping-system. Although the authors focus on the interaction of physicians' diagnostic (testing) and referral choice - rather than diagnostic (testing) and treatment choice - the modelling approach and the results of this study provide relevant insights for this dissertation. In their model, some patients may only be cured by a specialist, whereas others may be cured by either provider. The GP can take costly diagnostic effort to learn whether a referral to a specialist is warranted. Diagnostic effort and diagnostic outcome are private information of the GP. Thus, a double moral hazard problem arises. The authors find that, in the first-best case, it may not always be optimal that the physician exerts diagnostic effort. Instead, there are circumstances, in which blind referral decisions are superior. In principle, a mixed contract, consisting of a capitation payment for performing the diagnosis and reward payments for successful treatment by the GP and for not providing primary care, respectively, is found to implement the first-best allocation. However, in the second-best case,

an information rent problem may arise from the GP's reservation utility constraint, leading to inefficient renunciation of diagnostic effort and over- or underreferral of the patient.

In summary, these theoretical studies show that the optimal use of a single diagnostic test (diagnostic effort) in a treatment or referral choice problem particularly depends on its value of information and that contracts, consisting of capitation payments and/or conditional FFS payments to aim at inducing appropriate treatment choice, provide optimal incentives in the first-best case. Yet, in the second-best case, moral hazard problems may force the regulator to trade-off information rent payments to the physician against allocative efficiency. Furthermore, the literature generally assumes that providers maximize their own expected utility, which does not account for patients' well-being. However, as already outlined in the seminal paper by Arrow (1963), health care markets are essentially characterized by physicians who are altruistic towards their patients. This may have an important effect on payment scheme design and market outcomes. Accordingly, the assumption of altruistic physicians has been adopted by various theoretical studies on the design of optimal payment contracts (Ellis and McGuire 1986; Jack 2005; Liu and Ma 2013). For instance, Ellis and McGuire (1986) find that in case a physician is a "perfect agent" for the patient, by equally accounting for his own and the patient's benefit, a pure prospective payment scheme implements the optimal outcome.

The study presented in Chapter 3 contributes to the previously discussed literature by combining the findings concerning the optimal utilization of two sequential diagnostic tests and the subsequent cost-efficient treatment choice in a situation with double moral hazard as well as the derivation of welfare maximizing payment contracts. In current practice, physicians usually apply some sort of pretest before costly (advanced) diagnostic tests, such as MRI scans of the lower back, are utilized. This study takes that observation into account and, to the best of my knowledge, is the first study on optimal payment contracts for two subsequent diagnostic tests in a treatment choice problem. The study further complements the literature by, firstly, showing that the incorporation of physicians' concern for the patients' well-being may mitigate, but not fully eliminate, the underlying incentive problem and by, secondly, explicitly modeling the sensitivity and specificity of the two tests. The latter allows for generalization to various diagnostic (testing) technologies.

Empirical evidence further indicates that, comparable to other expert professionals, physicians might be heterogenous in their diagnostic ability. For instance, Doyle et al. (2010) and, more recently, Chan et al. (2019), provide empirical evidence for heterogeneity in physicians' diagnostic ability. They studied physicians from various specialties in the inpatient sector and

radiologists for diagnosis of pneumonia and found significant differences in ability levels measured by diagnostic efficiency and accuracy, respectively. Assuming that diagnostic ability is private information to the physicians creates an adverse selection problem and, therefore, implementing optimal incentives for effective care becomes increasingly challenging. Essentially, unobserved heterogeneity in physicians' diagnostic ability should be accounted for when creating optimal incentives for physicians' diagnostic and treatment choice because heterogeneous physicians respond heterogeneously to incentives (Jelovac 2014).

While some of the above-mentioned theoretical studies on the interaction of physicians' diagnostic and treatment choice indeed address physicians' diagnostic ability by modelling endogenous, non-contractable diagnostic effort, they do not account for unobserved heterogeneity in diagnostic ability (Garcia Mariñoso and Jelovac 2003; Jelovac 2001).[3] However, there are two further studies that focus on referrals by heterogeneous gatekeeping GPs. Although the authors do not explicitly derive optimal incentives for diagnostic and treatment choice, their findings may still apply. Allard et al. (2011) and the companion paper by Allard et al. (2014) model a gatekeeping GP's referral problem, whereby GPs are heterogeneous in both diagnostic ability and altruism. Allard et al. (2011) assume that the exogenous ability of a GP determines the accuracy of the diagnostic signal. Patients suffer from a less severe or more severe illness; whereby less severely ill patients can be cured by both the GP and a specialist and more severely ill patients can only be cured by a specialist. If the latter patients are treated inappropriately by the GP, they are subsequently referred to the specialist and suffer from a health loss. The First-Best is determined by maximizing the patients' utility, subject to the GP's participation constraint. Therefore, in the first-best situation, the optimal referral strategy depends on the GPs' diagnostic ability: Only if the GP's diagnostic ability is sufficiently high, the GP should follow the outcome of the diagnosis for referral choice. This implies that in case concerns for wasteful referrals (for underreferrals) dominate, high-ability GPs should follow the diagnostic signal and low-ability should always treat the patient (refer the patient). Based on the first-best analysis, the authors investigate, which pure payment system - FFS, fundholding or capitation - is most efficient given a GP's ability-altruism-profile. Generally, the authors find that FFS payments and fundholding result in less referrals to costly specialist care than capitation payments. The identification of the most efficient payment

[3] As the focus of this dissertation lies on physicians' diagnostic and treatment choice and the associated problem of double moral hazard (and simultaneous adverse selection from heterogeneous diagnostic ability) and not on the problem of pure adverse selection, the literature on pure adverse selection is not further discussed. However, it should be emphasized that there is quite some evidence on pure adverse selection from physicians' heterogeneous ability or efficiency (i.a. Makris and Siciliani 2013) or, for instance, on simultaneous single moral hazard and adverse selection in quantity setting problems (Kantarevic and Kralj 2016).

scheme, however, particularly depends on the distribution of ability-profiles among the GPs. Focusing on the more relevant case for this dissertation, which is that concerns for wasteful referrals dominate (because the specialist's costs are relatively high and/or the probability of a severe illness is low), the capitation scheme never dominates as the GP would always refer. The FFS scheme implements the First-Best and tends to be superior if most GPs are highly able and not too altruistic such that GPs follow the outcome of their diagnosis. By contrast, fundholding implements the First-Best and strictly dominates if most GPs are less able and at least fairly altruistic such that GPs choose to systematically treat patients. Allard et al. (2014) extend the analysis by considering GPs' endogenous diagnostic ability resulting from endogenous effort. Hence, a situation with adverse selection from unobserved ability is assumed. By modelling GPs' behavior when they are paid by pure FFS, pure capitation or can self-select into either payment system, they show that GPs' self-selection is never optimal: GPs choose the level of effort that maximizes their expected utility given a payment scheme and a treatment/referral strategy. Under the FFS-follow-signal strategy, the most altruistic physicians endogenously become the most able physicians. In all other cases, the GP will choose to exert the least possible level of effort as the patients are systematically referred or treated anyway. Now, the highly able (highly altruistic) GPs have an incentive to mimic the low-ability GPs by, given their type, opting for the inefficient capitation system. The authors essentially conclude that this calls for a more sophisticated payment mechanism.

Further analyses by Alger and Ma (2003) reinforce the particular importance of the distribution of types in a market as noted by Allard et al. (2011). The authors examine optimal contracts for physicians who differ in their decision to collude with the patient when reporting to the payer. In case the payer offers a single policy contract, the distribution of the types essentially determines, whether a collusion- or a non-collusion-proof contract is superior. Only if the share of the dishonest physicians is sufficiently low, it is best to tolerate collusion. Moreover, they find that implementing a menu of policies would increase the overall welfare. Generally, classical principal-agent theory suggests that in situations with more than one type, contracts consisting of a menu of policies, which aim at inducing self-selection among the agents, provide optimal incentives (Laffont and Tirole 1993; Laffont and Martimort 2002). However, Gottlieb and Moreira (2014) show that for incentive problems in insurance and procurement with simultaneous adverse selection and moral hazard, offering a menu of policies may be inefficient. Moral hazard entails additional incentive-compatibility constraints (ICs) and agents, who exert low effort, may earn positive rents from mimicking high effort types. Because the

ICs are binding, implementing a menu would imply a further effort distortion. This contrasts the *no distortion at the top* property of screening contracts.

The study presented in Chapter 5 of this dissertation contributes to the discussed literature by deriving payment contracts that implement optimal diagnostic and treatment choice under the quite realistic, yet also more sophisticated, assumption of unobserved heterogeneity in physicians' diagnostic ability. In particular, to the best of my knowledge, it is the first paper that examines optimal incentives for physicians' diagnostic and treatment choice given simultaneous double moral hazard and adverse selection. The study further complements the literature by, firstly, analyzing the interaction of the incentive effects from double moral hazard and from adverse selection on physicians' behavior and the implications for optimal payment scheme design and by, secondly, considering different operationalizations of physicians' diagnostic ability – that is diagnostic accuracy and diagnostic efficiency. While the former leads to a deeper understanding of the underlying incentive problems, the latter allows for generalization to a broad range of applications.

2.2 The Provision of Advanced Diagnostic Tests, in Particular MRI Scans of the Female Breast, and the Impact on Cost and Outcomes

Physicians' decision to provide advanced diagnostic (test) services, in particular MRI scans of the female breast for breast cancer treatment, may generally be influenced by a variety of factors. Focusing on countries with national health services or compulsory health insurance schemes, regulators usually decide on the allocation of limited health care resources (as outlined at the very beginning of this chapter). Thus, inter alia, depending on the medical indication, individual patient characteristics or patients' medical history, regulators evaluate and enact, which health care services are covered and reimbursed. In addition, evidence-based clinical practice guidelines, published by expert societies, serve as further decision aids for physicians and patients alike (Niessen et al. 2000). However, evidence suggests that physicians may ignore the regulations and reimbursement conditions defined by the regulator, leading to overprovision of costly diagnostic imaging services for breast cancer treatment (Baxi et al. 2017 for an international survey). Facilitated by information asymmetries between physicians, the public insurer and patients, physicians may provide the amount of services that is in their own best interest rather than adhering to the regulator's requirements (McGuire 2000).

In the literature, several potential reasons for overprovision of health care services are noted. First and foremost, the incentives that are placed on physicians by a payment system influence

their actions (Epstein et al. 1986; Gosden et al. 2001 for a systematic review; McGuire 2000). Regarding the provision of outpatient MRI scans of the breast in Germany, incentive problems arise at two stages of the care process: In Germany, most outpatient care physicians, including gynecologists, are predominantly reimbursed by capitation, whereas radiologists belong to the few specialties of outpatient physicians that tend to be reimbursed by FFS rather than by capitation (Federal Association of Statutory Health Insurance Physicians 2021). Thus, the attending gynecologists have an incentive to minimize the number of consultations in a given time period and to refer the patient to receive an MRI scan at a radiologist instead of, for instance, taking an effort and explaining to the patient, why the MRI scan is not indicated (Gosden et al. 2001). Radiologists subsequently have an incentive to provide the MRI scan as they bear high fixed cost for the provision of breast MRIs and, therefore, benefit from each additional scan they bill.

Furthermore, physicians may fear malpractice liability cost. Since clinical outcomes are sometimes hard to measure, resulting in unclear legal situations, physicians may practice defensive medicine. Hence, especially diagnostic imaging services are prone to overuse because a lack of provision is often perceived to have more serious consequences for patients' health than overprovision (Baicker et al. 2007; Danzon 2000; Lucas et al. 2010). Moreover, physicians may be subject to peer effects (Chandra and Staiger 2007) or react to competition. Theoretical evidence shows that in case providers have to compete to attract patients, concerns for inefficient overuse of diagnostic tests should be higher (Brandt and Cassou 2018). Essentially, physicians may feel pressured by patients' expectations to receive advanced diagnostic services (Little et al. 2004). In Germany, and in numerous other OECD-countries, patients are practically fully insured, which gives rise to a moral hazard problem. Because patients do not have to bear the marginal cost of care provision, they may ask physicians for allegedly beneficial, yet, ineffective services or even misreport on their condition to receive these services (Einav and Finkelstein 2018). Finally, physicians may be convinced that the diagnostic service is clinically indicated, despite knowing that the regulator's requirements are not met. Or, the physician simply does not know, which alternative service to provide to help the patient. Motivated by the Hippocratic Oath and the concern for the patient's well-being, an ineffective amount of services may be provided (Brownlee 2007).

The identification of overprovision of advanced diagnostic services, in particular of breast MRI scans, is beneficial for public insurers in the short-term, because it may form the basis for regress claims. Yet, from a societal perspective, the more relevant question would be, whether overprovision is indeed ineffective and should be targeted by the regulator (and insurers alike).

As already discussed at the beginning of this chapter, coverage and reimbursement regulations are not always based on high quality (clinical) evidence.

Indeed, international evidence suggests that, when applied for primary diagnosis or for surveillance of breast cancer treatment, breast MRI scans may generally have controversial effects on patient outcomes and health care cost: To begin with, it is unclear whether breast MRI leads to improved disease-specific survival. While evidence from randomized controlled trials (RCTs) is not yet available, most observational studies based on clinical data, cancer registries and/or Medicare claims from the United States (U.S.) indicate that breast MRI has no significant effect on survival (Onega et al. 2018; Parsyan et al. 2013 for a systematic review). Yet, Wang et al. (2018), for instance, also find that a subgroup of women, who received preoperative breast MRI but no radiotherapy, benefits from a significantly lower risk of breast cancer mortality than similar women who did not receive such MRI. The authors analyzed patient outcomes from the Surveillance, Epidemiology, and End Results (SEER)-Medicare database, adjusted for patient characteristic (inter alia, social status, Elixhauser comorbidity groups and tumor characteristics) and applied a log-rank test to compare survival functions.

Moreover, partly due to possible false-positive results and over-diagnosis, breast MRI is associated with increased probability of further diagnostic investigation such as repeat MRI, biopsy and additional physician referrals (Brennan et al. 2010; Padia et al. 2016). Further, patients who received an MRI are more likely to receive extensive surgeries, while re-operation rates do not seem to decline (Houssami et al. 2017). All those procedures are found to result in significantly increased cost of care (Bedrosian et al. 2016; Hayes et al. 2016). Finally, based on single and multi-center studies from Canada and France, respectively, there is mixed evidence on whether breast MRI scans also lead to increased patient distress and anxiety (Brédart et al. 2012; Spiegel et al. 2011). In particular, Padia et al. (2016) review breast MRI scans that were provided at a single institution in the U.S. including newly diagnosed patients and patients for surveillance. They report that false-positive findings result in further imaging (in 51.6 % of all cases), additional laboratory testing (8.6%), biopsies (5.7%) and referrals (5.7%). Based on the Medicare FFS schedule and applying Welch's t-test, the authors conclude that these procedures add up to significantly increased cost of $328 per patient. Houssami et al. (2017) conduct a meta-analysis of RCTs and comparative studies on the effect of preoperative MRI. Applying a random-effects logistic regression, they find that preoperative MRI is associated with significantly increased odds of receiving mastectomy. Further, Wang et al. (2016) use the SEER-Medicare database in order to explore the effects of preoperative MRI. After adjusting for observed heterogeneity in patient characteristics by applying propensity score matching and

subsequently developing a Markov model, they find that breast MRI leads to over-diagnosis as increased detection rates, attributable to MRI, are not offset by a decrease of subsequent occurrence among women with early-stage BC. Finally, Bedrosian et al. (2016) conduct a retrospective cost study in the U.S. using institutional billing records. Running linear regressions with log-transformed costs, the authors find that women who received a preoperative breast MRI scan on average cause significantly higher medical cost (+30.1%) in the after period than women who did not. The authors further show that the results remain robust when additionally controlling for clinical factors.

The study presented in Chapter 4 of this dissertation contributes to the discussed literature in several ways. Firstly, to the best of my knowledge, it constitutes the first study that investigates whether outpatient breast MRI scans in Germany are provided compliant with the reimbursement conditions of the social health insurance. Secondly, the study comprehensively investigates the effect of breast MRI scans on a total of 27 outcomes across various outcome categories. Thirdly, as the study is based on administrative data from Germany, women aged 18 and older can be observed and data can be analyzed across institutions and care sectors. By contrast, most of the related studies only analyze outcomes from a single category, consider data from single institutions and/or focus on data from other OECD-countries, where diagnosis and treatment patterns, decision-making factors, health service / insurance structures and fiscal considerations may be different and not generalizable to the German population. Fourthly, propensity score matching and a difference-in-difference approach are applied to overcome the limitation of observed and unobserved heterogeneity from administrative data.

References

Alger, Ingela; Ma, Ching-to Albert (2003). Moral hazard, insurance, and some collusion. Journal of Economic Behavior & Organization 50 (2), pp. 225–247. DOI: 10.1016/S0167-2681(02)00049-5.

Allard, Marie; Jelovac, Izabela; Léger, Pierre Thomas (2011). Treatment and referral decisions under different physician payment mechanisms. Journal of Health Economics 30 (5), pp. 880–893. DOI: 10.1016/j.jhealeco.2011.05.016.

Allard, Marie; Jelovac, Izabela; Léger, Pierre-Thomas (2014). Payment mechanism and GP self-selection: capitation versus fee for service. International Journal of Health Care Finance and Economics 14 (2), pp. 143–160. DOI: 10.1007/s10754-014-9143-z.

Alonso-Coello, Pablo; Schünemann, Holger J.; Moberg, Jenny; Brignardello-Petersen, Romina; Akl, Elie A.; Davoli, Marina et al. (2016). GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ (Clinical research ed.) 353, i2016. DOI: 10.1136/bmj.i2016.

Arrow, Kenneth J. (1963). Uncertainty and the Welfare Economics of Medical Care. The American Economic Review 53 (5), pp. 941–973.

Attia, John (2003). Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. Australian Prescriber 26 (5).

Baicker, Katherine; Fisher, Elliott S.; Chandra, Amitabh (2007). Malpractice liability costs and the practice of medicine in the Medicare program. Health Affairs (Project Hope) 26 (3), pp. 841–852. DOI: 10.1377/hlthaff.26.3.841.

Baxi, Shrujal S.; Kale, Minal; Keyhani, Salomeh; Roman, Benjamin R.; Yang, Annie; Derosa, Antonio P.; Korenstein, Deborah (2017). Overuse of Health Care Services in the Management of Cancer: A Systematic Review. Medical Care 55 (7), pp. 723–733. DOI: 10.1097/MLR.0000000000000734.

Bedrosian, Isabelle; Xing, Yan; Abdel Rahman, Shereen; Allen, Lisa; Le-Petross, Huong; Whitman, Gary J. et al. (2016). A Cost Analysis of Preoperative Breast MRI Use for Patients with Invasive Lobular Cancer. Annals of Surgical Oncology 23 (1), pp. 23–29. DOI: 10.1245/s10434-015-4710-1.

Brandt, Nikolai M.; Cassou, Matthieu (2018). Incomplete Information and Economic Incentives in Health Care. Unpublished Paper. Paper presented at the 19th European Health Economics Workshop, Paris, May 2018.

Brédart, Anne; Kop, Jean-Luc; Fall, Mouhamadou; Pelissier, Sandra; Simondi, Cécile; Dolbeault, Sylvie et al. (2012). Anxiety and specific distress in women at intermediate and high risk of breast cancer before and after surveillance by magnetic resonance imaging and mammography versus standard mammography. Psycho-Oncology 21 (11), pp. 1185–1194. DOI: 10.1002/pon.2025.

Brennan, Sandra; Liberman, Laura; Dershaw, D. David; Morris, Elizabeth (2010). Breast MRI screening of women with a personal history of breast cancer. American Journal of Roentgenology 195 (2), pp. 510–516. DOI: 10.2214/AJR.09.3573.

Brosig-Koch, Jeannette; Hennig-Schmidt, Heike; Kairies-Schwarz, Nadja; Wiesen, Daniel (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. Journal of Economic Behavior & Organization 131, pp. 17–23. DOI: 10.1016/j.jebo.2015.04.011.

Brosig-Koch, Jeannette; Hennig-Schmidt, Heike; Kairies-Schwarz, Nadja; Wiesen, Daniel (2017). The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence. Health economics 26 (2), pp. 243–262. DOI: 10.1002/hec.3292.

Brownlee, Shannon (2007). Overtreated. Why Too Much Medicine Is Making Us Sicker and Poorer. New York: Bloomsbury.

Chalkley, Martin; Listl, Stefan (2018). First do no harm - The impact of financial incentives on dental X-rays. Journal of Health Economics 58, pp. 1–9. DOI: 10.1016/j.jhealeco.2017.12.005.

Chalkley, Martin; Malcomson, James M. (1998). Contracting for health services when patient demand does not reflect quality. Journal of Health Economics 17 (1), pp. 1–19. DOI: 10.1016/S0167-6296(97)00019-2.

Chan, David; Gentzkow, Matthew; Yu, Chuan (2019). Selection with Variation in Diagnostic Skill: Evidence from Radiologists. NBER Working Paper Series. Working Paper No. 26467. DOI: 10.3386/w26467.

Chandra, Amitabh; Staiger, Douglas O. (2007). Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. Journal of Political Economy 115, pp. 103–140. DOI: 10.1086/512249.

Danzon, Patricia M. (2000). Liability for medical malpractice. In : Handbook of Health Economics. Volume 1, Part B: Elsevier, pp. 1339–1404. DOI: 10.1016/S1574-0064(00)80039-6.

Darby, Michael R.; Karni, Edi (1973). Free Competition and the Optimal Amount of Fraud. The Journal of Law and Economics 16 (1), pp. 67–88. DOI: 10.1086/466756.

Doyle, Joseph J.; Ewer, Steven M.; Wagner, Todd H. (2010). Returns to physician human capital: evidence from patients randomized to physician teams. Journal of Health Economics 29 (6), pp. 866–882. DOI: 10.1016/j.jhealeco.2010.08.004.

Dranove, David; Satterthwaite, Mark A. (2000). The industrial organization of health care markets. In : Handbook of Health Economics, Volume 1, Part B: Elsevier, pp. 1039–1139. DOI: 10.1016/S1574-0064(00)80033-5.

Dulleck, Uwe; Kerschbamer, Rudolf (2006). On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. Journal of Economic Literature 44 (1), pp. 5–42. DOI: 10.1257/002205106776162717.

Einav, Liran; Finkelstein, Amy (2018). Moral Hazard in Health Insurance: What We Know and How We Know It. Journal of the European Economic Association 16 (4), pp. 957–982. DOI: 10.1093/jeea/jvy017.

Ellis, Randall P.; McGuire, Thomas G. (1986). Provider Behaviour under Prospective Reimbursement: Cost-sharing and Supply. Journal of Health Economics 5 (2), pp. 129–151. DOI: 10.1016/0167-6296(86)90002-0.

Ellis, Randall P.; McGuire, Thomas G. (1993). Supply-side and demand-side cost sharing in health care. The journal of economic perspectives : a journal of the American Economic Association 7 (4), pp. 135–151. DOI: 10.1257/jep.7.4.135.

Epstein, Arnold M.; Begg, Colin B.; McNeil, Barbara J. (1986). The use of ambulatory testing in prepaid and fee-for-service group practices. Relation to perceived profitability. The New England journal of medicine 314 (17), pp. 1089–1094. DOI: 10.1056/NEJM198604243141706.

Fainman, Emily Zhu; Kucukyazici, Beste (2020). Design of financial incentives and payment schemes in healthcare systems: A review. Socio-Economic Planning Sciences 72, p. 100901. DOI: 10.1016/j.seps.2020.100901.

Federal Association of Statutory Health Insurance Physicians (2021). Online Version of the EBM 2021. Berlin. Available online at https://www.kbv.de/html/online-ebm.php, checked on 1/28/2021.

Felder, Stefan; Mayrhofer, Thomas (2017). Medical Decision Making. 2nd Edition. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-53432-8.

Garber, Alan M. (2000). Advances in cost-effectiveness analysis of health interventions. In : Handbook of Health Economics, Volume 1, Part A: Elsevier, pp. 181–221. DOI: 10.1016/S1574-0064(00)80163-8.

Garcia Mariñoso, Begoña; Jelovac, Izabela (2003). GPs' payment contracts and their referral practice. Journal of Health Economics 22 (4), pp. 617–635. DOI: 10.1016/S0167-6296(03)00008-0.

Ghamat, Salar; Zaric, Gregory S.; Pun, Hubert (2017). Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment. Production and Operations Management 63 (5), p. 1606. DOI: 10.1111/poms.12780.

Gosden, T.; Forland, F.; Kristiansen, I. S.; Sutton, M.; Leese, B.; Giuffrida, A. et al. (2001). Impact of payment method on behaviour of primary care physicians: a systematic review. Journal of Health Services Research & Policy 6 (1), pp. 44–55. DOI: 10.1258/1355819011927198.

Gottlieb, Daniel; Moreira, Humberto (2014). Simultaneous Adverse Selection and Moral Hazard. Working Paper. Available online at https://epge.fgv.br/users/humberto/wp-content/uploads/2015/07/ASMH_09072014.20.pdf, checked on 1/17/2021.

Gottschalk, Felix; Mimra, Wanda; Waibel, Christian (2020). Health Services as Credence Goods: a Field Experiment. The Economic Journal 130 (629), pp. 1346–1383. DOI: 10.1093/ej/ueaa024.

Hayes, Lisa M.; Frebault, Julia S.; Landercasper, Jeffrey; Borgert, Andrew J.; Vang, Choua A.; Johnson, Jeanne M.; Linebarger, Jared H. (2016). Extramammary findings in diagnostic breast magnetic resonance imaging among patients with known breast cancer: incidence and cost analysis. American Journal of Surgery 212 (6), pp. 1194–1200. DOI: 10.1016/j.amjsurg.2016.09.020.

Hennig-Schmidt, Heike; Selten, Reinhard; Wiesen, Daniel (2011). How payment systems affect physicians' provision behaviour--an experimental investigation. Journal of Health Economics 30 (4), pp. 637–646. DOI: 10.1016/j.jhealeco.2011.05.001.

Holmström, Bengt (2017). Pay For Performance and Beyond. American Economic Review 107 (7), pp. 1753–1777. DOI: 10.1257/aer.107.7.1753.

Holmström, Bengt; Milgrom, Paul (1991). Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. The Journal of Law, Economics, and Organization 7 (special_issue), pp. 24–52. DOI: 10.1093/jleo/7.special_issue.24.

Houssami, Nehmat; Turner, Robin M.; Morrow, Monica (2017). Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer. Breast Cancer Research and Treatment 165 (2), pp. 273–283. DOI: 10.1007/s10549-017-4324-3.

Jack, William (2005). Purchasing health care services from providers with unknown altruism. Journal of Health Economics 24 (1), pp. 73–93. DOI: 10.1016/j.jhealeco.2004.06.001.

Jelovac, Izabela (2001). Physicians' payment contracts, treatment decisions and diagnosis accuracy. Health economics 10 (1), pp. 9–25. DOI: 10.1111/poms.12780.

Jelovac, Izabela (2014). Primary Care, Gatekeeping, and Incentives. In : Encyclopedia of Health Economics: Elsevier, pp. 142–145. DOI: 10.1016/B978-0-12-375678-7.01311-0.

Kantarevic, Jasmin; Kralj, Boris (2016). Physician Payment Contracts in the Presence of Moral Hazard and Adverse Selection: The Theory and Its Application in Ontario. Health economics 25 (10), pp. 1326–1340. DOI: 10.1002/hec.3220.

Kerschbamer, Rudolf; Sutter, Matthias (2017). The Economics of Credence Goods – a Survey of Recent Lab and Field Experiments*. CESifo Economic Studies 63 (1), pp. 1–23. DOI: 10.1093/cesifo/ifx001.

Kralj, Boris; Kantarevic, Jasmin (2013). Quality and quantity in primary care mixed-payment models: evidence from family health organizations in Ontario. Canadian Journal of Economics/Revue canadienne d'économique 46 (1), pp. 208–238. DOI: 10.1111/caje.12003.

Laffont, Jean-Jacques; Martimort, David (2002). The Theory of Incentives. The Principal-Agent Model. Princeton: Princeton University Press. DOI: 10.2307/j.ctv7h0rwr.

Laffont, Jean-Jacques; Tirole, Jean (1993). A Theory of Incentives in Procurement and Regulation. 5th ed. Cambridge, Massachusetts: The MIT Press.

Little, Paul; Dorward, Martina; Warner, Greg; Stephens, Katharine; Senior, Jane; Moore, Michael (2004). Importance of patient pressure and perceived pressure and perceived

medical need for investigations, referral, and prescribing in primary care: nested observational study. BMJ (Clinical research ed.) 328 (7437), p. 444. DOI: 10.1136/bmj.38013.644086.7C.

Liu, Fang; Rasch, Alexander; Schwarz, Marco A.; Waibel, Christian (2020). The Role of Diagnostic Ability in Markets for Expert Services. CESifo Working Paper No. 8704. Available online at SSRN: https://ssrn.com/abstract=3736486, checked on 12/29/2020.

Liu, Ting; Ma, Ching-to Albert (2013). Health insurance, treatment plan, and delegation to altruistic physician. Journal of Economic Behavior & Organization 85, pp. 79–96. DOI: 10.1016/j.jebo.2012.11.002.

Lucas, Frances Lee; Sirovich, Brenda E.; Gallagher, Patricia M.; Siewers, Andrea E.; Wennberg, David E. (2010). Variation in cardiologists' propensity to test and treat: is it associated with regional variation in utilization? Circulation. Cardiovascular quality and outcomes 3 (3), pp. 253–260. DOI: 10.1161/CIRCOUTCOMES.108.840009.

Ma, Ching-to Albert (1994). Health Care Payment Systems: Cost and Quality Incentives. J Economics Management Strategy 3 (1), pp. 93–112. DOI: 10.1111/j.1430-9134.1994.00093.x.

Makris, Miltiadis; Siciliani, Luigi (2013). Optimal Incentive Schemes for Altruistic Providers. Journal of Public Economic Theory 15 (5), pp. 675–699. DOI: 10.1111/jpet.12047.

McClellan, Mark (2011). Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality? Journal of Economic Perspectives 25 (2), pp. 69–92. DOI: 10.1257/jep.25.2.69.

McGuire, Thomas G. (2000). Physician Agency. In : Handbook of Health Economics, Volume 1, Part A: Elsevier, pp. 461–536. DOI: 10.1016/S1574-0064(00)80168-7.

Mossialos, Elias; Djordjevic, Ana; Osborn, Robin; Sarnak, Dana (2017). International Profiles of Health Care Systems. Australia, Canada, China, Denmark, England, France, Germany, India, Israel, Italy, Japan, the Netherlands, New Zealand, Norway, Singapore, Sweden, Switzerland, Taiwan, and the United States. New York: The Commonwealth Fund.

Niessen, Louis W.; Grijseels, Els W.M; Rutten, Frans F.H (2000). The evidence-based approach in health policy and health care delivery. Social Science & Medicine 51 (6), pp. 859–869. DOI: 10.1016/S0277-9536(00)00066-6.

OECD (2021). Magnetic resonance imaging (MRI) exams. Available online at https://www.oecd-ilibrary.org/social-issues-migration-health/magnetic-resonance-imaging-mri-exams/indicator/english_1d89353f-en, checked on 4/24/2021.

Onega, Tracy L.; Zhu, W.; Weiss, J. E.; Goodrich, M.; Tosteson, A. N. A.; DeMartini, W. et al. (2018). Preoperative breast MRI and mortality in older women with breast cancer. Breast Cancer Research and Treatment 170 (1), pp. 149–157. DOI: 10.1007/s10549-018-4732-z.

Padia, Shilpa A.; Freyvogel, Mary; Dietz, Jill; Valente, Stephanie; O'Rourke, Colin; Grobmyer, Stephen R. (2016). False-positive Extra-Mammary Findings in Breast MRI: Another Cause for Concern. The Breast Journal 22 (1), pp. 90–95. DOI: 10.1111/tbj.12524.

Parsyan, Armen; Alqahtani, Awadh; Mesurolle, Benoit; Meterissian, Sarkis (2013). Impact of preoperative breast MRI on surgical decision making and clinical outcomes: a systematic review. World Journal of Surgery 37 (9), pp. 2134–2139. DOI: 10.1007/s00268-013-2077-7.

Pflum, Kevin E. (2015). Physician Incentives and Treatment Choice. Journal of Economics and Management Strategy 24 (4), pp. 712–751. DOI: 10.1111/jems.12117.

Schünemann, Holger J.; Oxman, Andrew D.; Brozek, Jan; Glasziou, Paul; Jaeschke, Roman; Vist, Gunn E. et al. (2008). Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ (Clinical research ed.) 336 (7653), pp. 1106–1110. DOI: 10.1136/bmj.39500.677199.AE.

Shih, Terry; Chen, Lena M.; Nallamothu, Brahmajee K. (2015). Will Bundled Payments Change Health Care? Examining the Evidence Thus Far in Cardiovascular Care. Circulation 131 (24), pp. 2151–2158. DOI: 10.1161/CIRCULATIONAHA.114.010393.

Spiegel, Tirtza N.; Esplen, Mary Jane; Hill, Kimberley A.; Wong, Jiahui; Causer, Petrina A.; Warner, Ellen (2011). Psychological impact of recall on women with BRCA mutations undergoing MRI surveillance. The Breast 20 (5), pp. 424–430. DOI: 10.1016/j.breast.2011.04.004.

Street, Andrew David; O'Reilly, Jacqueline; Ward, Padraic; Mason, Anne (2011). DRG-based hospital payment and efficiency: theory, evidence, and challenges. In : Diagnosis Related Groups in Europe: moving towards transparency, efficiency and quality in hospitals. Maidenhead: Open University Press, pp. 93–114.

Wang, Shi-Yi; Long, Jessica B.; Killelea, Brigid K.; Evans, Suzanne B.; Roberts, Kenneth B.; Silber, Andrea; Gross, Cary P. (2016). Preoperative Breast Magnetic Resonance Imaging and Contralateral Breast Cancer Occurrence Among Older Women With Breast Cancer. JCO 34 (4), pp. 321–328. DOI: 10.1200/JCO.2015.62.9741.

Wang, Shi-Yi; Long, Jessica B.; Killelea, Brigid K.; Evans, Suzanne B.; Roberts, Kenneth B.; Silber, Andrea L. et al. (2018). Associations of preoperative breast magnetic resonance imaging with subsequent mastectomy and breast cancer mortality. Breast Cancer Research and Treatment 172 (2), pp. 453–461. DOI: 10.1007/s10549-018-4919-3.

Chapter 3

Optimal Financial Incentives for Physician's Sequential Diagnostic Testing and Treatment Choice

Katharina Beenk Mathias Kifmann

Abstract

Many medical contexts leave ample scope for the utilization of advanced diagnostic tests and for treatment choice. Yet, policy makers are concerned that, facilitated by remuneration schemes, current medical practice leads to an extensive use of diagnostic tests and cost-inefficient treatment choice. This paper examines the use of diagnostic tests to improve cost-efficient treatment choice. In a first-best analysis, we show that it can be optimal to apply an imperfect test first and to use an expensive subsequent test only conditional on the outcome of the first test. We then determine contracts, which implement optimal sequential diagnostic testing and treatment choice given the physician's effort for the first test is not observable and the test's results are private information to the physician. We show that a mixed payment contract, consisting of a capitation payment, fee-for-service payments with strictly positive mark-ups conditional on adequate treatment choice and a fee for the utilization of the expensive subsequent test, provides optimal incentives for the physician. This finding contrasts the less restrictive payments contracts, as commonly used in numerous OECD countries, which do not put sufficient responsibility for ordering expensive diagnostic tests on the attending physician and allow for unconditional FFS payments.

3.1 Introduction

Overuse of medical services has led to cost concerns among policymakers because limited health care resources are wasted and diverted from the budget for beneficial services. Indeed, there is strong empirical evidence that, across countries, especially the increased utilization of advanced diagnostic tests is worrisome (Brownlee et al. 2017). Magnetic resonance imaging (MRI), for instance, has become a widely employed diagnostic device. In Germany, the number of MRI exams in inpatient as well as outpatient care on average increased by more than 5% a year from 2010 until 2017, reaching 149 exams per 1,000 inhabitants in 2017 (OECD 2021). In international comparison, France and the United States show a similarly high quota (OECD 2021). While these tests often provide superior diagnostics compared to other tests, a key question from an economic perspective is whether they are cost-efficient or not. In particular, it should be considered to what extent the use of expensive diagnostic tests contributes to improved treatment choice. Evidence suggests that there are gaps between actual physician practice and evidence-based clinical practice guidelines, which recommend a prudent use of MRI, e.g. for the management of lower back pain (Foster et al. 2018). An additional consideration is the availability of other, less expensive diagnostic tests.

There are many potential reasons for the extensive use of expensive diagnostic tests, including physicians' concern about litigation and poor diagnostic ability, availability of the tests as well as peer effects and patients' expectations (Baras and Baker 2009; Baicker et al. 2007; Chandra and Staiger 2007; Doyle et al. 2010; Little et al. 2004). Further, the remuneration scheme plays an important role. In the principal-agent framework, the incentives that are placed on physicians by a payment system essentially determine their actions (Ellis and McGuire 1986, 1990; McClellan 2011; McGuire 2000). Due to the lack of cost accountability under fee-for-service (FFS) schemes, simply ordering an expensive diagnostic test may be viewed as more expedient than exerting diagnostic effort or explaining to a patient why the expensive test is not advisable (Levin et al. 2017).

In this paper, we analyze a situation, in which diagnostic tests are helpful in choosing between two treatment alternatives under diagnostic risk. Patients may be mildly or severely ill. Performing a basic treatment is only useful in case the patient is mildly ill. By contrast, an advanced treatment cures the patient regardless of the severity of illness. An example is the treatment of lower back pain. While basic treatment corresponds to conservative treatments, such as advising the patient to remain active or educating him about his condition, advanced treatment involves invasive procedures, such as spinal injections or surgery (Foster et al. 2018).

Overall, two tests are available for diagnosis: Firstly, an imperfect pretest can be performed, which requires diagnostic effort. Secondly, a superior test can be ordered. Because the superior test is assumed to be more expensive, it is not always optimal to use it. In particular, the optimal use of expensive diagnostic tests may involve the use of the pretest. For instance, advanced treatment should be performed right away if the pretest indicates a severe illness; and the costly diagnostic test should only be used for further investigation if the pretest indicates a mild illness. Focusing on the implementation of situations, in which it is optimal to apply an imperfect pretest first and to use an expensive subsequent test only conditional on the outcome of the pretest, we present a principal-agent model with a public insurer (the payer) and an outpatient physician. A key assumption is that the payer can neither observe whether the pretest was performed nor the resulting diagnostic signal. Regarding the example of lower back pain, it may be impossible or prohibitively costly for the payer to verify how thoroughly an orthopedist performed a physical examination. This creates a double moral hazard problem. In a FFS system, for instance, the orthopedist's profit-maximizing choice is to substitute diagnostic effort by ordering the utilization of a superior, yet more costly diagnostic device such as an MRI exam. We determine an optimal payment scheme which takes these incentives into account.

While there is vast literature on optimal incentives for diagnostic (test) choice and on treatment choice in particular (see, for instance, Felder and Mayrhofer 2017; Jelovac 2001), research on the interaction of both aspects is still scarce. The papers most closely related to our work are Garcia Mariñoso and Jelovac (2003), Pflum (2015), Ghamat et al. (2017) as well as Brandt and Cassou (2018). Garcia Mariñoso and Jelovac (2003) study the referral decision of a general practitioner (GP) in a gatekeeping-system. Some patients may only be cured by a specialist, whereas others may be cured by either provider. In their model, the GP can take diagnostic effort to learn whether a referral to a specialist is warranted. Diagnostic effort and diagnostic outcome are private information of the GP; thus, a moral hazard problem arises. The authors show that a mixed contract, consisting of a capitation payment for performing a diagnosis as well as reward payments for successful treatment by the GP and for not providing primary care, respectively, provides optimal incentives. Yet, in the second-best case, a rent-efficiency trade-off may result. Pflum (2015) considers the test and treatment choice of a physician who has to compete for patients. Patients may be cured with two mutually exclusive treatments. Initially, the physician observes a private signal on the patient's state of illness. Based on the signal, the physician decides on the use of an optional, costly diagnostic test that reveals the patient's true state of illness and, thus, enables a profit-maximizing treatment choice. If the physician's private cost from utilizing the test are sufficiently high, a mixed contract, consisting of a

capitation payment and fees depending on test and treatment choice, provides optimal incentives. Otherwise, additional payment instruments will be needed. Ghamat et al. (2017) also model the test and treatment decision of a physician. Some patients may only be cured with advanced treatment, while others may be cured with either basic or advanced treatment. As in Pflum (2015), the physician has private information about a patient's state of illness and decides on the use of an optional diagnostic test, which determines whether advanced treatment is actually needed. Test results are private information to the physician, while the test itself is directly billed to the payer. They find that a contract consisting of a fixed bundled payment for advanced treatment as well as reward payments for stable health after basic and advanced treatment, respectively, is optimal in the second-best case. Brandt and Cassou (2018) study how a profit-maximizing hospital can be incentivized to make welfare-maximizing diagnostic test and treatment decisions in the presence of private information on the patient's primary symptoms. Flat-rate payments that are conditioned on the entire episode of care are found to be optimal. Yet, whenever some patients should be tested, the implementation of the first-best diagnostic and treatment strategies can cause information rents. The second-best implementation may even call for over- or undertesting.

To the best of our knowledge, we are the first to study optimal payment contracts for two subsequent tests in a treatment choice problem. Building on the model by Garcia Mariñoso and Jelovac (2003), we show that, in the first-best case, an optimal diagnosis and treatment course can consist in applying the pretest first and using the expensive test only conditional on the outcome of the pretest. We find that a mixed payment contract provides optimal incentives for the implementation of a diagnosis and treatment course that entails the selective utilization of an expensive diagnostic test only conditional on the results from a less costly pretest. The mixed contract consists of a capitation payment for performing a diagnosis, FFS payments with strictly positive mark-ups conditional on adequate treatment choice as well as a fee for the utilization of the expensive subsequent test. By contrast, a less restrictive contract, comprising only a capitation payment and standard FFS payments, as common in numerous OECD countries, is not incentive compatible. Furthermore, we show that implementing the optimal diagnosis and treatment course can create information rents for the physicians. These can make alternative diagnosis and treatment courses superior.

The paper is organized as follows: In Section 3.2, we set up the model. Section 3.3 presents the first-best allocation as a benchmark case. In Section 3.4, we focus on the implementation of the first-best solution. Section 3.5 considers an alternative payment contract. Section 3.6 determines the second-best allocation. Finally, Section 3.7 discusses and concludes the paper.

3.2 The Model

3.2.1 The Patient

A patient who experiences symptoms of illness visits a physician. He suffers from a severe illness with probability $p \in (0; 1)$ and from a mild illness with probability $(1 - p)$.[1] Probability p may be interpreted as the prevalence of the severe illness in the population of patients who see a physician. In case the patient suffers from the mild illness, he may be cured with either a basic treatment T^B or an advanced treatment T^A. In case he suffers from the severe illness though, he may only be cured with T^A as T^B is insufficient. Either way, the patient eventually needs to receive sufficient treatment. Thus, if a severely ill patient is initially treated with T^B, he always returns to the physician to receive T^A on top and he suffers from a health loss with a monetary equivalent of $\ell > 0$.[2] The health loss may be interpreted as the waiting cost related to delays in obtaining the sufficient treatment T^A (Garcia Mariñoso and Jelovac 2003). We assume that patients are fully insured and play a passive role, in the sense that they always consent to the physician's actions.

3.2.2 The Physician

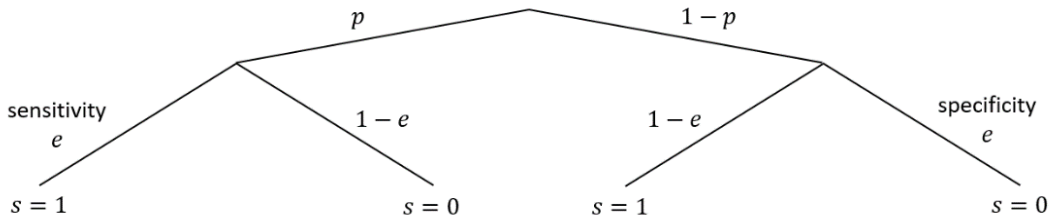
When initially seeing a patient, the physician cannot tell whether the patient suffers from the mild or the severe illness. She is only aware of the prevalence p of the severe illness in the patient population. To cure the patient, the physician takes three choices: Firstly, for each patient, she decides on the utilization of a pretest (test 1). The pretest yields an informative, yet imperfect, signal on the patient's state of illness and incurs effort cost on the physician, such as a detailed physical examination for the diagnosis of lower back pain. Secondly, she decides on the utilization of a subsequent test that yields a more precise signal (test 2). Test 2 is more expensive than test 1 and involves no effort but monetary cost. This may apply to an MRI exam of the lower back, which is provided by a radiologist. The physician may make use of none, both or either one of the two tests. We assume that the decision on test 2 is made conditional on the decision and, if applicable, on the results from test 1. Thirdly, conditional on the results from the two diagnostic tests, she decides on treatment intensity. In the following, we will illustrate the physician's three choices in detail.

[1] For ease of exposition, we adopt the linguistic convention that the patient and the payer are male, whereas the physician is female.
[2] We assume that the patient will not fall ill during the game, but only before. Hence, in case the patient returns to the physician, it may be traced back directly to the physician' false treatment choice.

Test 1: The physician decides whether to exert diagnostic effort k. In case she exerts effort, $k = 1$, otherwise $k = 0$.[3] Exerting diagnostic effort gives her a noisy signal $s \in \{0, 1\}$ on the patient's state of illness. Diagnostic effort comes at utility cost of $v > 0$ for the physician. We assume that the utilization of test 1 is hidden action and the generated signal is hidden information. Consequently, the payer cannot determine whether the physician took test 1 and in case she did, which signal the test resulted in. The informativeness of signal s depends on the sensitivity (se) and specificity (sp) of test 1. We follow Garcia Mariñoso and Jelovac (2003) and assume that $se = sp = e$ and $0.5 < e < 1$.[4] Thus, the test is indeed informative but imperfect. The test characteristics are depicted in Figure 3-1. Bayesian updating yields the following probabilities for the patient to suffer from the severe illness:

$$\begin{aligned} &\text{For signal } s = 0: \quad p_0^s = \frac{p(1-e)}{p(1-e)+(1-p)e} < p \\ &\text{For signal } s = 1: \quad p_1^s = \frac{pe}{pe+(1-p)(1-e)} > p \end{aligned} \tag{1}$$

Figure 3-1: Characteristics of Test 1



Note: This figure depicts the characteristic of test 1. p denotes the prevalence of the severe illness and $(1 - p)$ the prevalence of the mild illness in the population of patients who see a physician. e stands for the test's sensitivity and specificity. s is the imperfect private signal on the patient's state of illness, which is received by the physician after exerting diagnostic effort.

Test 2: Conditional on her decision regarding test 1 and, if applicable, conditional on the test's results, the physician decides on the sequential utilization of an expensive diagnostic test $g \in \{0, 1\}$. In case she does (not) utilize the test, $g = 1$ ($g = 0$). We assume that utilizing the expensive diagnostic test gives the physician a perfect signal $z \in \{0,1\}$ on the patient's state

[3] This test corresponds to the diagnostic test in Garcia Mariñoso and Jelovac (2003). While they allow for a continuous effort variable, for simplicity, we model diagnostic effort as a binary choice.
[4] This assumption simplifies the mathematical exposition but is not essential for the results of this paper.

of illness. Thus, $se = sp = 1$. Independent of her decision on test 1 and, if applicable, the resulting signal s, test 2 always gives her the following updated probabilities for the patient to suffer from the severe illness:

$$\begin{aligned} &\text{For signal } z = 0: \quad p_0^z = 0 < p_0^s < p \\ &\text{For signal } z = 1: \quad p_1^z = 1 > p_1^s > p \end{aligned} \tag{2}$$

The cost for the utilization of test 2 amount to c_M and exceed the physician's effort cost from exerting test 1 ($c_M > v$). We assume that the physician is not responsible for the cost of test 2. She refers the patient to receive test 2 at another provider who is equipped with the medical device to perform test 2. The other provider directly balances accounts with the payer. Hence, the physician acts as a gate-keeper for the expensive diagnostic test. Contrary to test 1, the utilization of test 2 as well as the resulting signal may be observed by the payer. Thus, test 2 is superior concerning the generated information on the patient's state of illness but is more costly for the payer.

Treatment choice: Conditional on the tests' results, the physician decides on treatment intensity $\tau \in \{T^A, T^B\}$. She may either choose a basic treatment T^B, for which cost are normalized to zero or she may choose an advanced treatment T^A. As compared to T^B, the additional cost for T^A amount to $h > 0$. The physician's objective is to maximize her expected utility, which is defined by her expected income from the payments she receives from the payer minus her expected cost. The latter correspond to the effort cost from test 1 (see equation (4) below).

3.2.3 The Payer

The payer's objective is to maximize expected social welfare. Therefore, he accounts for the expected overall cost of care provision as well as the monetary equivalent of the patient's expected utility from a diagnosis and treatment course. Expected overall cost of care arise from the physician's effort cost from exerting test 1, the additional cost for T^A compared to T^B , the cost for the utilization of test 2 and the payments to the physician. Apart from a possible health loss from delayed T^A, the patient's utility from a diagnosis and treatment course is normalized to zero. Thus, the payer's objective can be stated as minimizing the sum of the expected cost from providing optimal diagnosis and treatment to the patient and the monetary equivalent of the patient's expected health loss. Hereinafter, we refer to this as the expected social cost (ESC) as outlined in equation (3) below.

To put his objective into practice, the payer designs a payment contract (D, B, A, M) for the physician, which consists of four non-negative instruments. When the physician sees a patient, she is paid a capitation D for performing a diagnosis. Moreover, she receives FFS payments with a cost-based component for the respective treatment and mark-ups B and A conditional on adequate treatment choice. Treatment is presumed to be adequate if it cures the patient right away. Hence, in case the physician applies T^B and this turns out to be insufficient to cure the patient because he is severely ill, the physician receives no mark-up on the treatment cost since she has to apply T^A on top of T^B to cure the patient. As we show in Section 3.5, a scheme which would pay an unconditional mark-up for providing a service cannot create proper incentives. For T^B and T^A mark-ups B and A are paid, respectively. In case $B = A = 0$, the physician solely receives the cost-based FFS component for the respective treatment. In case $B, A > 0$, the physician additionally earns a mark-up on the respective treatment cost. Lastly, the physician has to pay a fee M in case she utilizes test 2, to possibly incentivize optimal diagnostic choice. Hence, the payer's objective function is defined by:

$$\min_{(D,B,A,M)} ESC = EC(v, h, c_M, D, B, A, M) + EL(\ell) \tag{3}$$

Given the payment contract (D, B, A, M), the physician will choose the diagnosis and treatment course that maximizes her expected income net of her expected cost:

$$\max_{(k,g,\tau)} EU = EI(D, B, A, M) - EC(v) \tag{4}$$

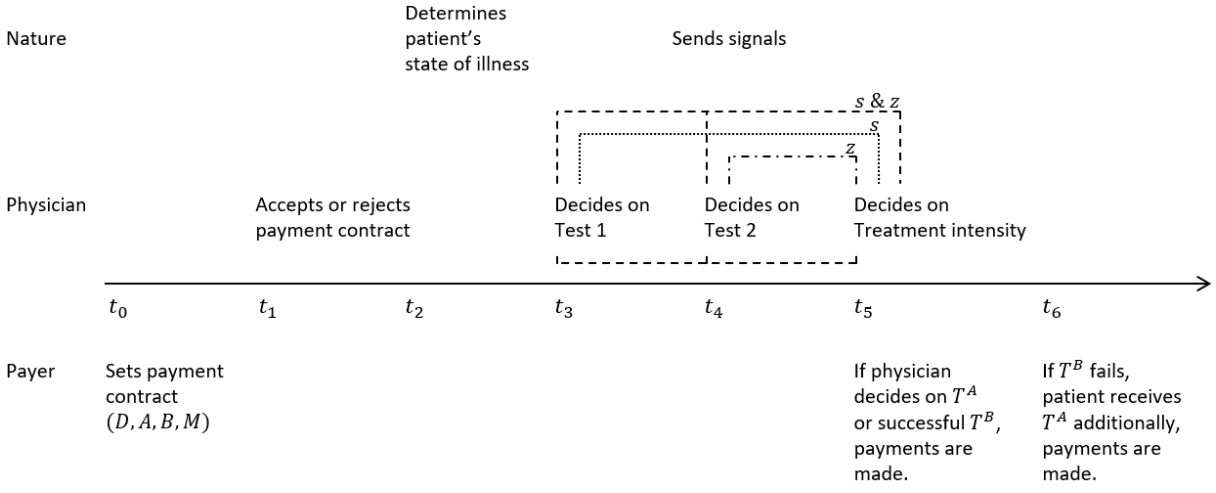
We assume that the payer respects a participation constraint (PC) in the sense that the physician's expected utility must be at least equal to her reservation utility ω. Here, ω captures that the physician has some negotiation power when signing a contract due to alternative income options. Furthermore, we require the capitation payment D to be non-negative. This assumption is motivated by the observation that payments by physicians to treat a patient do not exist. The most likely reason is that physicians will refuse contracts with $D < 0$ even if their PC is met.[5]

[5] Mougeot and Naegelen (2005) provide a similar argument in the hospital context.

3.2.4 The Sequence of Events

The sequence of events is depicted in Figure 3-2. At time t_0, the payer sets the physician's payment contract (D, B, A, M). The physician either accepts or rejects the contract (t_1). If the contract is rejected, the game ends. Otherwise, the game continues in t_2 and nature determines the patient's state of illness. In t_3, the physician sees the patient and decides on whether to utilize test 1 or not. In case the former applies, nature sends the physician an imperfect signal s on the patient's state of illness and in case the latter applies, no signal is sent. Subsequently, the physician decides on the utilization of test 2. If the physician orders test 2, nature sends a perfect signal z on the patient's illness, otherwise no signal is sent (t_4). Lastly, the physician decides on treatment intensity. In case the patient receives advanced treatment right away, or in case the patient initially receives basic treatment and is cured, payments are made and the game ends (t_5). In case basic treatment was not adequate, the patient additionally receives advanced treatment. The patient is cured, and payments are made (t_6).

Figure 3-2: The Sequence of Events



Note: This figure shows the sequence of events. The payment contract consists of a capitation D for performing a diagnosis, FFS payments with mark-ups B and A on the respective treatment cost of basic and advanced treatment as well as a fee M for the utilization of test 2. s is the imperfect signal on the patient's state of illness received by the physician after exerting test 1. z is the perfect, verifiable signal on the patient's state of illness after exerting test 2. T^B and T^A stand for basic and advanced treatment, respectively.

3.3 First-Best Solution: Optimal Diagnosis and Treatment Courses

First-best solutions are defined by diagnosis and treatment courses that minimize the expected social cost in the absence of incentive problems. Every patient receives at least T^B, for which cost are normalized to zero. Thus, the expected social cost are determined by the share of patients receiving T^A at additional cost h, the share of patients with a severe illness incurring a health loss due to delayed treatment with the monetary equivalent of ℓ as well as the shares of patients receiving test 1 and test 2 at cost v and c_M, respectively. We identify six optimal diagnosis and treatment courses (see Figure 3-3). All other courses are dominated by one of the six. For the complete set of courses see Figure 3-8 in Appendix A.

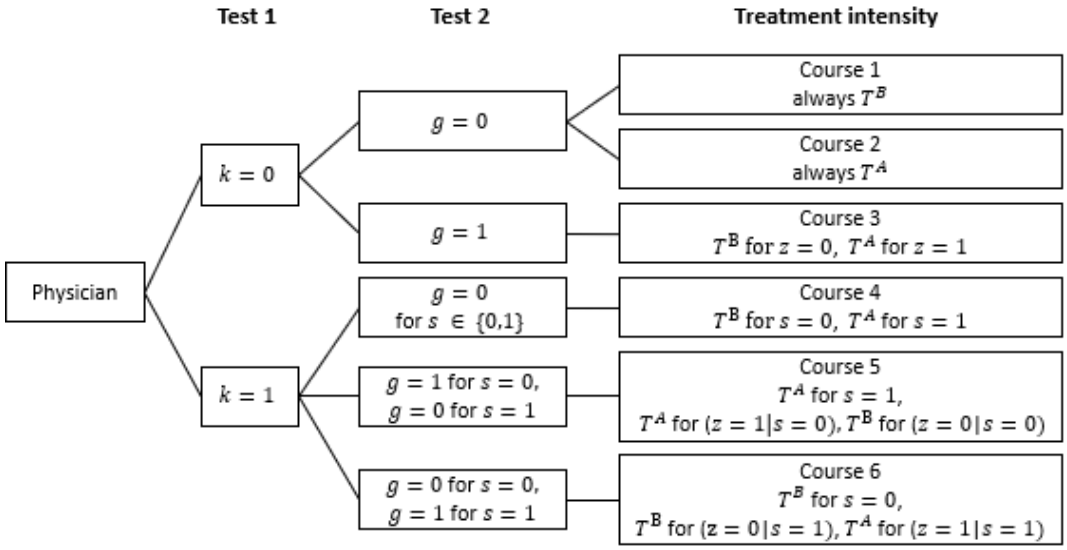
Courses without any tests:

- *Course 1: No test, T^B for all.* With probability $(1-p)$, the patient suffers from a mild illness and is cured with T^B at no cost. With probability p, however, the patient suffers from the severe illness. Because T^B is not adequate, the patient incurs a health loss ℓ and additionally needs T^A at a cost of h. This adds up to $ESC_1 = p(h+\ell)$.
- *Course 2: No test, T^A for all.* T^A cures all patients and leads to $ESC_2 = h$.

Courses with only one test:

- *Course 3: Test 2 only.* As the patient always receives test 2, cost c_M arise. Test 2 is assumed to be a perfect test and test results are followed. With probability p, the patient suffers from the severe illness and receives T^A at additional cost of h. This yields $ESC_3 = ph + c_M$.
- *Course 4: Test 1 only.* Because the patient always receives test 1, effort cost v incur. Test results are followed so that for signal $s = 0$, the patient receives T^B and for signal $s = 1$, he receives T^A. With probabilities e and $(1-e)$, respectively, the physician receives signal $s = 1$ for a severely (mildly) ill patient and therefore cost h incur from T^A. Moreover, with probability $(1-e)$ signal $s = 0$ is received for a severely ill patient and hence, T^B is not adequate. The patient suffers from health loss ℓ and additionally needs T^A at cost of h. This leads to $ESC_4 = v + peh + (1-p)(1-e)h + p(1-e)(h+\ell)$.

Figure 3-3: The First-Best Diagnosis and Treatment Courses



Note: The figure depicts the potential candidates for a first-best diagnosis and treatment course. For each candidate, the physician's choices are outlined. k denotes whether the physician exerts test 1. $k = 1$ ($k = 0$) implies that the physician does (not) use the test. s is the private imperfect signal on the patient's state of illness received by the physician after exerting test 1, whereby $s = 0$ ($s = 1$) indicates a mild (severe) illness. g denotes whether the physician uses test 2. $g = 1$ ($g = 0$) implies that the physician does (not) use the test. z is the verifiable signal on the patient's state of illness after exerting test 2, whereby $z = 0$ ($z = 1$) indicates a mild (severe) illness with certainty. T^B and T^A stand for basic and advanced treatment, respectively.

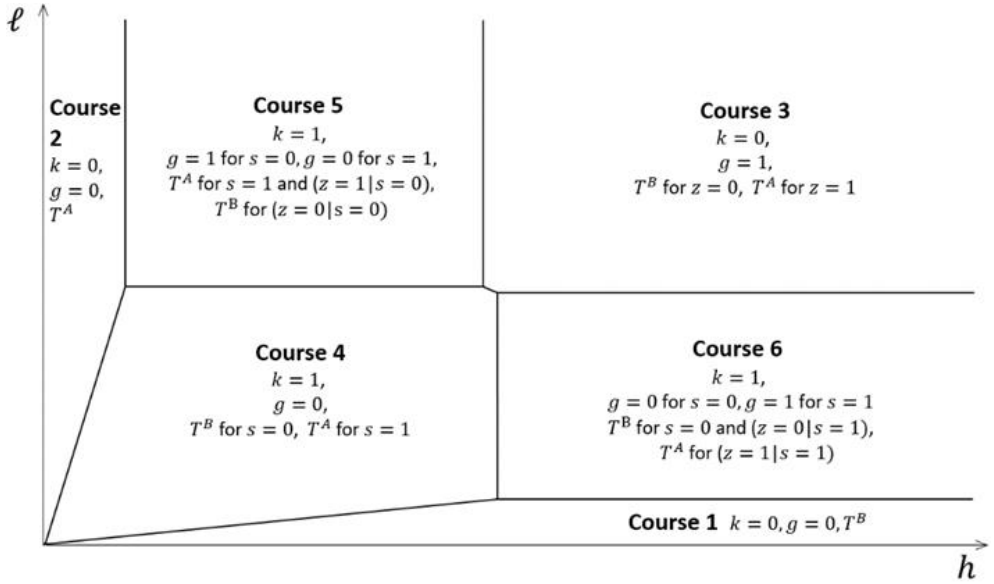
Courses with test 2 conditional on the signal from test 1 (composite courses):

- *Course 5: Disjunctive positivity criterion.* Only if the results of both tests are negative (indicate a mild illness), the patient receives T^B, otherwise T^A is applied. Since the patient always receives test 1, effort cost v arise. With probability $p(1-e) + (1-p)e$, signal $s = 0$ is received and test 2 is additionally utilized to ensure that following the signal and, thus, applying T^B is indeed adequate. This incurs cost c_M. Following the signal from test 2, with probability $p(1-e)$, the patient receives T^A and causes cost h and with probability $(1-p)e$, he receives T^B and is cured. With probability $pe + (1-p)(1-e)$, signal $s = 1$ is received and, therefore, following the signal, the patient is cured with T^A at cost h. This implies $ESC_5 = v + [1-(1-p)e]h + [p(1-e) + (1-p)e]c_M$.
- *Course 6: Conjunctive positivity criterion.* Only if the results of both tests are positive (indicate a severe illness), the patient receives T^A, otherwise T^B is applied. As the physician always utilizes test 1, effort cost v arise. With probability $p(1-e) + (1-p)e$, signal $s = 0$ is received and treatment T^B is applied. With probability $(1-p)e$ the patient suffers from the mild illness, is cured with T^B and causes no further cost. With probability

$p(1-e)$, however, signal $s=0$ is falsely received for a severely ill patient. Since T^B does not cure him, he additionally needs T^A and suffers a health loss from treatment delay. With probability $pe+(1-p)(1-e)$, the patient receives signal $s=1$ and, therefore, test 2 is additionally utilized to ensure that T^A is necessary, which incurs cost c_M. With probability pe, he has the severe illness and gets T^A. With probability $(1-p)(1-e)$, he receives T^B and is cured. Thus, $ESC_6 = v + p(1-e)(h+l) + peh + [pe+(1-p)(1-e)]c_M$.

In case Courses 1 to 4 are followed, test utilization is identical for all patients, in fact either one of the two tests or none is applied. In Courses 5 and 6, however, test utilization is not the same for all patients; conditional on the respective outcome of test 1, patients only selectively receive test 2. Figure 3-4 displays the allocation of first-best diagnosis and treatment courses depending on the subsequent cost of the physician's treatment choice, i.e. for values of the additional cost h for T^A and for the monetary equivalent of the patient's expected utility loss ℓ from delayed T^A. The proof is in Appendix A.

Figure 3-4: The First-Best Solution



Note: The figure shows the allocation of the first-best diagnosis and treatment courses in the $h-\ell$ space. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. k denotes whether the physician exerts test 1. $k=1$ ($k=0$) implies that the physician does (not) use the test. s is the private imperfect signal on the patient's state of illness received by the physician after exerting test 1, whereby $s=0$ ($s=1$) indicates a mild (severe) illness. g denotes whether the physician uses test 2. $g=1$ ($g=0$) implies that the physician does (not) use the test. z is the verifiable signal on the patient's state of illness after exerting test 2, whereby $z=0$ ($z=1$) indicates a mild (severe) illness with certainty. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of $p=0.5$, $e=0.85$, $v=2$, $c_M=200$ and $\omega=0$.[6]

[6] Laslett et al. (2005), for instance, find that a physical examination for diagnosis of sacroiliac joint pain as a source of lower back pain results in a sensitivity and specificity of 0.94 and 0.78, respectively. This motivates our choice of $e=0.85$ for illustration.

We obtain the following results:

- Course 1 tends to be optimal across values for h and for low values of ℓ. The health loss ℓ is small and T^A is relatively expensive. Thus, it is optimal to first give T^B a try, without any preceding tests.
- Course 2, by contrast to Course 1, tends to be optimal for low values of h and across values for ℓ. T^A's additional cost are small, while the negative effects of delayed treatment are relatively high. Thus, it is optimal to apply T^A right away.
- Course 3 tends to be optimal for high values of h and ℓ. Since the stakes are high, it is worthwhile to employ the more expensive but superior test 2.
- Course 4 tends to be optimal for intermediate values of h and ℓ. As the stakes are lower than in the previous course, utilizing test 1 allows to take an informed treatment choice at comparatively little cost.
- Course 5 tends to be optimal for intermediate values of h and high values of ℓ. Since the health loss is high, it is optimal to ensure that no patient suffers from delayed treatment and that if T^B is applied, it is adequate. However, T^A is fairly costly and, thus, not all patients should receive T^A. The cost-effective approach is to treat patients with signal $s = 1$ from test 1 with T^A right away and to additionally apply test 2 to those with test result $s = 0$, to ensure that T^B is adequate. Compared to Course 3, T^A is employed more frequently but testing cost are lower.
- Course 6, by contrast to Course 5, tends to be optimal for high values of h and intermediate values of ℓ. Because T^A is costly, here the emphasis is put on avoiding unnecessary high-cost treatment. But, since the health loss is also fairly high, not all patients simply receive T^B. It is cost-effective to apply T^A only to those patients for whom both tests indicate the severe illness.

Note that Courses 3 to 6 require that both T^B and T^A are used conditional on the test results. Testing is important to determine who should receive which treatment and, therefore, contributes to a cost-effective treatment choice.

3.4 Implementation of the First-Best Courses

In this section, we consider the implementation of the first-best courses in the context of asymmetric information. We assume that the utilization of test 1 is hidden action and the generated signal is hidden information. The implementation of the courses, for which test 1 is not required, is straightforward. For Courses 1 and 2, paying zero mark-ups implements the first-best. The physician is indifferent between treatments and will opt for the first-best. As no mark-ups are paid, there is no incentive to perform test 1. Course 3 can be implemented by requiring the physician to always utilize test 2 and to act according to the test results. Again, the first-best solution can be implemented without paying any mark-ups.

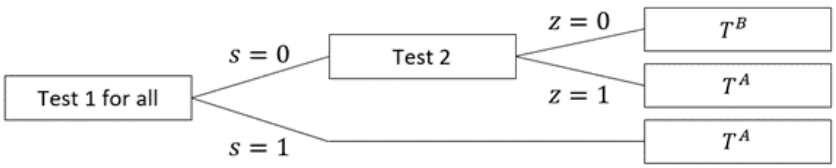
The implementation of Course 4 is more demanding. As Garcia Mariñoso and Jelovac (2003) have shown, positive mark-ups are necessary to induce the physician to undertake test 1. The optimal payment scheme requires $B > A > 0$. Furthermore, physicians may earn an information rent. The reader is referred to Appendix B for the details concerning the implementation of Courses 1 to 4.

Our focus is on the implementation of the composite courses. In these courses, test 2 should be used conditional on the results of test 1. However, if test 2 is freely available, overuse of this test can be expected. Therefore, incentives must be provided such that (i) test 1 is always performed by the physician and (ii) test 2 is utilized conditional on the results of test 1.

3.4.1 Implementation of Course 5 "Disjunctive Positivity Criterion"

Course 5 initially requires the utilization of test 1. Patients with signal $s = 1$ from test 1 should always receive the advanced treatment right away. For those with signal $s = 0$, test 2 should be applied for further investigation. Patients with signal $(z = 0|s = 0)$ should receive the basic and patients with signal $(z = 1|s = 0)$ the advanced treatment (see Figure 3-5).

Figure 3-5: Characteristics of Course 5



Note: The figure depicts the physician's diagnostic and treatment choices in Course 5. s is the private imperfect signal on the patient's state of illness received by the physician after exerting test 1, whereby $s = 0$ ($s = 1$) indicates a mild (severe) illness. z is the verifiable signal on the patient's state of illness after exerting test 2, whereby $z = 0$ ($z = 1$) indicates a mild (severe) illness with certainty. T^B and T^A stand for basic and advanced treatment, respectively.

The proposed payment contract (D, B, A, M) needs to align the physician's privately optimal choices with the socially optimal choices. This is achieved by anticipating the physician's diagnostic and treatment decisions at each stage of the game, conditional on the preceding pattern of her decisions, and by inciting her to take the socially optimal decisions at each stage of the game. Starting with the physician's last choice on treatment intensity, the game is solved by backwards induction.[7]

Inciting the optimal treatment intensity: The physician takes the treatment decision conditional on the signals from the two preceding diagnostic tests by comparing the expected additional payoffs from T^A and T^B. Because she must not bill both treatments, she will choose the more profitable one in expected terms. In case the physician uses test 2, the payer can just dictate to the physician to always follow the test's signals as these are verifiable.[8] If the physician receives signal $z = 0$ from test 2, she has to apply T^B. If signal $z = 1$ results, she has to apply T^A. By contrast to test 2, the diagnostic signals from test 1 are private information to the physician. Turning to the situation in which the physician only takes the pretest and receives signal $s = 1$, the expected payoff is A for choosing T^A which always heals the patient and $(1 - p_1^s)B$ for choosing T^B (see equation (1) for the definitions of p_1^s and p_0^s). Incentive compatibility requires:

$$A \geq (1 - p_1^s)B \tag{5}$$

Inciting the optimal utilization of test 2: The physician takes the decision concerning the utilization of test 2 conditional on the resulting signals from the preceding diagnostic test 1, yielding the updated probabilities $\tilde{p}^s \in \{p_0^s, p_1^s\}$. She compares the additional payoffs from either utilizing test 2 or not. In case test 2 is used, with probability $(1 - \tilde{p}^s)$, the outcome is that patients are mildly ill. For mildly ill patients the physician has to apply T^B and receives payoff B. For the share of severely ill patients, $\tilde{p}^s$, she has to apply T^A, which leads to payment A. For the utilization of test 2 the physician always has to pay the fee M. Thus, utilizing test 2 yields an additional payoff of $\tilde{p}^s A + (1 - \tilde{p}^s)B - M$, depending on the outcome of test 1. In case test 2 is not utilized, she chooses the treatment that yields the higher expected payoff.

The expected additional payoff is $\max\{A, (1 - \tilde{p}^s)B\}$. Thus, we obtain the following incentive constraints (ICs):

[7] Hereinafter, we assume that with indifference, the physician will opt for the first-best. Yet, the results remain unchanged if we instead assumed that with indifference the physician would decide on behalf of the patient.
[8] The results remain robust to a change in the verifiability of the signals from test 2.

$$p_0^s A + (1 - p_0^s)B - M \geq \max\{A, (1 - p_0^s)B\} \quad (6)$$

$$p_1^s A + (1 - p_1^s)B - M \leq \max\{A, (1 - p_1^s)B\} \quad (7)$$

Taking into consideration (5) and noting that A may be greater or smaller than $(1 - p_0^s)B$, (6) and (7) may be rewritten as:

$$(1 - p_0^s)(B - A) - M \geq 0 \quad (8)$$

$$p_0^s A - M \geq 0 \quad (9)$$

$$(1 - p_1^s)(A - B) + M \geq 0 \quad (10)$$

Inciting the optimal utilization of test 1: The physician takes the decision on the utilization of test 1 independent of any preceding events. In case she exerts effort, and the above ICs are met, Course 5 yields an expected utility of

$$EU_5 = D + [p + (1 - p)(1 - e)]A + (1 - p)eB - [p(1 - e) + (1 - p)e]M - v \quad (11)$$

For each patient, the physician sees, she is paid capitation D. All patients with signal $s = 1$ resulting from test 1 receive T^A right away. Thus, she receives mark-up A for patient share $pe + (1 - p)(1 - e)$. Severely ill patients with signal $s = 0$ receive T^A. Thus, for patient share $p(1 - e)$ she is also paid mark-up A. Patients with a mild illness with signal $s = 0$ receive T^B. Hence, for patient share $(1 - p)e$ she receives mark-up B. The physician must bear cost M for patient share $[p(1 - e) + (1 - p)e]$ because all patients with signal $s = 0$ resulting from test 1 additionally receive test 2. As the physician always utilizes test 1, she incurs effort cost v.

In case she does not utilize test 1, she may either treat all patients with T^B first (Course 1), treat all patients with T^A right away (Course 2) or comply to test 2 for treatment choice (Course 3). Her respective expected utilities amount to:

$$EU_1 = D + (1 - p)B \quad (12)$$

$$EU_2 = D + A \quad (13)$$

$$EU_3 = D + pA + (1 - p)B - M \quad (14)$$

Course 5 requires that the physician always utilizes test 1. This is also in her private best interest, if and only if contract (B, A, M) fulfills the condition $EU_5 \geq \max\{EU_1, EU_2, EU_3\}$. Plugging in for EU_1, EU_2, EU_3 and EU_5 as well as regrouping yields the following three conditions:

$$[p + (1-p)(1-e)]A + (1-p)(e-1)B - [p(1-e) + (1-p)e]M \geq v \quad (15)$$

$$-(1-p)eA + (1-p)eB - [p(1-e) + (1-p)e]M \geq v \quad (16)$$

$$(1-p)(1-e)A + (1-p)(e-1)B + [pe + (1-p)(1-e)]M \geq v \quad (17)$$

Using the definitions of p_0^S and p_1^S (see equation (1)) to express everything in terms of p and e, conditions (5), (8), (9) and (10) may be rewritten and regrouped as:

$$[pe + (1-p)(1-e)]A + (1-p)(e-1)B \geq 0 \quad (18)$$

$$-(1-p)eA + (1-p)eB - [p(1-e) + (1-p)e]M \geq 0 \quad (19)$$

$$p(1-e)A - [p(1-e) + (1-p)e]M \geq 0 \quad (20)$$

$$(1-p)(1-e)A + (1-p)(e-1)B + [pe + (1-p)(1-e)]M \geq 0 \quad (21)$$

We obtain that (16) implies (19) and (17) implies (21).

Now, we can determine the least costly combination of payments (D, B, A, M) that incites the physician to take Course 5. Note that the capitation payment D has no incentive effect. Furthermore, low values of D should be avoided because physicians may refuse to accept negative capitation payments. Therefore, we focus on values (B, A, M) which minimize the expected payments to the physician s.t. the incentive-compatibility constraints. The payer will set D just only as high as necessary to make sure that the physician's participation constraint $EU_5 \geq \omega$ is met, possibly constrained by $D \geq 0$.[9] Minimizing the sum of the expected payments to the physician yields the following problem:

$$\min_{(A,B,M)} \{[p + (1-p)(1-e)]A + (1-p)eB - [p(1-e) + (1-p)e]M\}$$

subject to (15), (16), (17), (18) and (20).

The ICs ensure that it is in the physician's best interest to choose the socially optimal diagnosis and treatment course. (15) to (17) are effort constraints and ensure that test 1 is always utilized

[9] This constraint can lead to an information rent for physicians. We consider this problem in Section 3.6.

in the first place. IC (18) is a treatment choice constraint and ensures that all patients with signal $s = 1$ resulting from test 1 receive T^A. IC (20) is a constraint on the use of the expensive diagnostic test and ensures that only for patients with signal $s = 0$ resulting from test 1, test 2 is utilized additionally. Thus, only selected patients are tested with test 2.

In order to simplify the optimization problem, we check whether one or several of the ICs have to be binding. Since increasing B furthers IC (16), but counters all other ICs, (16) has to binding at the optimum. Furthermore, we find that increasing M furthers IC (17), but counters all other ICs, therefore (17) also has to be binding at the optimum. Lemma 1 states our findings.

Lemma 1: *In Course 5, effort constraints (16) and (17) are binding at the optimum.*

Thus, we obtain:

$$-(1-p)eA + (1-p)eB = v + [p(1-e) + (1-p)e]M$$

$$-(1-p)(1-e)A + (1-p)(1-e)B = -v + [pe + (1-p)(1-e)]M$$

Adding up the equations and simplifying yields $M = (1-p)(B-A)$. Inserting M for $(1-p)(B-A)$ into either equation (16) or (17) results in:

$$M = \frac{v}{p(2e-1)} \tag{22}$$

The implementation of Course 5 requires a strictly positive utilization fee M for the utilization of test 2. This is necessary to counter the incentive of the physician to replace test 1 by test 2. For this reason, the size of M depends on parameters related to test 1. M increases in the effort cost v for test 1. M decreases in e because once test 1 gets more informative, the incentives to replace test 1 with test 2 or to skip both tests diminish. Furthermore, M decreases in p. Intuitively, a payment contract which is to implement a diagnosis and treatment course that entails the use of T^B only if test 2 turns out to be negative has to account for the fact that this outcome is less likely, the higher the share of severely ill patients. Thus, the incentive to use test 2 is decreasing in p.

Having solved for M, we turn to the solution for mark-up payments A and B. With a strictly positive M, a necessary condition for ICs (15) and (20) to hold, is $A > 0$. Furthermore, from Lemma 1, we know that ICs (16) and (17) are binding. A necessary condition for (16) and (17) to bind is $B > A$.

Lemma 2: *Setting $B > A > 0$ is a necessary condition to provide optimal incentives for the implementation of Course 5.*

Clearly, if $B \leq A$, there would be no incentive to do any test as the physician would always earn at least as much by providing treatment T^A instead of treatment T^B. Furthermore, A must be strictly positive to reward the physician for utilizing test 2 for patients with signal $s = 0$.

We consider the remaining ICs (15), (18) and (20) to solve for A and B. Rewriting $(1-p)(B-A) = M$ yields

$$B = A + \frac{M}{1-p} \tag{23}$$

Inserting (23) into ICs (15), (18) and (20) and regrouping gives:

$$pA - [(1-e) + p(1-e) + (1-p)e]M \geq v \tag{24}$$

$$peA - (1-e)M \geq 0 \tag{25}$$

$$p(1-e)A - [p(1-e) + (1-p)e]M \geq 0 \tag{26}$$

We are looking for the lowest value of A fulfilling these incentive constraints. We find that (26) has to be binding, since, in order to hold, it requires a strictly larger A than ICs (24) and (25). IC (26) ensures that the physician is indifferent towards the utilization of test 2 for patients with signal $s = 0$. We conclude with

Lemma 3: *In Course 5, the constraint on the use of the expensive diagnostic test, IC (26), is binding if payment A is to be chosen as low as possible in an incentive-compatible solution.*

Proof. See Appendix C.

Since IC (26) is binding, we can rewrite the IC and insert for M from (22). This results in

$$A = \frac{v(p(1-e)+(1-p)e)}{p^2(1-e)(2e-1)} \tag{27}$$

Inserting for A and M in (23) yields

$$B = \frac{v\left(e+\left(p^2-2p\right)(2e-1)\right)}{p^2(1-p)(1-e)(2e-1)} \tag{28}$$

We find that mark-up A is strictly larger than M. Both mark-ups, A and B, are strictly increasing in effort cost v. The higher the physician's effort cost for test 1, the higher have to be her gains from the mark-up payments at the optimum, otherwise she will skip test 1. Moreover, A strictly decreases in p. B, by contrast, is non-monotonous in p. Both mark-up payments are also non-monotonous in e. Proposition 1 summarizes our findings.

Proposition 1: *A mixed payment contract provides incentives for the implementation of a diagnosis and treatment course that applies the disjunctive positivity criterion. This entails the selective utilization of an expensive diagnostic test only for patients who received a negative/ reassuring signal* $(s = 0)$ *from a pretest. If the capitation payment* D *is set as low as possible to fulfill the physician's participation constraint, the contract consists of FFS payments with mark-ups* $B > A > 0$ *on the respective treatment cost of* T^B *and* T^A*, conditional on adequate treatment choice, as well as a utilization fee* $M > 0$ *for the utilization of the expensive subsequent test. The optimal payment instruments are given by:*

$$A = \frac{v(p(1-e)+(1-p)e)}{p^2(1-e)(2e-1)}, \qquad B = \frac{v\left(e+(p^2-2p)(2e-1)\right)}{p^2(1-p)(1-e)(2e-1)} \quad \textit{and} \quad M = \frac{v}{p(2e-1)}.$$

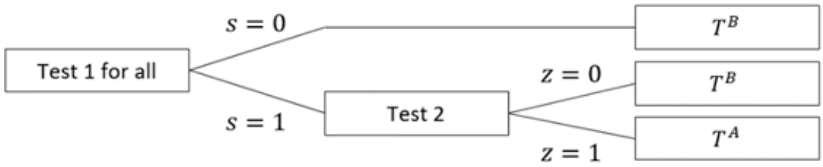
A potential problem is that even if mark-ups A and B and the utilization fee M are chosen as in Proposition 1, the physician's participation constraint may require a negative capitation payment D. If a negative capitation payment cannot be enforced, the physician will earn an information rent. This is analyzed in Section 3.6. Next, we consider the implementation of Course 6.

3.4.2 Implementation of Course 6 "Conjunctive Positivity Criterion"

As in Course 5, Course 6 initially requires the utilization of test 1. However, by contrast to Course 5, patients with signal $s = 0$ resulting from test 1 should always receive basic treatment right away and for patients with signal $s = 1$, test 2 should additionally be applied. Patients with signal $(z = 0|s = 1)$ should also receive basic and only patients with signal $(z = 1|s = 1)$ the advanced treatment (see Figure 3-6).

The payer's objective is to determine the least costly combination of payments (D, B, A, M) that incites the physician to take Course 6, s.t. incentive-compatibility constraints. As before, the capitation payment D has no incentive effect and will be set only just as high as necessary to make sure that the physician's participation constraint $EU_6 \geq \omega$ is fulfilled, possibly constrained by $D \geq 0$.

Figure 3-6: Characteristics of Course 6



Note: The figure depicts the physician's diagnostic and treatment choices in Course 6. s is the private imperfect signal on the patient's state of illness received by the physician after exerting test 1, whereby $s = 0$ ($s = 1$) indicates a mild (severe) illness. z is the verifiable signal on the patient's state of illness after exerting test 2, whereby $z = 0$ ($z = 1$) indicates a mild (severe) illness with certainty. T^B and T^A stand for basic and advanced treatment, respectively.

The incentive-compatibility constraints for Course 6 are derived analogous to Course 5. Therefore, the reader is referred to Appendix D for the derivation of ICs (29) to (33) .

For inciting the optimal treatment intensity, we get:

$$-[p(1-e)+(1-p)e]A+(1-p)eB \geq 0 \tag{29}$$

The optimal utilization of test 2 calls for:

$$(1-p)(e-1)A+(1-p)(1-e)B+[(1-p)(e-1)-pe]M \geq 0 \tag{30}$$

Finally, the optimal utilization of test 1 yields the following constraints:

$$peA+[(1-p)(e-1)-pe]M \geq v \tag{31}$$

$$(pe-1)A+(1-p)B+[(1-p)(e-1)-pe]M \geq v \tag{32}$$

$$p(e-1)A+[p(1-e)+(1-p)e]M \geq v \tag{33}$$

As above, we focus on values (B, A, M), which minimize the expected payments to the physician. This yields the following problem:

$$\min_{(A,B,M)} \{peA+(1-p)B-[pe+(1-p)(1-e)]M\}$$

subject to (29), (30), (31), (32) and (33).

The problem is solved analogously to the previous section and, thus, delegated to Appendix D. We derive Proposition 2.

Proposition 2: *In the second-best case, a mixed payment contract provides optimal incentives for the implementation of a diagnosis and treatment course that applies the conjunctive positivity criterion. This entails the selective utilization of an expensive diagnostic test only for patients who received a positive/alarming signal* $(s = 1)$ *from a pretest. If the capitation payment* D *is set as low as possible to fulfill the physician's participation constraint, the contract consists of FFS payments with mark-ups* $B > A > 0$ *on the respective treatment cost of* T^B *and* T^A*, conditional on adequate treatment choice, as well as a utilization fee* $M > 0$ *for the utilization of the expensive subsequent test. The optimal payment instruments are given by*

$$A = \frac{v}{p(1-p)(2e-1)}, \qquad B = \frac{v\left(1-e+p^2(2e-1)\right)}{p(1-p)^2(1-e)(2e-1)} \qquad \textit{and} \qquad M = \frac{v}{(1-p)(2e-1)}.$$

Analogous to the implementation of Course 5, a positive utilization fee M is part of the optimal solution. If M was not strictly positive, the selective utilization of test 2 would not be feasible. Based on the same reasoning as outlined above, M strictly increases in effort cost v and decreases in the informativeness e of test 1. Yet, in contrast to Course 5, M strictly increases in the prevalence of the severe illness p since the physician's gain from utilizing test 2 increases in p. This is because, in Course 6, the physician should not use test 2 in case test 1 yields signal $s = 0$. Intuitively, a payment scheme which aims at providing incentives to use T^A only if test 2 is positive, has to account for the fact that this outcome is more likely, the higher the share of severely ill patients. Hence, the physician's incentive to use test 2 is increasing in p.

As before, we find that mark-up B has to be strictly larger than mark-up A. This ensures that test 1 is always used in the first place and that all patients with signal $s = 0$ from test 1 receive T^B. If B was not strictly larger than A, a differentiated treatment choice would be infeasible. The physician would skip both tests and all patients would receive T^A right away. Furthermore, A must be positive to reward the physician for utilizing test 2 for patients with signal $s = 1$.

A potential problem, as for Course 5, is that even if mark-ups A and B are chosen as low as possible and M is chosen as high as possible, the physician's participation constraint may require a negative capitation payment D. If this cannot be enforced, the physician will earn an information rent if Courses 5 and 6, respectively, are to be implemented. This is analyzed in Section 3.6. Before, we consider an alternative payment contract.

3.5 Why Unconditional FFS Payments Cannot Implement the First-Best Solution

In numerous OECD countries, such as Australia, Germany, Switzerland, the Netherlands and the United States, physicians delivering primary and specialist outpatient care are predominantly reimbursed by a combination of capitation and FFS payments, while average earnings from the latter clearly outweigh those from the former (see Mossialos et al. 2017 for an international survey). Our proposed payment contracts significantly differ from these common payment contracts in the sense that FFS mark-up payments A and B need to be conditional on adequate treatment choice. Only if the chosen treatment is adequate, in the sense that the treatment intensity is sufficient to cure the patient right away, mark-up A or B is paid. Furthermore, the physician must bear cost for the utilization test 2. If one relaxed one of these restricting assumptions, the implementation of Courses 5 and 6 would be infeasible.

The latter result is evident from Propositions 1 and 2, which established that the utilization fee M must be strictly positive to implement the first-best solution. The former result is based on the recognition that unconditional mark-up payments allow the physician to follow a trial-and-error strategy, which does not punish her for inadequate treatment choice. To ensure incentive-compatibility, the constraints on the physician's treatment choice after performing test 1 have to be adjusted. These adjusted ICs are not compatible with the constraints on the utilization of tests 1 and 2 though. We establish the following proposition:

Proposition 3: *A mixed contract, consisting of unconditional FFS mark-up payments, always fails to provide optimal incentives for the implementation of Courses 5 and 6.*

Proof. See Appendix E.

For Course 5, the intuition for Proposition 3 is as follows: to incentivize the physician to choose T^A right away for patients with signal $s = 1$ from test 1, the mark-up for T^A must be larger than the mark-up for T^B. Yet, in that case, there cannot be any incentive to utilize test 2 only for patients with signal $s = 0$ as this requires that the physician has a higher reimbursement in case the patient turns out to be mildly ill. For Course 6, the mark-up for T^B must be sufficiently high compared to the mark-up for T^A to incentivize the physician to choose T^B right away for patients with signal $s = 0$ from test 1. With these strong incentives to provide T^B, however, incentives to utilize test 2 only for patients with signal $s = 1$ cannot be established. It would be more attractive for the physician to follow the trial-and-error strategy.

3.6 Implementation and Information Rents

In principal-agent situations, in which the agent cannot be punished because the principal's instruments are limited to reward payments only, the principal may have to grant the agent some information rent. This gives rise to a rent-efficiency trade-off (Laffont and Martimort 2002, pp. 155–157). A similar issue arises in our model. The physician has to be incentivized to always utilize test 1 first and, based on the private signals from test 1, to only selectively utilize test 2 and to take the optimal treatment choice. The mark-up payments, as outlined above, serve as rewards, while the utilization fee M punishes the physician for the utilization of test 2. However, M may not be sufficient to completely extract the physician's expected gain from the rewards. Furthermore, the capitation payment cannot be negative. Thus, the physician's participation constraint $EU \geq \omega$ may not bind and she will earn an information rent.

3.6.1 Cost of Implementation for Course 5

Inserting the optimal values (A, B, M) from Proposition 1 into (11) and simplifying yields the physician's expected utility from adhering to Course 5:

$$EU_5 = D + v\frac{p(1-e)+e(1-p)}{p^2(1-e)(2e-1)} > D \tag{34}$$

Depending on the size of the second summand, implementation may require paying an information rent to the physician. If it is smaller or equal to reservation wage ω, Course 5 may be implemented without information rent. If, however, it exceeds ω, even setting $D = 0$ is not sufficient to extract all remaining rents. Taking the partial derivatives with respect to parameters (v, p, e) for a zero-capitation based on (34) shows that the information rent strictly increases in the effort cost, strictly decreases in the prevalence of the severe illness p and is non-monotonous in the informativeness of test 1.

3.6.2 Cost of Implementation for Course 6

We proceed in the same manner. Inserting the optimal values (A, B, M) from Proposition 2 into the physician's expected utility from adhering to Course 6 (see equation (49) in Appendix D) and simplifying gives:

$$EU_6 = D + v\frac{1-e+p^2(2e-1)}{p(1-p)(1-e)(2e-1)} > D \tag{35}$$

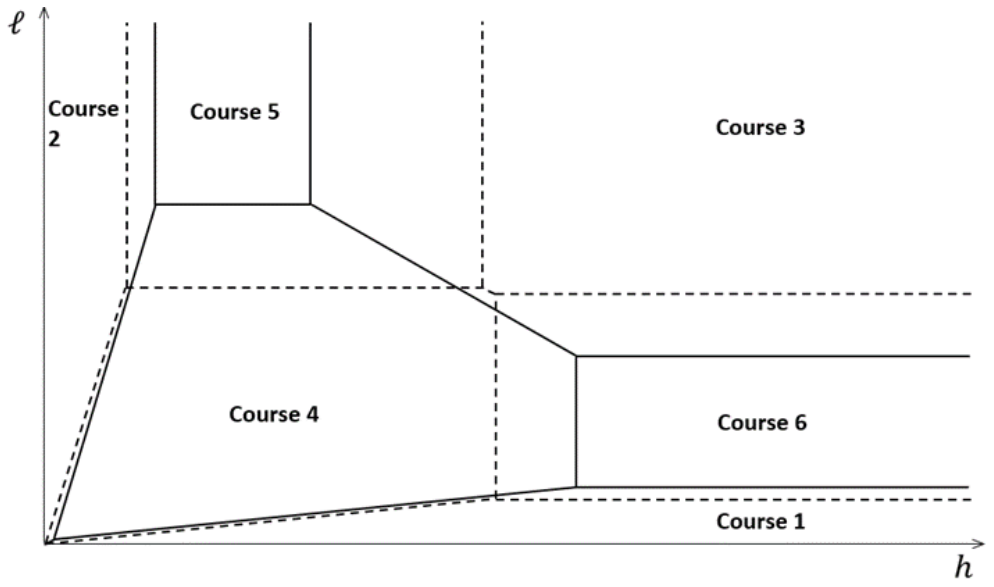
Again, the size of the second summand determines whether implementation may require paying an information rent to the physician. If it is smaller or equal to ω, Course 6 may be implemented

without information rent. However, if it exceeds ω, even setting $D = 0$ is not sufficient to extract all remaining rents. Analyzing the second term reveals that the potential information rent increases in effort cost v and is non-monotonous in parameters (p, e).

3.6.3 Implementation with Information Rents

When determining the second-best solution, potential information rents need to be accounted for. Besides Courses 5 and 6, the implementation of Course 4 may also require paying an information rent to the physician. Implementing Courses 1 to 3, by contrast, is costless to the payer (see Appendix B). Thus, we can expect that, in the second-best situation, the parameter space for which Courses 4 to 6 are optimal is smaller than in the first-best. Figure *3-7* depicts the second-best solution for the same set of parameters that were chosen for the presentation of the first-best solution in Figure 3-4 above.

Figure 3-7: The Second-Best Solution



Note: The figure shows the allocation of the second-best diagnosis and treatment courses in the $h - \ell$ space. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. The dashed lines indicate the first-best solution (compare Figure 3-4). The figure is exemplary for values of $p = 0.5$, $e = 0.85$, $v = 2$, $c_M = 200$ and $\omega = 0$.

We find that each of the six courses may also be second-best optimal. For reasons of illustration, we set the physician's reservation utility ω equal to zero. Now, even we assumed that $D = 0$ and, therefore, extracted as much rent as possible, the information rents that must be paid to implement Courses 4, 5 and 6 would still remain strictly positive. Consequently, the expected

social cost for Courses 4, 5 and 6 increase and Courses 1, 2 and 3 become second-best optimal for regions, for which Courses 4, 5 and 6 were found to be first-best optimal. If v increased, the regions for which Courses 4, 5 and 6 are second-best optimal would shrink even further because the information rent strictly increases in v in all three courses.

3.7 Discussion

This paper determined contracts that implement optimal sequential diagnostic and treatment choice in the presence of asymmetric information. We show that it is not always socially optimal to utilize a perfectly informative, but also more expensive diagnostic test for all patients right away. Particularly, there are circumstances in which such a test should only be used conditional on the diagnostic signal from a less costly, though imperfect pretest. Whether a selective utilization is optimal or not depends not only on the cost of the expensive device but also on the consequences for treatment choice. This includes the monetary equivalent of the patient's health loss from delayed adequate treatment and the cost difference between available treatments. We focused on two situations. In the first case, physicians should apply the disjunctive positivity criterion, which indicates that they should provide a basic treatment if and only if both tests show that the patient is not severely ill, otherwise an advanced treatment is provided. In the second case, physicians should apply the conjunctive positivity criterion, which calls for the use of an advanced treatment if and only if both tests show that the patient is severely ill, otherwise a basic treatment is used.

In the presence of asymmetric information concerning the physician's utilization of the pretest and her adherence to private diagnostic signals, we find that designing a mixed payment contract may incentivize physicians to follow these courses. Hence, our work supports the ongoing debate about shifting away from contracts dominated by unconditional FFS payments and instead implementing mixed payment contracts to encourage efficient decision making. In this paper, we determined an optimal payment system consisting of three components: The first component consists of a capitation for performing a diagnosis. The second component comprises conditional and differentiated FFS payments with strictly positive mark-ups. More specifically, to support cost-effective treatment choice, the mark-up for the less costly treatment must be higher than the mark-up for the more costly treatment. The third component is a fee for the utilization of the expensive diagnostic test. It should be emphasized that the physician has to share responsibility for the cost of expensive test, otherwise overuse can be expected. Implementation with our proposed payment contracts may come at a cost though. The FFS

mark-up payments must reach a certain level to initially incite diagnostic effort. Hence, the physician extracts an information rent if her earnings from treatment exceed her reservation utility and the capitation payment must not be negative. These implementation cost may render the selective utilization of the expensive device suboptimal.

Although our study makes several important contributions to existing research, it is not without limitations. In our model, we have made several simplifying assumptions. To begin with, physicians were assumed to be concerned only with their own income. If we instead supposed that they at least partially internalize the patient's health loss, the selective utilization of an expensive test could still be first-best, and the structure of the optimal payment contracts would remain unchanged. While the utilization fee for the expensive test would remain unaffected, the FFS mark-up payments would both decrease by the exact same amount the physician internalizes. As the physician accounts for the possible health loss from delayed advanced treatment for severely ill patients, rewards for adequate treatment choice may be lowered. This may alleviate, but not eliminate, the rent-efficiency problem.[10] We further assumed that physicians are risk-neutral. If physicians were risk-averse though, we expect that the incentive problem would be reinforced. A risk-averse physician would rather opt for the expensive test and the advanced treatment because these allow her to avoid the situation with no earnings from treatment. Thus, compared to our main results, increasing the utilization fee and the spread between the FFS payments should provide optimal incentives. Another limitation is that physicians were modeled with similar diagnostic abilities. If ability was heterogeneous and the payer was only able to contract on the observed average, the designed payment contract might lead to over- or underprovision of the expensive diagnostic test depending on the physician's ability type.

Patients were modeled as passive because our aim was to focus on the decision-making problem of the attending physician. However, one could argue that patients actively take part in diagnostic test and treatment decisions. While we included diagnostic risk in our model, we did not account for therapeutic risk. Given a patient's state of illness, treatments were assumed to be successful. Yet, especially advanced treatment may come at a risk, as it may fail or even be harmful to the patient. Moreover, the expensive diagnostic device was assumed to give perfect signals. It may be interesting to extend our analysis to a more general model, which relaxes the latter assumptions.

[10] Appendix F provides proof for this intuitive argumentation. We assume that double counting may be ruled out (Chalkley and Malcomson 1998). Therefore, the first-best solution remains unchanged.

Despite the limitations, we conjecture that the principal finding of this paper remains robust. A physician, who decides on a diagnosis and treatment course that may entail the selective ordering of a superior expensive diagnostic service from a third provider, should be made accountable for the cost of that service; she should share cost responsibility for the entire episode of care. By contrast, less restrictive contracts, as commonly used in numerous OECD countries, are not incentive compatible as they do not put sufficient responsibility for ordering expensive diagnostic tests on the attending physician and allow for unconditional FFS payments.

References

Baicker, Katherine; Fisher, Elliott S.; Chandra, Amitabh (2007). Malpractice Liability Costs and The Practice of Medicine in the Medicare Program. Health Affairs (Millwood) 26 (3), pp. 841–852. DOI: 10.1377/hlthaff.26.3.841.

Baras, Jacqueline D.; Baker, Laurence C. (2009). Magnetic resonance imaging and low back pain care for Medicare patients. Health Affairs (Project Hope) 28 (6), w1133-40. DOI: 10.1377/hlthaff.28.6.w1133.

Brandt, Nikolai M.; Cassou, Matthieu (2018). Incomplete Information and Economic Incentives in Health Care. Unpublished Paper. Paper presented at the 19th European Health Economics Workshop, Paris, May 2018.

Brownlee, Shannon; Chalkidou, Kalipso; Doust, Jenny; Elshaug, Adam G.; Glasziou, Paul; Heath, Iona et al. (2017). Evidence for overuse of medical services around the world. The Lancet 390 (10090), pp. 156–168. DOI: 10.1016/S0140-6736(16)32585-5.

Chalkley, Martin; Malcomson, James M. (1998). Contracting for health services when patient demand does not reflect quality. Journal of Health Economics 17 (1), pp. 1–19. DOI: 10.1016/S0167-6296(97)00019-2.

Chandra, Amitabh; Staiger, Douglas O. (2007). Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. Journal of Political Economy 115, pp. 103–140. DOI: 10.1086/512249.

Doyle, Joseph J.; Ewer, Steven M.; Wagner, Todd H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. Journal of Health Economics 29, pp. 866–882. DOI: 10.1016/j.jhealeco.2010.08.004.

Ellis, Randall P.; McGuire, Thomas G. (1986). Provider Behaviour under Prospective Reimbursement: Cost-sharing and Supply. Journal of Health Economics 5 (2), pp. 129–151. DOI: 10.1016/0167-6296(86)90002-0.

Ellis, Randall P.; McGuire, Thomas G. (1990). Optimal Payment Systems for Health Services. Journal of Health Economics 9 (4), pp. 375–396. DOI: 10.1016/0167-6296(90)90001-J.

Felder, Stefan; Mayrhofer, Thomas (2017). Medical Decision Making. 2nd Edition. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-53432-8.

Foster, Nadine E.; Anema, Johannes R.; Cherkin, Dan; Chou, Roger; Cohen, Steven P.; Gross, Douglas P. et al. (2018). Prevention and treatment of low back pain: evidence, challenges, and promising directions. The Lancet 391 (10137), pp. 2368–2383. DOI: 10.1016/S0140-6736(18)30489-6.

Garcia Mariñoso, Begoña; Jelovac, Izabela (2003). GPs' payment contracts and their referral practice. Journal of Health Economics 22 (4), pp. 617–635. DOI: 10.1016/S0167-6296(03)00008-0.

Ghamat, Salar; Zaric, Gregory S.; Pun, Hubert (2017). Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment. Production and Operations Management 63 (5), p. 1606. DOI: 10.1111/poms.12780.

Jelovac, Izabela (2001). Physicians' payment contracts, treatment decisions and diagnosis accuracy. Health economics 10 (1), pp. 9–25. DOI: 10.1002/1099-1050(200101)10:1<9::AID-HEC560>3.0.CO;2-Y.

Laffont, Jean-Jacques; Martimort, David (2002). The Theory of Incentives. The Principal-Agent Model. Princeton: Princeton University Press. DOI: 10.2307/j.ctv7h0rwr.

Laslett, Mark; Aprill, Charles N.; McDonald, Barry; Young, Sharon B. (2005). Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Manual therapy 10 (3), pp. 207–218. DOI: 10.1016/j.math.2005.01.003.

Levin, David C.; Parker, Laurence; Palit, Charles D.; Rao, Vijay M. (2017). After Nearly A Decade Of Rapid Growth, Use And Complexity Of Imaging Declined, 2008-14. Health Affairs (Project Hope) 36 (4), pp. 663–670. DOI: 10.1377/hlthaff.2016.0836.

Little, Paul; Dorward, Martina; Warner, Greg; Stephens, Katharine; Senior, Jane; Moore, Michael (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. BMJ (Clinical research ed.) 328 (7437), p. 444. DOI: 10.1136/bmj.38013.644086.7C.

McClellan, Mark (2011). Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality? Journal of Economic Perspectives 25 (2), pp. 69–92. DOI: 10.1257/jep.25.2.69.

McGuire, Thomas G. (2000). Physician Agency. In : Handbook of Health Economics, Volume 1, Part A: Elsevier, pp. 461–536. DOI: 10.1016/S1574-0064(00)80168-7.

Mossialos, Elias; Djordjevic, Ana; Osborn, Robin; Sarnak, Dana (2017). International Profiles of Health Care Systems. Australia, Canada, China, Denmark, England, France, Germany, India, Israel, Italy, Japan, the Netherlands, New Zealand, Norway, Singapore, Sweden, Switzerland, Taiwan, and the United States. New York: The Commonwealth Fund.

Mougeot, Michel; Naegelen, Florence (2005). Hospital price regulation and expenditure cap policy. Journal of Health Economics 24 (1), pp. 55–72. DOI: 10.1016/j.jhealeco.2004.04.007.

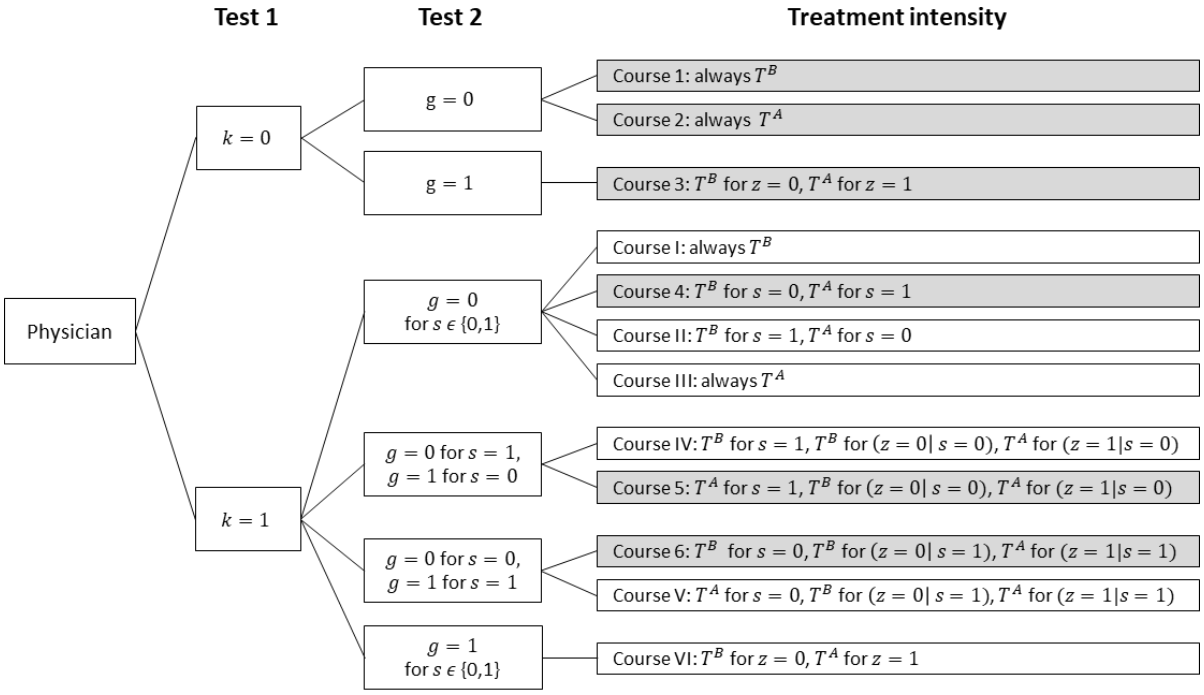
OECD (2021). Magnetic resonance imaging (MRI) exams (indicator). Available online at https://www.oecd-ilibrary.org/social-issues-migration-health/magnetic-resonance-imaging-mri-exams/indicator/english_1d89353f-en, checked on 4/24/2021.

Pflum, Kevin E. (2015). Physician Incentives and Treatment Choice. Journal of Economics and Management Strategy 24 (4), pp. 712–751. DOI: 10.1111/jems.12117.

Appendix A

The complete set of diagnosis and treatment courses is depicted in Figure 3-8. The six candidates for a first-best diagnosis and treatment course, which we focused on in the main part of the paper, are shaded in grey and referred to with Arabic numerals. All other courses, referred to with Roman numerals, are dominated by one of the six.

Figure 3-8: The Complete Set of Diagnosis and Treatment Courses



Note: The figure depicts the complete set of diagnosis and treatment courses. The six candidates for a first-best diagnosis and treatment course are shaded in grey and referred to with Arabic numerals. All other courses, referred to with Roman numerals, are dominated by one of the six. k denotes whether the physician exerts test 1, where $k = 1$ ($k = 0$) implies that the physician does (not) use test 1. s is the signal on the patient's state of illness received by the physician after exerting test 1. g indicates whether the physician uses test 2, where $g = 1$ ($g = 0$) implies that the physician does (not) use test 2. z is the signal on the patient's state of illness received by the physician after exerting test 2. T^B and T^A stand for basic and advanced treatment, respectively.

In our model, three items need to be considered to determine the optimal diagnosis and treatment course: (i) is it better to perform none, only one or both tests, (ii) if test 1 is applied, is it better to follow, to oppose or to ignore the signal and (iii) if both tests are used, which positivity criterion should be applied. We identified six potential candidates for a first-best diagnosis and treatment course (referred to with Arabic numerals and shaded in grey). All other courses, referred to with Roman numerals, are dominated by one of the six:

- It is never optimal to utilize the costly, informative test 1 and to afterwards ignore the generated signals. Thus, Courses I and III are always dominated by Courses 1 and 2, respectively.
- Because test 1 is informative ($e > 0.5$), it is better to follow the test's signal unlike to oppose the signal. Therefore, Courses II, IV and V are dominated by Courses 4, 5 and 6, respectively.
- If the physician utilized the perfect test 2 in addition to the imperfect test 1 for all patients, test 1 would be redundant. As test 1 incurs effort cost v, the physician will always refrain from utilizing test 1. Consequently, Course VI is dominated by Course 3.

The pairwise comparison of the six first-best candidates, gives the first-best solution:

$$
FB = \begin{cases}
\text{Course 1} & \text{if } \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{(1-p)(1-e)h+v}{pe}, \frac{c_M}{p}, \frac{(1-p)(1-e)h+p_{s=0}c_M+v}{p}, \frac{p_{s=1}c_{M+v}}{pe}\right\}; \\
\text{Course 2} & \text{if } h \leq \min\left\{\frac{p\ell}{(1-p)}, \frac{p(1-e)\ell+v}{e(1-p)}, \frac{c_M}{(1-p)}, \frac{p_{s=0}c_M+v}{e(1-p)}, \frac{p(1-e)\ell+p_{s=1}c_{M+v}}{(1-p)}\right\}; \\
\text{Course 3} & \text{if } \ell \geq \max\left\{\frac{c_M}{p}, \frac{p_{s=0}c_{M-v}}{p(1-e)}\right\}, \\
 & \quad h \geq \max\left\{\frac{c_M}{(1-p)}, \frac{-p(1-e)\ell+c_M-v}{(1-p)(1-e)}, \frac{p_{s=1}c_{M-v}}{(1-p)(1-e)}\right\}; \\
\text{Course 4} & \text{if } \ell \leq \min\left\{\frac{(1-p)eh-v}{p(1-e)}, \frac{(1-p)(e-1)h+c_M-v}{p(1-e)}, \frac{p_{s=0}c_M}{p(1-e)}\right\}, \\
 & \quad h \leq \min\left\{\frac{pe\ell-v}{(1-p)(1-e)}, \frac{p_{s=1}c_M}{(1-p)(1-e)}\right\}; \\
\text{Course 5} & \text{if } \ell \geq \max\left\{\frac{(1-p)(1-e)h+p_{s=0}c_M+v}{p}, \frac{p_{s=0}c_M}{p(1-e)}\right\}, \\
 & \quad h \geq \frac{p_{s=0}c_M}{p(1-e)}, h \leq \min\left\{\frac{p_{s=1}c_{M-v}}{(1-p)(1-e)}, \frac{p(1-e)\ell+(1-2p)(1-2e)c_M}{(1-p)(1-e)}\right\}; \\
\text{Course 6} & \text{if } \ell \geq \frac{p_{s=1}c_{M+v}}{pe}, \ell \leq \frac{p_{s=0}c_{M-v}}{p(1-e)}, \\
 & \quad h \geq \max\left\{\frac{p(1-e)\ell+p_{s=1}c_{M+v}}{(1-p)}, \frac{p_{s=1}c_M}{(1-p)(1-e)}, \frac{p(1-e)\ell+(1-2p)(1-2e)c_M}{(1-p)(1-e)}\right\};
\end{cases}
$$

where $p_{s=1} \coloneqq \Pr(s=1), p_{s=0} \coloneqq 1 - \Pr(s=1) = \Pr(s=0)$.

Appendix B

As the utilization of test 2 is verifiable, the payer can dictate to the physician not to use test 2 in Courses 1, 2 and 4. Alternatively, setting M arbitrarily high or refusing to reimburse the test cost to the provider of test 2, creates optimal incentives.

Implementation of Course 1 "No test, basic treatment for all" and Course 2 "No test, advanced treatment for all":

The physician's payoffs from the respective treatments yield:

Preceding test(s)	Signal	Additional payoff from T^B	Additional payoff from T^A
none	n. a.	$(1-p)B$	A

As we assume that, when indifferent, the physician will opt for the first-best action, Course 1 will only be chosen if $(1-p)B \geq A$. For the implementation of Course 2, $(1-p)B \leq A$, accordingly.

The least costly option to implement Courses 1 and 2, respectively, is to set $B = A = 0$. The physician is indifferent between treatments and will opt for the first-best. If the physician utilized test 1 and followed the signal, her payoff would decrease due to effort cost v. Hence, she will refrain from utilizing test 1. For both courses, her expected payoff from treatment yields zero, thus for the PC to hold, $D = \omega \geq 0$.

Implementation of Course 3 "Test 2 only":

As the utilization of test 2 as well as the test's generated signals are verifiable, the payer can simply dictate to the physician to always utilize test 2 and to follow the signals. Alternatively, setting $M = 0$ provides optimal incentives.

The least costly option to implement Course 3, is to set $B = A = 0$. For any signal, the physician is indifferent between treatments and will opt for the first-best. If the physician utilized test 1 before test 2, her expected payoff from treatment would remain unchanged and because test 1 incurs effort cost v, she will always refrain from utilizing test 1. Her expected payoff from treatment amounts to zero, thus for the PC to hold, $D = \omega \geq 0$.

Implementation of Course 4 "Test 1 only":

The physician's payoffs from the respective treatments yield:

Preceding test(s)	Signal	Additional payoff from T^B	Additional payoff from T^A
test 1	$s = 0$	$\boldsymbol{(1 - p_0^s)B}$	A
	$s = 1$	$(1 - p_1^s)B$	$\boldsymbol{A}$

In Course 4, the physician has to be incentivized to comply with the outcome of test 1 for treatment choice. She will only choose T^B if $(1 - p_0^s)B \geq A$ and only choose T^A if $(1 - p_1^s)B \leq A$. Furthermore, she needs to be incited to utilize test 1 in the first place. Therefore, at the optimum, her expected utility from exerting diagnostic effort has to be at least equal to her expected utility from alternative courses that do not entail diagnostic effort. Course 3 may be ignored as it requires the utilization of test 2. Thus, the following constraints need to hold:

$$(1 - p_0^s)B \geq A \tag{36}$$

$$(1 - p_1^s)B \leq A \tag{37}$$

$$EU_4 \geq \max\{EU_1, EU_2\} \tag{38}$$

Simplifying (36), (37) and (38) and inserting for p_0^s and p_1^s from equation (1) yields the following incentive constraints:

$$-[p(1 - e) + (1 - p)e]A + (1 - p)eB > 0 \tag{39}$$

$$[pe + (1 - p)(1 - e)]A + (1 - p)(e - 1)B \geq 0 \tag{40}$$

$$[pe + (1 - p)(1 - e)]A + (1 - p)(e - 1)B \geq v \tag{41}$$

$$-[p(1 - e) + (1 - p)e]A + (1 - p)eB \geq v \tag{42}$$

IC (41) strictly dominates IC (40) and IC (42) strictly dominates IC (39). Thus, we are left with ICs (41) and (42). A and B have to be chosen such that both ICs bind. Adding up (41) and (42) and solving for A yields $A = (1 - p)B$. Inserting for A into one of the two equations gives

$$B = \frac{v}{p(2e-1)(1-p)} \qquad \text{and} \qquad A = \frac{v}{p(2e-1)}.$$

At the optimum, the physician's expected payoff amounts to $\frac{v}{p(2e-1)}$. From the physician's PC we find that $D \leq \omega - \frac{v}{p(2e-1)}$. This may require a negative capitation fee D (compare Garcia Mariñoso and Jelovac 2003).

Appendix C

Proof for Lemma 3:

Rearranging equations (24), (25) and (26) yields:

$$A \geq \frac{v+[(1-e)+(1-e)p+e(1-p)]M}{p} \quad (24')$$

$$A \geq \frac{(1-e)}{pe} M \quad (25')$$

$$A \geq \frac{(1-e)p+e(1-p)}{(1-e)p} M \quad (26')$$

(26') requires a strictly larger A in order to hold than (25'):

$\frac{(1-e)(1-e)}{pe(1-e)} M < \frac{(1-e)pe+e^2(1-p)}{pe(1-e)} M$ ↔

$(1-e)(1-e) < (1-e)pe + e^2(1-p)$ ↔

$1 > pe$

(26') requires a strictly larger A in order to hold than (24'):

$\frac{e(1-e)v+e(1-e)[(1-e)+(1-e)p+e(1-p)]M}{pe(1-e)} < \frac{(1-e)pe+e^2(1-p)}{pe(1-e)}$ ↔

$eM - 2e^2M - e^2pM + pe^3M < v(e-1)$ ↔

Rearranging and inserting from (22) for v gives

$eM - 2e^2M - e^2pM + pe^3M < p(2e-1)M(e-1)$ ↔

$1 - p - 2e + 2pe < 0$ ↔

$1 < 2e$

□

Appendix D

Derivation of the incentive-compatibility constraints for Course 6:

Inciting the optimal treatment intensity: As in Course 5, the payer can dictate to the physician to follow the results of test 2. In the situation, in which the physician only took the pretest though and received signal $s = 0$, her expected payoff from choosing T^A, which always cures the patient, is A and her expected payoff from T^B is $(1 - p_0^s)B$. Incentive compatibility calls for:

$$A \leq (1 - p_0^s)B \tag{43}$$

Inciting the optimal utilization of test 2: In case the physician uses test 2, for the share of mildly ill patients, $(1 - \tilde{p}^s)$, she has to apply T^B and receives payoff B. For the share of severely ill patients, $\tilde{p}^s$, she has to apply T^A, which this leads to payment A. For the utilization of test 2 the physician always has to pay the fee M. Hence, utilizing test 2 yields an additional payoff of $\tilde{p}^s A + (1 - \tilde{p}^s)B - M$, depending on the outcome of test 1. In case test 2 is not utilized, she chooses the treatment that yields the higher expected payoff. The expected additional payoff is $\max\{A, (1 - \tilde{p}^s)B\}$. Test 2 should be applied if and only if test 1 results in signal $s = 1$. This yields the following ICs:

$$p_0^s A + (1 - p_0^s)B - M \leq \max\{A, (1 - p_0^s)B\} \tag{44}$$

$$p_1^s A + (1 - p_1^s)B - M \geq \max\{A, (1 - p_1^s)B\} \tag{45}$$

Taking into consideration IC (43), ICs (44) and (45) may be rewritten as:

$$p_0^s A \leq M \tag{46}$$

$$(1 - p_1^s)(B - A) \geq M \tag{47}$$

$$p_1^s A \geq M \tag{48}$$

Inciting the optimal utilization of test 1: The physician takes the decision on the utilization of test 1 independent of any preceding events. In case she exerts effort, and the above ICs are met, Course 6 yields an expected utility of:

$$EU_6 = D + peA + (1 - p)B - [pe + (1 - p)(1 - e)]M - v \tag{49}$$

For each patient, the physician sees, she is paid capitation D. Only patients with signal $s = 1$ resulting from test 1 and signal $z = 1$ resulting from test 2 receive T^A, thus she receives mark-up A for patient share pe. Patients with a mild illness with signal $s = 0$ receive T^B, hence for patient share $(1-p)e$ she receives mark-up B. Furthermore, patients with signals $s = 1$ and $z = 0$ obtain T^B, generating a mark-up B for patient share $(1-p)(1-e)$. Since all patients with signal $s = 1$ resulting from test 1 additionally receive test 2, the physician has to bear cost M for patient share $[pe + (1-p)(1-e)]$. As the physician always utilizes test 1, she incurs a disutility of effort v.

In case she does not utilize test 1, she may treat all patients with T^B first (Course 1), treat all patients with T^A right away (Course 2) or utilize test 2 and base the treatment choice on the perfect diagnostic signal z (Course 3). This yields the condition $EU_6 \geq \max\{EU_1, EU_2, EU_3\}$. Plugging in for EU_1, EU_2, EU_3 from (12) to (14) and regrouping yields the following three conditions:

$$peA + [(1-p)(e-1) - pe]M \geq v \tag{50}$$

$$(pe-1)A + (1-p)B + [(1-p)(e-1) - pe]M \geq v \tag{51}$$

$$p(e-1)A + [p(1-e) + (1-p)e]M \geq v \tag{52}$$

Using the definitions of p_0^s and p_1^s (see equation (1)) to express everything in terms of p and e, conditions (43), (46), (47) and (48) may be rewritten and regrouped as:

$$-[p(1-e) + (1-p)e]A + (1-p)eB \geq 0 \tag{53}$$

$$p(e-1)A + [p(1-e) + (1-p)e]M \geq 0 \tag{54}$$

$$(1-p)(e-1)A + (1-p)(1-e)B + [(1-p)(e-1) - pe]M \geq 0 \tag{55}$$

$$peA + [(1-p)(e-1) - pe]M \geq 0 \tag{56}$$

We obtain that (50) implies (56) and (52) implies (54).

Solution to the optimization problem:

To solve the problem, we first check whether one or several of the ICs have to be binding. Since increasing A furthers IC (50), but counters all other ICs, (50) has to be binding at the optimum. Furthermore, we find that increasing M furthers IC (52), but counters all other ICs, therefore (52) has to be binding at the optimum.

Lemma 4: *In Course 6, effort constraints (50) and (52) are binding at the optimum.*

Since ICs (50) and (52) are binding, we can rewrite both ICs and obtain:

$$-peA = -v + [(1-p)(e-1) - pe]M \tag{57}$$

$$p(e-1)A = v - [p(1-e) + (1-p)e]M \tag{58}$$

Adding up ICs (57) and (58) and solving for M yields $M = pA$. Inserting M for pA into either IC (57) or (58) and solving leads to the unique solution

$$M = \frac{v}{(1-p)(2e-1)} \tag{59}$$

Turning to the solution for mark-up payments A and B, we know from Lemma 4 that ICs (50) and (52) are binding at the optimum. A necessary condition for these to bind is $A > 0$. Moreover, ICs (51) and (55) only hold, if we set $B > A$. We conclude with Lemma 5.

Lemma 5: *Setting $B > A > 0$ is a necessary condition to provide optimal incentives for the implementation of Course 6.*

Rewriting $M = pA$ yields:

$$A = M/p \tag{60}$$

Inserting for M from (59), we obtain:

$$A = \frac{v}{p(1-p)(2e-1)} \tag{61}$$

For an optimal solution, mark-up B must be chosen as low as possible to meet the remaining constraints (51), (53) and (55). Inserting for A from equation (61) and regrouping yields:

$$(pe-1)M + p(1-p)B + [(1-p)(e-1) - pe]pM \geq vp \tag{62}$$

$$-[p(1-e) + (1-p)e]M + p(1-p)eB \geq 0 \tag{63}$$

$$(1-p)(e-1)M + p(1-p)(1-e)B + [(1-p)(e-1) - pe]pM \geq 0 \tag{64}$$

We find that, at the optimum, IC (64) has to be binding, since, in order to hold, it requires a strictly larger B than (62) and (63). Solving for B yields

$$B = \frac{v\left(1-e+p^2(2e-1)\right)}{p(1-p)^2(1-e)(2e-1)} \tag{65}$$

Appendix E

Proof for Proposition 3:

Denote by $\tilde{A}$ and $\tilde{B}$ the mark-ups for providing T^A and T^B, respectively, which can be billed unconditionally. Furthermore, note that $\tilde{A} = \tilde{B} = 0$ cannot implement either course because there would be no incentives to perform test 1.

Course 5 requires that after test 1 with test result $s = 1$, the physician chooses T^A right away, which would lead to payoff $\tilde{A}$. Alternatively, the physician could try T^B first. Yet, as this only cures the patient with probability $(1 - p_1^s)$, she has to apply T^A on top with probability p_1^s. Thus, the payoff is $\tilde{B} + p_1^s\tilde{A}$. We obtain the condition $\tilde{B} + p_1^s\tilde{A} \leq \tilde{A} \Leftrightarrow \tilde{B} \leq (1 - p_1^s)\tilde{A}$ implying $\tilde{B} < \tilde{A}$. Furthermore, Course 5 requires that test 2 should be applied if and only if test 1 results in signal $s = 0$. This yields the following ICs:

$$p_0^s\tilde{A} + (1 - p_0^s)\tilde{B} - M \geq \max\{\tilde{A}, \tilde{B} + p_0^s\tilde{A}\}$$
$$p_1^s\tilde{A} + (1 - p_1^s)\tilde{B} - M \leq \max\{\tilde{A}, \tilde{B} + p_1^s\tilde{A}\}$$

Using $\tilde{B} + p_1^s\tilde{A} \leq \tilde{A}$, we obtain:

$$(1 - p_0^s)(\tilde{B} - \tilde{A}) - M \geq 0$$
$$(1 - p_1^s)(\tilde{B} - \tilde{A}) - M \leq 0$$

This implies:

$$(p_1^s - p_0^s)(\tilde{B} - \tilde{A}) \geq 0$$

Now, $p_1^s > p_0^s$ requires $\tilde{B} > \tilde{A}$ which, however, contradicts the requirement above. Hence, unconditional FFS mark-up payments fail to implement Course 5.

Course 6 requires that after test 1 with test result $s = 0$, the physician chooses T^B. This would lead to payoff $\tilde{B} + p_0^s\tilde{A}$. Alternatively, the physician could apply T^A, which would lead to payoff $\tilde{A}$. We obtain the condition $\tilde{B} + p_0^s\tilde{A} \geq \tilde{A}$. Also, Course 6 requires that test 2 should be applied if and only if test 1 results in signal $s = 1$. This yields the following ICs:

$$p_0^s\tilde{A} + (1 - p_0^s)\tilde{B} - M \leq \max\{\tilde{A}, \tilde{B} + p_0^s\tilde{A}\}$$
$$p_1^s\tilde{A} + (1 - p_1^s)\tilde{B} - M \geq \max\{\tilde{A}, \tilde{B} + p_1^s\tilde{A}\}$$

Using $\tilde{B} + p_0\tilde{A} \geq \tilde{A}$, we obtain:

$$-p_0^s\tilde{B} - M \leq 0$$
$$-p_1^s\tilde{B} - M \geq 0$$

This implies $p_0^s\tilde{B} \geq p_1^s\tilde{B}$. Due to $p_1^s > p_0^s$, this is possible only if $\tilde{B} = 0$, requiring also $\tilde{A} = 0$. Without any mark-ups, however, there cannot be any incentive to perform test 1 as there is no financial gain from the test result. Hence, unconditional FFS mark-up payments fail to implement Course 6.

Appendix F

Implementation for (partially) altruistic physicians

If we supposed that the physician also accounted for the patients' well-being, by at least partially internalizing the patients' health loss from receiving an initially insufficient treatment, her objective function would change as follows (compare equation (4)):

$$\max_{k,g,\tau} \widehat{EU} = EI\left(\hat{D}, \hat{B}, \hat{A}, \hat{M}\right) - EC(v) - \alpha EL(\ell) \tag{66}$$

As in the main part of the paper, $\hat{D}$ is a capitation payment, $\hat{B}$ and $\hat{A}$ denote the conditional mark-up payments and $\hat{M}$ is the fee for the utilization of test 2. The physician's expected utility is defined by her expected income, minus her expected effort cost from test 1 and, now additionally, minus the monetary equivalent of her expected disutility from the patients' health loss ℓ. The degree of altruism is restricted to $\alpha \in (0,1]$, whereby $\alpha < 1$ ($\alpha = 1$) indicates that the physician partially (fully) internalizes the patients' health loss. We proceed analogously to the situation without altruism to solve the incentive problem.

Implementation of Course 5 for altruistic physicians:

Inciting the optimal treatment intensity: Given the physician only takes the pretest and receives signal $s = 1$, her expected payoff is $\hat{A}$ for choosing T^A and $(1 - p_1^s)\hat{B} - p_1^s \alpha \ell$ for choosing T^B. With probability p_1^s, the patient suffers from the severe illness and T^B is insufficient. Consequently, the patient experiences health loss ℓ from delayed adequate treatment, which, in turn, is internalized by the physician. Incentive compatibility requires:

$$\hat{A} \geq (1 - p_1^s)\hat{B} - p_1^s \alpha \ell \tag{67}$$

Inciting the utilization of test 2: In case the physician utilizes test 2 and follows its perfect signals, the patient always receives the adequate treatment. Therefore, the physician's pay-off remains unaffected by her concern for the patients' well-being (see the left-hand side of equations (68) and (69)). In case test 2 is not utilized, the physician chooses the treatment that yields her the higher expected payoff. This may lead to an insufficient treatment choice and her expected additional payoff is $\max\left\{\hat{A}, (1 - \tilde{p}^s)\hat{B} - \tilde{p}^s \alpha \ell\right\}$. The resulting ICs are given by:

$$p_0^s \hat{A} + (1 - p_0^s)\hat{B} - \hat{M} \geq \max\left\{\hat{A}, (1 - p_0^s)\hat{B} - p_0^s \alpha \ell\right\} \tag{68}$$

$$p_1^s \hat{A} + (1 - p_1^s)\hat{B} - \hat{M} \leq \max\left\{\hat{A}, (1 - p_1^s)\hat{B} - p_1^s \alpha \ell\right\} \tag{69}$$

Inciting the utilization of test 1: In case the physician exerts effort and ICs (67) to (69) are met, her expected utility from Course 5, $\widehat{EU}_5$, remains unaffected by her concern for the patients' well-being (compare equation (11)). Since the disjunctive positivity criterium is applied, no patient receives insufficient treatment. In case she does not exert effort, she may either treat all patients with T^B first (Course 1), treat all patients with T^A right away (Course 2) or comply to test 2 for treatment choice (Course 3). Her expected utilities from Courses 2 and 3, $\widehat{EU}_2$ and $\widehat{EU}_3$, remain unaffected (compare equations (13) and (14), respectively). In Course 1, the patient suffers from the severe illness with probability p. This makes T^B insufficient and leads to health loss ℓ for the patient. Accordingly, the physician's expected utility amounts to

$$\widehat{EU}_1 = \hat{D} + (1-p)\hat{B} - p\alpha\ell \tag{70}$$

Utilizing test 1 is in the physician's private best interest, if and only if contract $(\hat{B}, \hat{A}, \hat{M})$ fulfills the condition $\widehat{EU}_5 \geq \max\{\widehat{EU}_1, \widehat{EU}_2, \widehat{EU}_3\}$. Plugging in and regrouping yields:

$$p\alpha\ell + [p + (1-p)(1-e)]\hat{A} + (1-p)(e-1)\hat{B} - [p(1-e) + (1-p)e]\hat{M} \geq v \tag{71}$$

$$-(1-p)e\hat{A} + (1-p)e\hat{B} - [p(1-e) + (1-p)e]\hat{M} \geq v \tag{72}$$

$$(1-p)(1-e)\hat{A} + (1-p)(e-1)\hat{B} + [pe + (1-p)(1-e)]\hat{M} \geq v \tag{73}$$

Note that ICs (72) and (73) are similar to ICs (16) and (17), respectively.

Using (1) and IC (67), ICs (67) to (69) may be rewritten and regrouped as:

$$[pe + (1-p)(1-e)]\hat{A} + (1-p)(e-1)\hat{B} + p\alpha\ell \geq 0 \tag{74}$$

$$-(1-p)e\hat{A} + (1-p)e\hat{B} - [p(1-e) + (1-p)e]\hat{M} \geq 0 \tag{75}$$

$$p(1-e)\hat{A} + p(1-e)\alpha\ell - [p(1-e) + (1-p)e]\hat{M} \geq 0 \tag{76}$$

$$(1-p)(1-e)\hat{A} + (1-p)(e-1)\hat{B} + [pe + (1-p)(1-e)]\hat{M} \geq 0 \tag{77}$$

Note that IC (75) is similar to (19) and is implied by (72), while IC (77) is similar to (21) and is implied by (73).

Minimizing the sum of the expected payments to the physician yields the following problem:

$$\min_{(\hat{A},\hat{B},\hat{M})} \{[p + (1-p)(1-e)]\hat{A} + (1-p)e\hat{B} - [p(1-e) + (1-p)e]\hat{M}\}$$

subject to (71), (72), (73), (74) and (76).

ICs (71) to (73) are effort constraints, (74) is the treatment constraint and (76) is the constraint on the utilization of test 2. Compared to the initial situation without altruism, the ICs are less restrictive. To avoid the expected disutility from the patients' health loss, it is relatively more attractive for the physician to exert diagnostic effort than to try T^B first, to provide T^A right away and to utilize the perfect test 2.

As in the situation without altruism, ICs (72) and (73) are binding at the optimum. Adding up the equations and simplifying yields $\widehat{M} = (1-p)\left(\hat{B} - \hat{A}\right)$. Hence, we find:

$$\widehat{M} = \frac{v}{p(2e-1)}$$

We consider the remaining ICs (71), (74) and (76) to solve for $\hat{A}$ and $\hat{B}$. Rewriting $(1-p)\left(\hat{B} - \hat{A}\right) = \widehat{M}$ yields $\hat{B} = \hat{A} + \frac{\widehat{M}}{1-p}$. Inserting for $\hat{B}$ into each of the ICs, rewriting and solving for $\hat{A}$ gives:

$$\hat{A} \geq \frac{v+[(1-e)+(1-e)p+e(1-p)]\widehat{M}}{p} - \alpha\ell \tag{78}$$

$$\hat{A} \geq \frac{(1-e)}{pe}\widehat{M} - \alpha\ell \tag{79}$$

$$\hat{A} \geq \frac{(1-e)p+e(1-p)}{(1-e)p}\widehat{M} - \alpha\ell \tag{80}$$

We look for the lowest value of $\hat{A}$ fulfilling the above incentive constraints. Referring to Appendix C, it is straightforward that (80) has to be binding. Inserting for $\widehat{M}$ into (80) and rewriting results in:

$$\hat{A} = \frac{v(p(1-e)+(1-p)e)}{p^2(1-e)(2e-1)} - \alpha\ell \tag{81}$$

Inserting for $\hat{A}$ and $\widehat{M}$ into $\hat{B} = \hat{A} + \frac{\widehat{M}}{1-p}$ yields:

$$\hat{B} = \frac{v\left(e+\left(p^2-2p\right)(2e-1)\right)}{p^2(1-p)(1-e)(2e-1)} - \alpha\ell \tag{82}$$

Finally, $\widehat{D}$ will be set just only as high as necessary to ensure that the physician's participation constraint $\widehat{EU_5} \geq \omega$ is met, possibly constrained by $\widehat{D} \geq 0$.

For the implementation of Course 6, we proceed in the same manner. The results will be briefly outlined.

Implementation of Course 6 for altruistic physicians:

Inciting the optimal treatment intensity: Given the physician only takes the pretest and receives signal $s = 0$, her expected payoff is $\hat{A}$ for choosing T^A and $(1 - p_0^s)\hat{B} - p_0^s\alpha\ell$ for choosing T^B as, with probability p_0^s, the patient suffers from the severe illness. Incentive compatibility requires:

$$\hat{A} \leq (1 - p_0^s)\hat{B} - p_0^s\alpha\ell \tag{83}$$

Inciting the utilization of test 2: The incentive constraints for the optimal utilization of test 2 are derived based on the same rationale as for Course 5 and are given by:

$$p_0^s\hat{A} + (1 - p_0^s)\hat{B} - \widehat{M} \leq \max\left\{\hat{A}, (1 - p_0^s)\hat{B} - p_0^s\alpha\ell\right\} \tag{84}$$

$$p_1^s\hat{A} + (1 - p_1^s)\hat{B} - \widehat{M} \geq \max\left\{\hat{A}, (1 - p_1^s)\hat{B} - p_1^s\alpha\ell\right\} \tag{85}$$

Inciting the utilization of test 1: In case the physician exerts effort and ICs (83) to (85) are met, her expected utility from Course 6 amounts to

$$\widehat{EU}_6 = \widehat{D} + pe\hat{A} + (1 - p)\hat{B} - [pe + (1 - p)(1 - e)]\widehat{M} - p(1 - e)\alpha\ell - v \tag{86}$$

Since the conjunctive positivity criterium is applied, it is straightforward that patients, for whom the physician initially receives signal $s = 0$, are treated with T^B right away. Yet, with probability $p(1 - e)$, T^B is insufficient to cure the patient. The patient suffers from health loss ℓ due to delayed adequate treatment, which is internalized by the physician. In case the physician does not exert effort, she follows Course 1, Course 2 or Course 3. Again, her expected utilities from Courses 2 and 3, $\widehat{EU}_2$ and $\widehat{EU}_3$, remain unaffected by her concern for the patients' well-being and her expected utility from Course 1 is given by equation (70).

Utilizing test 1 is in the physician's private best interest, if and only if contract $(\hat{B}, \hat{A}, \widehat{M})$ fulfills the condition $\widehat{EU}_6 \geq \max\left\{\widehat{EU}_1, \widehat{EU}_2, \widehat{EU}_3\right\}$. Plugging in and regrouping yields:

$$pe\hat{A} + [(1 - p)(e - 1) - pe]\widehat{M} + pe\alpha\ell \geq v \tag{87}$$

$$(pe - 1)\hat{A} + (1 - p)\hat{B} + [(1 - p)(e - 1) - pe]\widehat{M} - p(1 - e)\alpha\ell \geq v \tag{88}$$

$$p(e - 1)\hat{A} + [p(1 - e) + (1 - p)e]\widehat{M} + p(e - 1)\alpha\ell \geq v \tag{89}$$

Using (1) and IC (83), ICs (83) to (85) may be rewritten and regrouped as:

$$-[p(1-e)+(1-p)e]\hat{A}+(1-p)e\hat{B}-p(1-e)\alpha\ell \geq 0 \quad (90)$$

$$p(e-1)\hat{A}+[p(1-e)+(1-p)e]\widehat{M}+p(e-1)\alpha\ell \geq 0 \quad (91)$$

$$(1-p)(e-1)\hat{A}+(1-p)(1-e)\hat{B}+[(1-p)(e-1)-pe]\widehat{M} \geq 0 \quad (92)$$

$$pe\hat{A}+[(1-p)(e-1)-pe]\widehat{M}+pe\alpha\ell \geq 0 \quad (93)$$

We obtain that (89) implies (91) and (87) implies (93). Therefore, the optimization problem reduces to:

$$\min_{(\hat{A},\hat{B},\widehat{M})}\{pe\hat{A}+(1-p)\hat{B}-[pe+(1-p)(1-e)]\widehat{M}\}$$

subject to (87), (88), (89), (90), (92).

ICs (87) and (89) are binding. Adding up the ICs and solving for $\widehat{M}$ gives $\widehat{M}=p(\hat{A}+\alpha\ell)$. Inserting M for $p(\hat{A}+\alpha\ell)$ into either IC and solving leads to the unique solution:

$$\widehat{M}=\frac{v}{(1-p)(2e-1)}$$

Inserting for $\widehat{M}$ into $\widehat{M}=p(\hat{A}+\alpha\ell)$ and rewriting, we obtain:

$$\hat{A}=\frac{v}{p(1-p)(2e-1)}-\alpha\ell$$

For an optimal solution, mark-up $\hat{B}$ must be chosen as low as possible to meet the remaining ICs (88), (90) and (92). Inserting for $\hat{A}$ into these ICs and regrouping yields:

$$(pe-1)\widehat{M}+p(1-p)\hat{B}+[(1-p)(e-1)-pe]p\widehat{M}+p(1-p)\alpha\ell \geq vp \quad (94)$$

$$-[p(1-e)+(1-p)e]\widehat{M}+p(1-p)e\hat{B}+p(1-p)e\alpha\ell \geq 0 \quad (95)$$

$$(1-p)(e-1)\widehat{M}+p(1-p)(1-e)\hat{B}+[(1-p)(e-1)-pe]p\widehat{M}+p(1-p)(1-e)\alpha\ell \geq 0 \quad (96)$$

At the optimum, IC (96) has to be binding, because, in order to hold, it requires a strictly larger $\hat{B}$ than (94) and (95). Solving for $\hat{B}$ yields

$$\hat{B}=\frac{v\left(1-e+p^2(2e-1)\right)}{p(1-p)^2(1-e)(2e-1)}-\alpha\ell$$

For the implementation of both composite courses, the utilization fee for test 2 remains unaffected by the physician's concern for the patients' well-being. We have $\widehat{M}=M$. The FFS

mark-up payments, by contrast, both decrease by the exact same amount the physician internalizes, $\alpha\ell$. This implies $\hat{A} < A$ and $\hat{B} < B$.The physician considers the possible health loss from delayed advanced treatment for severely ill patients and, therefore, rewards for adequate treatment choice may be lowered. Her earnings from the reward payments are smaller than in the initial situation without altruism. Thus, a greater share of her gains may be extracted by the utilization fee for test 2. Depending on the size of the physician's reservation utility, capitation payment $\widehat{D}$ may be set larger, $\widehat{D} \geq D$, without having to grant an information rent to the physician. This may alleviate, but not eliminate, the rent-efficiency problem.

Chapter 4

The Provision of Breast MRIs in Adherence to Guideline-Based Reimbursement Conditions and the Impact on Cost and Outcomes: Evidence from Germany

Katharina Beenk Dirk Horenkamp-Sonntag
Udo Schneider Mathias Kifmann

Abstract

The provision of costly magnetic resonance imaging (MRI) for breast cancer treatment has increased in recent years. Clinical practice guidelines recommend a prudent use of breast MRIs and, in Germany, the Federal Joint Committee (G-BA) puts these guidelines into concrete terms by enacting under which conditions breast MRI scans are reimbursed by statutory health insurance (SHI). However, due to weak enforcement in combination with asymmetric information, providers may disregard the reimbursement conditions. Analyzing administrative data, we identify outpatient breast MRI scans in Germany, which are provided non-adherent to the reimbursement conditions of the SHI. Next, we examine the effect of a non-adherent provision on women's direct medical cost of care, medical resource utilization, personal well-being, breast cancer specific procedures and survival over up to two years. To allow for causal interpretation, we apply propensity score matching and a difference-in-difference approach. Results suggest that 34.3% of the breast MRI scans that were performed in 2016 may be classified as non-adherent to reimbursement conditions. Women who received such non-adherent MRI on average cause 26.7% higher direct medical cost per year than women in a control group. Significantly increased medical resource consumption and greater probabilities to receive additional diagnostic and treatment procedures drive the significant growth in cost. Effects on women's personal well-being are found to be insignificant. Concomitantly, the two-year all-cause survival probability significantly increases by 1.3 percentage points. Hence, it is not straightforward whether the overprovision of MRI scans is indeed ineffective and should be targeted by regulators. We conclude that further research is needed to inform regulators.

4.1 Introduction

Breast cancer (BC) is by far the most frequently diagnosed form of cancer in European women. In 2018, incidence rates varied between 60-155 per 100,000 women and, due to an aging population as well as increased pick-up of screening programs, rates are rising (Ferlay et al. 2018; Quante et al. 2016). Correspondingly, expenditures for BC treatment increase, which constitutes a burden on health care systems. In this light, the number of advanced imaging exams, such as magnetic resonance imaging (MRI) of the breast, and expenditures for said scans have increased in recent years (Jönsson et al. 2016; Placido et al. 2017; Stout et al. 2014). While international clinical practice guidelines propose a prudent provision of breast MRI, evidence indicates that there is a gap between these guidelines and actual physician practice (see Baxi et al. 2017 for an international survey).

In the course of BC treatment, MRI may be an effective tool for primary diagnosis and for surveillance. Compared to conventional imaging such as mammography (MG) and ultrasound (US), breast MRI is more sensitive and, therefore, the information gained from the MRI may influence the therapeutic strategy (Kuhl 2007; Kuhl et al. 2010). Yet, also due to its limited specificity, breast MRI is found to have controversial effects on patient outcomes such as survival and distress from further diagnostic procedures as well as follow-up cost (Bedrosian et al. 2016; Onega et al. 2018; Padia et al. 2016). Furthermore, MRI scans are many times more expensive than conventional imaging. Under the German statutory health insurance (SHI), the provision of breast MRI for the indication breast cancer is reimbursed with €223.27 in the first quarter of 2021, whereas MG and US of the breast are reimbursed with €60.98 and €16.69, respectively (Federal Association of Statutory Health Insurance Physicians 2021).

Applying a retrospective, population-based cohort study design, using administrative data from the largest German sickness fund from 2014 to 2018, this paper pursues two aims: Firstly, we identify outpatient breast MRIs, which are provided non-adherent to the reimbursement conditions of the SHI. Following national clinical practice guidelines, the Federal Joint Committee regulates the provision of outpatient breast MRIs by enacting under which conditions an MRI is reimbursed by the SHI. Amongst other things, conventional imaging (MG and US) must have preceded the MRI and a certain waiting period after breast conserving or reconstruction therapy has to be adhered to. However, due to asymmetric information, data privacy restrictions and weak enforcement during the relevant time period in our study, the reimbursement conditions may be disregarded.[1] By comparing clinical practice to the

[1] See Section 4.2 for a detailed exposition of the reimbursement conditions and the enforcement mechanisms.

reimbursement conditions, we identify women who received an MRI that was provided non-adherent to the reimbursement conditions of the SHI. Hereinafter, we also refer to these scans as non-conform MRIs. Secondly, we examine the effect of a non-conform MRI, which was performed for the purpose of primary diagnosis or surveillance of BC, on women's direct medical cost of care, medical resource consumption, personal well-being, BC specific procedures and survival over up to two years. Women with breast cancer who received a non-conform MRI are compared to similar women who did not receive any MRI. To provide robustness against observed and unobserved heterogeneity, we employ two risk adjustment procedures: differences in observed characteristics between the non-conform MRI and the control group are eliminated with propensity-score matching. To account for differences in time-invariant unobserved heterogeneity, we apply a difference-in-difference approach. We investigate 4,687 outpatient MRI scans for the indication of breast cancer that were performed in 2016, of which 34.3% are classified as non-adherent to the reimbursement conditions. Results of this multi-outcome study imply that women who received a non-conform MRI on average cause significantly higher direct medical cost in the following two years than similar women in the control group. Total cost increase by 26.7%. A significant growth in medical resource consumption, captured by, i.a., hospital days (+16.2%) and pharmaceutical prescriptions (+7.4%), mirrors the elevated cost. Further, women in the non-conform MRI group more likely receive BC specific diagnostic and treatment procedures such as biopsy (+4.5 ppts.) and breast conserving surgery (+8.7 ppts.), which explains the increased resource utilization. Concomitantly, for women in the non-conform MRI group all-cause mortality is found to be 1.3 ppts. (p=0.062) lower than in the control group.

Literature suggests that, when applied for primary diagnosis or for surveillance of BC, breast MRI may generally have controversial effects on patient outcomes and health care cost. First, it is unclear whether breast MRI leads to improved disease-specific survival. While evidence from randomized controlled trials (RCTs) is not yet available, most observational studies find that breast MRI has no significant effect on survival (Onega et al. 2018; Parsyan et al. 2013; Wang et al. 2018). Moreover, partly due to possible false-positive results and over-diagnosis, breast MRI is associated with increased probability of further diagnostic investigation such as repeat MRI, biopsy and additional physician referrals (Brennan et al. 2010; Padia et al. 2016). There is mixed evidence on whether MRI, in turn, also results in increased patient distress and anxiety (Brédart et al. 2012; Spiegel et al. 2011). Further, patients who received an MRI are more likely to receive extensive surgeries, while re-operation rates do not seem to decline (Houssami et al. 2017). All those procedures are reflected by increased cost of care (Bedrosian

et al. 2016; Hayes et al. 2016). Padia et al. (2016), for instance, review breast MRIs that were provided at a single institution including newly diagnosed patients and patients for surveillance. They report that false-positive findings result in further imaging (in 51.6 % of all cases), additional laboratory testing (8.6%), biopsies (5.7%) and referrals (5.7%). These procedures lead to additional cost of $ 328 per patient. Brennan et al. (2010) perform a retrospective study of breast MRIs and discover that, if used for surveillance, MRI leads to biopsy in 31% of all cases, yet only 12% of all cases were subsequently diagnosed with second BC. Houssami et al. (2017) conduct a meta-analysis of RCTs and comparative studies on the effect of pre-operative MRI. They find that pre-operative MRI is associated with increased odds of receiving mastectomy. Further, Wang et al. (2016) argue that pre-operative MRI may lead to over-diagnosis as increased detection rates, attributable to MRI, are not offset by a decrease of subsequent occurrence among women with early-stage BC. Finally, Bedrosian et al. (2016) conducted a retrospective study using institutional billing records and found that women who received pre-operative MRI on average cause 30.1% higher cost than women who did not.

The related studies predominantly consider data from single institutions in the United States, where BC diagnosis and treatment patterns, decision-making factors, health service/ insurance structures and fiscal considerations may be different and not generalizable to the German population. Besides, findings from the United States are mainly based on the Medicare population, which only comprises women aged 65 and older. In comparison to the above-mentioned studies, this paper has several strengths: It is the first German study that exams the impact of MRI scans, which were provided non-adherent to the reimbursement conditions of the SHI, on a large variety of outcomes. Because administrative data serve as the basis of the analyses, we can observe women aged 18 and older, who were treated at multiple institutions, and are able to analyze data across care sectors.

The remainder of this paper is organized as follows: Section 4.2 gives background information on the provision of outpatient breast MRI scans in Germany. Sections 4.3 and 4.4 present the data and outline the empirical strategy. Section 4.5 shows the main results, robustness checks and analyses of heterogeneous effects. Finally, Section 4.6 discusses the findings and concludes.

4.2 Provision of Outpatient Breast MRI Scans in Germany

In Germany, evidence- and consensus-based clinical practice guidelines specify the appropriate utilization of breast MRI in screening, primary diagnosis and surveillance of BC (Guideline

Programs Oncology 2017). In brief, for screening of patients with a high lifetime risk of breast and ovarian cancer (patients with BRCA1/2 gene mutation or a lifetime risk ≥ 30%) periodic MRI is recommended, whereas for screening of patients with an average lifetime risk of BC it is not. For primary diagnosis of symptomatic patients and for surveillance after primary treatment, MRI is recommended only adjuvant to both MG and US, in case the latter do not allow for an accurate diagnosis. In this respect, German guidelines are mostly consistent with international guidelines from OECD countries such as Italy, Norway, the United Kingdom and the United States (Associazione Italiana di Oncologia Medica 2018; Helsedirektoratet 2018; NICE 2018; The American Society of Breast Surgeons 2018).

Following the guidelines, the Federal Joint Committee regulates the provision of outpatient breast MRIs by enacting under which conditions an outpatient MRI is reimbursed by the SHI.[2] Firstly, the provider needs to meet certain structural quality requirements. These comprise equipment and training requirements as well as minimum volume standards. Secondly, there need to be certain medical indications. According to the doctor's fee schedule position (GOP) "GOP 34431 - MRI(s) of the female mammary gland" in the reimbursement catalogue, breast MRI is billable only for two indications. Either, "to rule out a relapse (at the earliest 6 months after surgery or 12 months after completion of radiation therapy) of a histologically secured mamma carcinoma after breast conserving therapy, also after reconstruction surgery, in case previous MG and US could not clarify the dignity of the suspected relapse". Or, "to search for a primary tumor in case of axillary lymph node metastases (in case their histological morphology does not rule out a mamma carcinoma), if the primary tumor could neither be described clinically, nor with MG and US" (Federal Association of Statutory Health Insurance Physicians 2021; Federal Joint Committee 2020a). Conversely, for all other medical indications, breast MRI would be provided non-adherent to the reimbursement conditions of the SHI. This study focuses on these medical indications.

Comparing the German clinical practice guidelines and the reimbursement conditions shows that the latter are slightly more stringent than the former. For screening of women with an average lifetime risk of BC, MRI is neither recommended by guidelines, nor adherent to reimbursement conditions. Only for screening of women with a high lifetime risk, MRI is recommended by guidelines, but is non-adherent to reimbursement conditions.[3] For primary

[2] MRI of the female breast according to magnetic resonance imaging agreement following §135 Abs. 2 SGB V

[3] However, it should be noted that sickness fund customized integrated care programs (IGV) on breast cancer prevention - which target high-risk women - account for this. Hence, breast MRIs, which are provided for high-risk women who participate in such IGV, may nevertheless be considered as adherent to the reimbursement conditions.

diagnosis and for diagnosis of a potential relapse in the course of surveillance, both guidelines and reimbursement conditions, state that an MRI is only indicated if the tumor cannot clearly be described with conventional imaging, which is MG and US. According to guidelines, this is a sufficient condition. For the MRI to be adherent to the reimbursement conditions, however, this is only a necessary condition. In case the MRI is provided for primary diagnosis, women additionally need to have an axillary lymph node metastases diagnosis and in case it is provided for surveillance, a mandatory waiting period of 6 (12) months after surgery (radiation therapy) has to be complied with for an MRI to be adherent. The reason given for the latter is that in case of a relapse, new tumor cells need a certain time to grow to be detected by the MRI.

For the year 2016, which corresponds to our treatment period, the Federal Joint Committee suspended the obligation of the Associations of Statutory Health Insurance Physicians (ASHIPs) to audit outpatient physicians that provide MRI scans, which also includes breast MRI (Federal Joint Committee 2015).[4] Furthermore, data privacy restrictions, the ASHIPs' lack of cross-sectoral data and the resulting information asymmetries enable physicians to disregard the reimbursement conditions. However, with administrative data, we are able to identify these non-conform breast MRIs. Attending physicians may refer patients and radiologists may subsequently provide the MRI even though they do not adhere to the reimbursement conditions. The attending physician may refer the patient for a variety of reasons. He may be convinced that for his patient MRI is clinically indicated, despite knowing that the reimbursement conditions are not met at that time, or he may be unaware of the reimbursement conditions in the first place. Asymmetric information between attending physician and patient regarding the treatment process may also serve as a potential reason. Physicians may fear malpractice liability cost and thus practice defensive medicine, they may be subject to peer effects or feel pressured

[4] For every outpatient physician in Germany, who provides services for the SHI, a membership in at least one of the 17 regional ASHIPs is compulsory. By law, the 17 ASHIPs are obligated to annually audit a random sample of at least 4% of their members that provide MRI scans in order to assure the quality of outpatient MRI scans (§ 135b Sec. 2 SGB V, § 136 Sec. 2 SGB V, § 92 Sec. 1 SGB V). The auditions comprise the verification of regulations, which apply to MRI scans in general (for instance, requirements regarding the quality of the image), but also the verification of requirements that are specific to selected services, such as the reimbursement conditions for breast MRI. The physicians have to justify their practice with the help of patient records. If the audits reveal any inconsistencies, in the worst case, physicians, amongst other things, are not paid for the service or have to refund payments they have already received. For the year 2016, the Federal Joint Committee for the first time suspended the obligation to audit MRI scans because of a planned refinement of the quality requirements for MRI scans in general and because the results from the previous annual auditions were satisfactory (Federal Joint Committee 2015). While the refinements also targeted data privacy restrictions, the medical indications for breast MRI remained unaffected. Thereupon, 12 of the 17 ASHIPs paused the auditions, 2 ASHIPs (North Rhine, Rhineland-Palatinate) continued the auditions with a reduced sample size and only 3 ASHIPs (Baden-Wuerttemberg, Saarland, Saxony-Anhalt) continued the auditions with a sample size of at least 4%. To what extent breast MRI scans in particular were still audited by the remaining 5 ASHIPs is not clear because the published quality report for 2016 only reports the number of audited physicians that provide MRI scans in general but not the concrete service. Yet, one of the 5 auditing ASHIPs raised severe quality concerns for breast MRI (Federal Association of Statutory Health Insurance Physicians 2017, p. 82).

by patients' expectations (Baicker et al. 2007; Chandra and Staiger 2007; Little et al. 2004). Not least, the method of payment may influence the attending physicians' referral practice. In Germany, most outpatient care physicians, particularly including gynecologists, are predominantly reimbursed by capitation. Thus, they have an incentive to minimize the number of consultations in a given period as well as diagnostic effort (Gosden et al. 2001). They may refer the patient directly for an MRI instead of pursuing the sequential approach and adhering to the waiting period, as this may be time-consuming[5] Once the attending physician referred the patient, radiologists may provide a non-conform MRI for several reasons. Due to asymmetric information, they may not know whether billing conditions are truly met. Furthermore, radiologists bear high fixed cost for the provision of breast MRIs and they belong to the few specialties of outpatient physicians that tend to be reimbursed by fee-for-service rather than by capitation. Hence, they benefit from each additional scan they bill. Lastly, the minimum volume standards may incentivize radiologists at the threshold to provide non-conform MRI scans.

4.3 Data

Our primary data source comprises administrative data from *Techniker Krankenkasse*, the largest sickness fund in Germany, covering more than 10 million insured across all parts of the country (Techniker Krankenkasse 2021). The dataset includes longitudinal patient-level information on socio-demographic status, direct medical cost, medical diagnoses, medical resource utilization across care sectors (i.a. inpatient stays and procedures, outpatient consultations and procedures, pharmaceutical prescriptions) and all-cause mortality from 2014 to 2018.[6] Our initial study sample consists of women aged 18 and older, who were continuously enrolled during the study-period. For each woman, the study-period lasts for four years: it starts two years prior to an individual index date and ends two years after. The former exclusion criterion of continuous enrollment does not apply to women who decease during the after period. This results in an initial study population of 3,633,742 women.

Moreover, we add county-level information from the German Federal Office for Building and Regional Planning and county- and region-level information from the Federal Association of

[5] In Germany, MG is commonly provided in centralized MG centers. Thus, a referral and a subsequent review of the results by the attending gynecologist is necessary. Most gynecologists provide US in their practice though, and they are reimbursed with a fee for the provision.
[6] In Germany, outpatient refers to office-based practices providing primary and specialist care.

Statutory Health Insurance Physicians to our data set in order to generate further conditioning variables. All information is retrieved for the year 2016, which is our treatment period.

4.3.1 Intervention and Control Group

The aim of our study is to examine the impact of a non-conform outpatient breast MRI, which was performed for the purpose of primary diagnosis or surveillance after primary BC, on women's cost and outcomes. Hence, we compare women who suffer from breast cancer and received a non-conform MRI, the intervention group, to a similar cohort of women who did not receive any MRI, the control group.

4.3.1.1 Identification of Non-Conform MRI Scans

In 2016, a total number of 4,687 outpatient breast MRIs were claimed for 4,400 women, whereby 4,130 women received exactly one MRI and 270 women received more than one MRI. Scans were provided by 361 different radiologists. For 3,864 scans we have additional information on the referring physician. 84% of these scans were ordered by gynecologists, 8% by general practitioners and the remaining 8% by other specialist physicians.

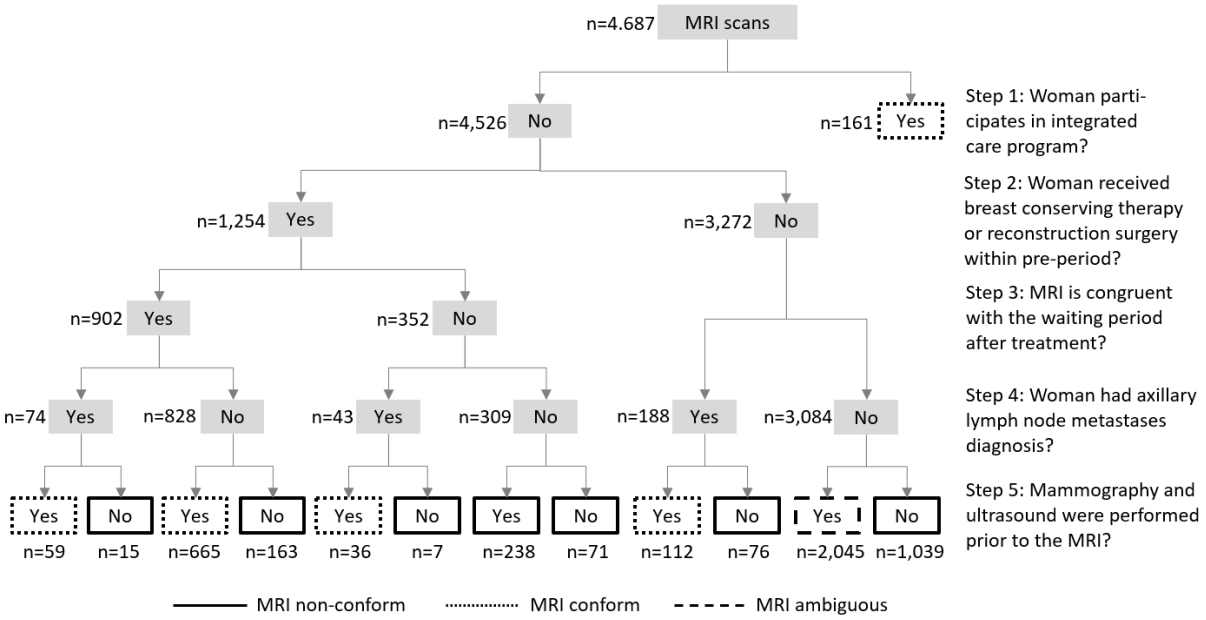
We compare actual clinical practice to the reimbursement conditions of the SHI to classify the observed MRI scans into three groups – non-conform, conform or ambiguous according to reimbursement conditions. Following doctor's fee schedule position "GOP 34431 - MRI(s) of the female mammary gland", the necessary conditions for a conform MRI scan are verified in five consecutive steps. Figure 4-1 depicts an overview of the classification. In case women participated in an integrated care program (IGV) for familial or genetic risk of breast cancer at the time of the MRI, we classify the MRI as conform (step 1). In case women did not participate in an IGV, we consider whether there is a medical indication which would justify the MRI: First, we check whether women received breast conserving therapy - that is breast conserving surgery or radiation therapy - or reconstruction surgery within two years prior to the MRI (step 2).[7] In this context, we also verify whether the mandatory waiting period was adhered to (step 3).[8] Second, we check whether a lymph node metastases diagnosis was documented prior to the MRI (step 4).[9]

[7] Breast conserving surgery and reconstruction surgery are operationalized through Operation and Procedure Code (OPS) 5870/5879 and OPS 5881/5882/5885/5886, respectively. For radiation therapy, we consider GOP 25320/25321/25310 and OPS 852 (German Institute of Medical Documentation and Information 2019b; Federal Association of Statutory Health Insurance Physicians 2021).

[8] As most women receive a sequence of radiation therapy, we always consider the first date of a sequence.

[9] A lymph node metastases diagnosis is identified using ICD-10-GM code C77 (German Institute of Medical Documentation and Information 2019a). Women needed to have ≥1 hospitalization or ≥1 outpatient claim with a secured diagnosis.

Figure 4-1: Classification of Outpatient MRI Scans



Note: This figure depicts the classification of the outpatient breast MRI scans. The classification results from the comparison of actual clinical practice to the reimbursement conditions of the SHI (see doctor's fee schedule position "GOP 34431 - MRI(s) of the female mammary gland").

The existence of one (or both) of the above medical indications is only a necessary condition though. For an MRI to be conform to reimbursement conditions, both MG and US must have additionally preceded the MRI (step 5).[10] Finally, adding up the cases from the thick solid boxes, we classify 34.3% (n=1,609) of the MRI scans as non-conform to the reimbursement conditions of the SHI. The majority of non-conform MRIs were classified as such, because the MRI was performed as a replacement for MG and/or US rather than as a complement (n=1,293). While in 868 (131) of these cases only MG (US) was missing prior to the MRI, in 294 cases both conventional imaging exams were missing. In 238 cases, MRI was performed prematurely after surgery/radiation therapy and in 78 cases, neither the order of the imaging exams nor the waiting period was adhered to.[11] Summing up the cases from the dotted boxes, we classify 22.0% (n=1,033) of the MRI scans as conform. The remaining cases in the dashed box, 43.7% (n=2,045) of all MRIs, are classified as ambiguous. Given neither of the two medical indications were present (see step 2 and step 4), but women received both MG and US prior to the MRI, it

[10] MG (US) is operationalized through GOP 34270/34272/34275 and OPS 3100/14943 (GOP 33041, OPS 3036).
[11] As detailed in Section 4.2, in 2016, one-third of the ASHIPs continued to audit the provision of MRI scans. Of the 4,687 MRI scans in our initial sample, 1,601 were provided by members from the auditing ASHIPs. When excluding these scans, we still find that of the remaining 3,086 MRI scans, which were performed by members from the non-auditing ASHIPs, 35.2% (n=1,086) may be classified as non-conform to the reimbursement conditions. This quota confirms the quota of non-conform MRI scans in our initial study sample (34.3%).

remains unclear whether women received breast conserving therapy or reconstruction surgery even prior to our observation period, in which case the MRI would also be conform.

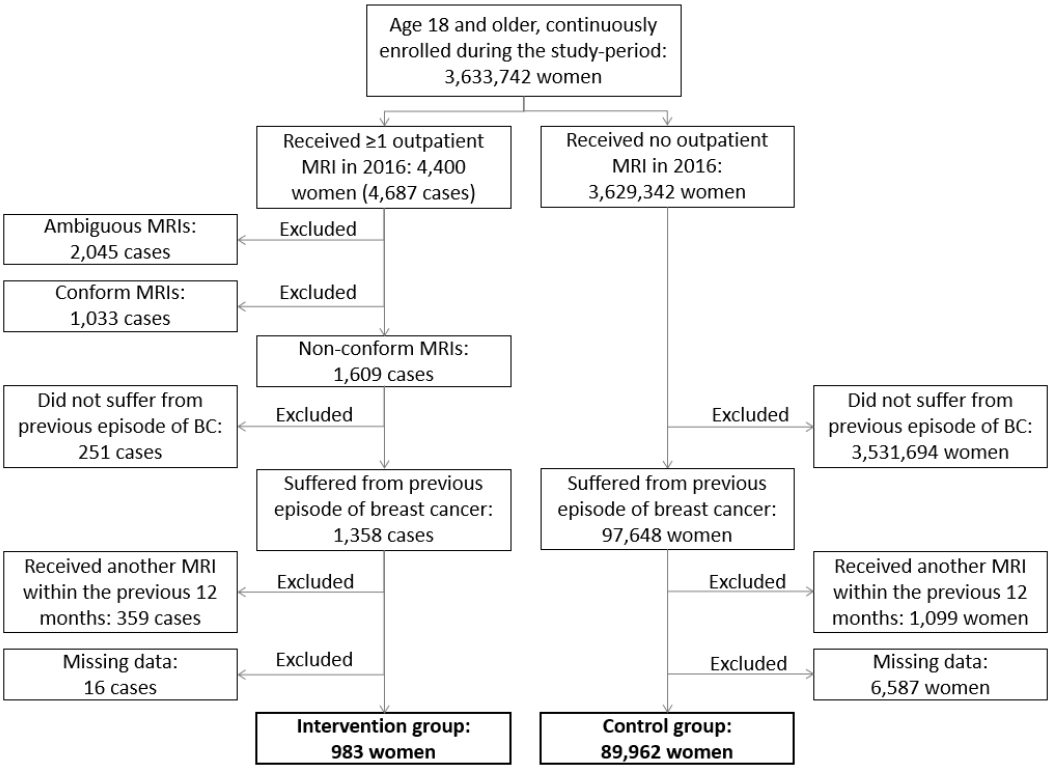
4.3.1.2 Assignment of Intervention and Control Group

An overview of the assignment of the intervention and control group is provided in Figure 4-2. The left-hand side of the figure shows how the intervention group is selected. We are interested in the contribution of an MRI to primary diagnosis and to surveillance BC. Therefore, we only focus on women for whom an episode of BC was documented prior to the MRI index date. Women with incident BC are identified using ICD-GM-10 diagnosis code "C50". They needed to have ≥1 hospitalization or ≥2 outpatient claims within one year with the secured diagnosis "C50" in primary position (Abraha et al. 2018; German Institute of Medical Documentation and Information 2019a). Due to data limitations, we can only observe medical diagnoses for up to two years prior to the MRI index date. Women, for whom a BC diagnosis was only documented earlier than two years prior to the index date are identified through their participation in the BC disease-management-program (DMP). Information on DMP participation is available since 2008.[12] Moreover, to isolate the effect of a single non-conform outpatient MRI, we only consider women who did not receive another MRI within 12 months prior to the index date and, finally, we must exclude 16 women due to missing data. This gives us a total number of 983 women, who serve as our intervention group. The assignment of the control group is depicted on the right-hand side. We choose women who did not receive an outpatient MRI in 2016 and should not have received an MRI given the medical indications.[13] Because there is no actual date of provision, a fictitious index date is fixed on December 31st, 2016. To ensure homogeneity between the two groups, we only consider women for whom an episode of BC was documented before the index date and who did not receive another in – or outpatient MRI within 12 months prior to the index date. 89,962 women meet these inclusion criteria.

[12] It should be noted that DMP participation is optional and BC diagnoses may not be coded reliably. Therefore, we may miss some women who suffered from BC prior to the index date. However, the inclusion criterion ensures a certain degree of homogeneity within the intervention group as well as between intervention and control group. In 251 of 1,609 cases (15.6%), no BC indication is documented prior to the MRI index. A possible explanation could be that these MRIs were performed for the purpose of screening.

[13] See Section 4.6 for a discussion on the choice of the control group.

Figure 4-2: Assignment of Intervention and Control Group



Note: The figure depicts the assignment of the intervention and the control group.

4.3.2 Study Outcomes

In this study, we consider a total of 27 outcomes, which were chosen based on the most frequently used outcomes in related studies and the German guidelines on BC treatment. Outcomes may be grouped into five categories: *(i)* direct medical cost of care, *(ii)* indicators for medical resource consumption, *(iii)* indicators for women's personal well-being, *(iv)* breast cancer specific procedures and *(v)* survival. In categories *(i)*, *(ii)*, *(iii)* and *(v)* outcomes are assessed over a follow-up period of two years, while for outcomes in category *(iv)* a follow-up period of 90 days is chosen.[14] Cost and countable medical resource consumption outcomes were calculated by averaging the annual values over the two year pre- and the two-year post-period, respectively. Binary outcomes were assessed over the complete two-year periods.

[14] Guidelines from professional societies do not make any concrete recommendations regarding the time from diagnosis to treatment; recommended timeframes vary from "several days" to "a few weeks". Analyzing more than 200,000 BC patients from the SEER-Medicare Cohort and the National Cancer Database in the United States, Bleicher et al. (2016) find that within 90 (60, 30) days after the diagnosis more than 98% (94%, 70%) of all women had surgery, which is often the first treatment procedure. Hence, a follow-up period of 90 days constitutes a compromise between capturing the initial treatment that follows the MRI and delimiting the effect of the MRI. In addition, the timeframe was validated by a gynecologist practicing in a German hospital.

(i) Direct medical cost of care: In order to capture the impact of a non-conform MRI on medical expenditures, we account for the total cost of care as well as the cost that arise in the inpatient, outpatient and pharmaceutical care sector, respectively.[15] Cost are calculated from the sickness fund perspective, reported in 2018 Euros and adjusted for inflation (European Central Bank 2019).

(ii) Indicators for medical resource consumption: We determine women's medical resource consumption in the three care sectors listed above.[16] This allows to further unravel the cost. For each sector, we chose indicators for the extensive as well as intensive all-cause and BC specific resource consumption: the share of patients with at least one hospitalization (due to BC), the number of hospital days (due to BC), the share of patients who consulted an outpatient physician at least once (due to BC), the number of outpatient consultations (due to BC), the share of patients with at least one pharmaceutical prescription (due to BC) and the number of pharmaceutical prescriptions (due to BC).[17]

(iii) Indicators for women's personal well-being: Based on the objectives of the German clinical practice guideline on BC treatment and related studies, we investigate the effect on women's personal well-being (Guideline Programs Oncology 2017, p. 246; Spiegel et al. 2011). Because women's personal well-being may not be captured directly with administrative data, we make use of two indicators: the share of patients with at least one depressive episode and the share of patients for whom at least one sickness absence was documented (Eaker et al. 2011; Mausbach and Irwin 2017).[18,19]

(iv) Breast cancer specific procedures: We assess BC specific diagnostic and treatment procedures that may follow an MRI in the short term (see, i.a., Houssami et al. 2017). For the

[15] For the non-conform MRI group, the cost for the MRI are included in the outpatient cost (€228.86 in 2016).
[16] Achelrod et al. (2016) use smililar indicators to examine the effect of the German DMP for chronic obstructive pulmonary disease (COPD).
[17] A hospitalization is categorized as BC specific in case ICD "C50" was reported as discharge diagnosis. As outpatient physicians are partially paid by capitation, the reported number of consultations may be viewed as a lower bound of the true number. A consultation is categorized as BC specific on the following conditions: the patient visited a gynecologist and a secured ICD C50 diagnosis and a surcharge for oncological care (GOP 08345) were recorded in the respective quarter of the visit, or the patient visited an oncologist and a secured ICD C50 diagnosis was recorded in the respective quarter, or the patient saw any outpatient physician in the course of the BC DMP. BC specific pharmaceutical prescriptions are identified using the Anatomical Therapeutic Chemical Codes (ATC) for anti HER2, chemo- and hormone therapy (see below).
[18] We follow Quan et al. (2005) and operationalize a depressive episode with ICD F20.4/F31.3F31.5/F32/F33/ F34.1/F41.2/F43.2. An episode is recorded, if a woman had ≥1 hospitalization or ≥2 outpatient claims with a secured diagnosis within one year.
[19] Information on sickness absences is only available for women who are employed or unemployed, but not for women who are self-employed, students, retired or receive welfare benefits. Therefore, this information is available for 58% of our study sample (see Table 4-8 in the Appendix).

former, we report the shares of patients receiving a biopsy and further imaging.[20] For the latter, we report the shares of patients receiving local and systemic treatments. Usually BC patients receive some combination of treatments (see Waks and Winer 2019 for a review). Most women get local treatment, that is breast conserving surgery, mastectomy and/or radiation therapy, to remove the tumor. This is complemented by (neo)adjuvant systemic treatments such as anti HER2, chemo- and/or hormone therapy.[21] Assessing the BC specific procedures allows for identifying differences in the treatment courses. Taking into consideration the entirety of treatment options also allows to determine the cancer yield after the non-conform MRI and to identify potentially false-positive findings.

(v) Survival: We further measure patient outcomes by capturing the effect of a non-conform MRI on all-cause mortality.

4.4 Empirical Strategy

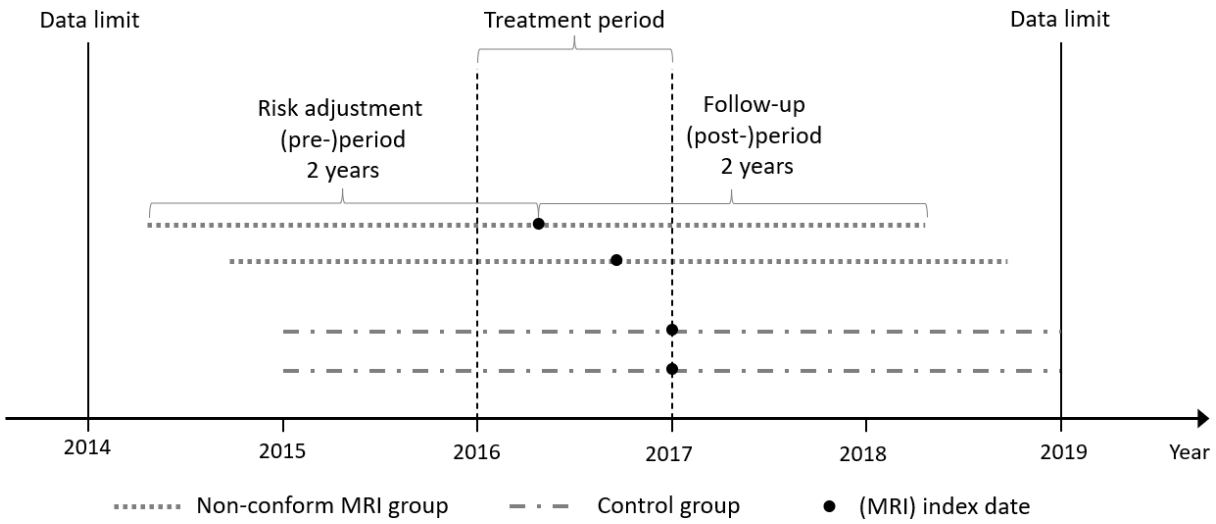
Due to the retrospective cohort study design with administrative data, the provision of the non-conform MRI did not occur at random. This may result in selection bias and unbalanced baseline characteristics in our patient cohorts. Thus, to isolate the causal effect of a non-conform MRI on outcomes – the average treatment effect on the treated – we apply two risk-adjustment techniques. For outcomes, for which a pre-post comparison is meaningful, we apply a combination of Propensity Score Matching (PSM) and Difference-in-Difference (DiD) to account for observed as well as time-invariant unobserved heterogeneity. For outcomes, for which a pre-post comparison is not expedient, we only apply PSM.

4.4.1 Risk Adjustment with Propensity Score Matching

The propensity score (PS) is defined as an individual's probability to be treated, conditional on a vector of observed baseline characteristics. The rationale is that adjusting for the scalar PS is sufficient to control for imbalance in all the observed baseline characteristics such that, based on the values of the propensity score, homogeneous matched sets of treated and untreated individuals may be formed (Austin 2009; Rosenbaum and Rubin 1983). Figure 4-3 shows a timeline of our study period.

[20] Biopsy is identified by GOP 08320 and OPS 1493/1494/1501, imaging by GOP 34431 and OPS 3807/3827.
[21] Breast conserving surgery is operationalized through OPS 5870/5879, mastectomy through OPS 5872/5874/ 5877, anti HER2 therapy through ATC L01XC03/L01XC07/L01XC13/L01XE07/L01XE10/L01XE33/L01XE42/ L04AA18, chemotherapy through GOP 86514/86516/86520, OPS 8542/ 8543/8544/8547 and ATC L01AA01/ L01BC05/L01BC06/L01CA04/L01DB01-L01DB11, hormone therapy through ATC L02AE01-L02AE05/ L02BA01-L02BA03/L02BG01L02BG06/L02BG09, radiation therapy through GOP 25320/ 25321/25310 and OPS 852.

Figure 4-3: Timeline of the Study Period



Note: This figure shows a timeline of our study period. Over a period of two years prior to the (MRI) index date, we measure baseline characteristics for risk-adjustment and over a follow-up period of up to two years post the index date, we assess the outcomes. For the non-conform MRI group (dotted lines), the index date varies as it corresponds to the actual date of provision of the MRI. For the control group (dashed lines), a fictitious index date is fixed on December 31st, 2016 as these individuals did not receive any MRI scan. Each line depicts an exemplary individual.

We measure outcomes and baseline characteristics for risk-adjustment over a period of two years prior to the (MRI) index date. Further, we assess the outcomes over a follow-up period of up to two years post the index date. For the intervention group, the index date varies as it corresponds to the actual date of provision. For the control group, a fictitious index date is fixed on December 31st, 2016.[22] We estimate the propensity score by fitting a logistic regression model in which the probability to receive a non-conform MRI is the dependent variable. As covariates, we choose a vast list of variables that we assume to be predictive of both treatment assignment and outcomes (Stuart 2010). On the individual level, we account for socio-demographic indicators such as age, person group, insurance status and insurance category.[23]

[22] For the control group, the data do not indicate an actual (MRI) index date of provision as women in this group did not receive any MRI scan. Hence, we chose December 31st, 2016, as a fictitious index date. By choosing December 31st, we only account for control subjects who also did not receive any other MRI scans within 12 months prior to the (fictitious) index date and, therefore, treatment and control group are mutually exclusive.

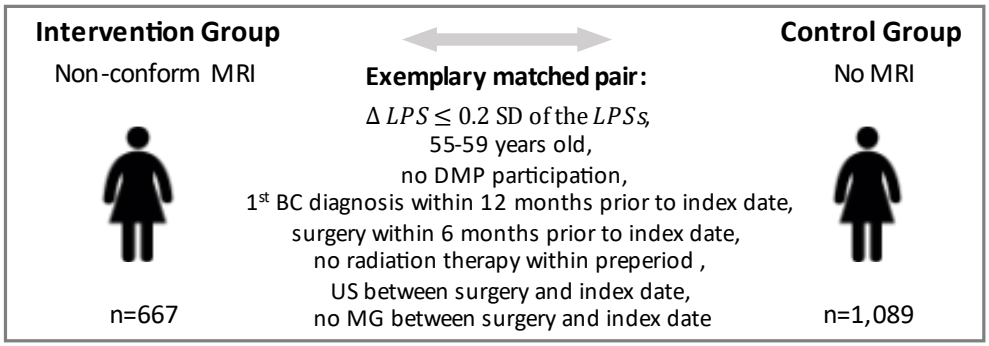
[23] Person group indicates whether individuals are employed, self-employed, a student, unemployed, a welfare recipient or a pensionary. Insurance status specifies whether individuals are a contributing member of the SHI or are covered by family members. Insurance category refers to individuals' enrolment in SHI as being mandatory, which is the case for employees below a certain income threshold, or voluntary, which is the case for, i.a., employees above this threshold and self-employed persons. Put simply, for historical reasons, higher-income individuals in Germany may choose between SHI and fully substitutive private health insurance. By interacting employee with insurance category, we can identify higher-income individuals (in 2016, with an annual gross income of at least €56,250). This serves as an indicator for individuals' socio-economic status.

Because comorbidities are found to have a great impact on BC outcomes, we control for DMP participation other than BC, Elixhauser comorbidity groups and pharmacy-based classes (Hong et al. 2015; Kuo et al. 2011; Quan et al. 2005). We further adjust for BC specific indicators by considering DMP participation, type of BC diagnosis, timing of the primary diagnosis, lymph node metastases diagnosis as well as diagnostic and treatment indicators, such as breast conserving surgery, mastectomy, anti HER2/chemo-/hormone/ radiation therapy, a running waiting period after breast conserving therapy, receipt of biopsy, MG, US and another in- or outpatient MRI.[24] In order to reproduce the trends in outcomes between non-conform MRI and control group, we further match on the pre-period levels of the outcomes (Ryan et al. 2015). The three months prior the index date are particularly predictive of the treatment assignment; therefore, we consider the pre-period levels in outcomes over segregated periods of 0-3, 4-12 and 13-24 months prior to the index date. Lastly, we account for county- and region-level information, such as average household income, share of inhabitants with an academic degree, number of hospital beds, information on gynecological and radiological care as well as on settlement structure. To ensure adequate model specification, we further include quadratic and interaction terms. Table 4-7 in the Appendix displays a full list and definitions of the covariates.

We apply a 2:1 ratio for greedy nearest neighbor matching without replacement. Matches were chosen within 0.2 standard deviations of the logit of the propensity score (LPS) and over the common support region only. 2:1 matching allows for increased precision and a threshold of 0.2 is found to perform well when at least some of the covariates are continuous (Austin 2010, 2011). Moreover, we match exactly on covariates that are assumed to be highly predictive of the outcomes and the treatment assignment: 5-year age groups, BC DMP participation, timing of the primary BC diagnosis, breast conserving surgery, radiation therapy, a running waiting period after surgery/radiation therapy as well as previous MG and US. Each subject from the intervention group is sequentially matched with up to the two nearest control subjects that satisfy the caliper and the exact matching criteria. This may not result in all intervention subjects being matched. In this study, 667 intervention subjects (68% of the total) are matched with 1,089 control subjects. Thus, our final study sample consists of 1,756 subjects. For an exemplary illustration, see Figure 4-4. Before applying PSM, the distributions of the measured baseline covariates differed significantly in 80 out of 187 variables. After PSM, the final sample is well-balanced over all covariates (see Table 4-8 in the Appendix).

[24] In- and outpatient MRI are operationalized through OPS 3807/3827 and GOP 34431, respectively.

Figure 4-4: Exemplary Matched Pair



Note: This figure shows an exemplary match of an intervention and a control subject. Generally, matches were chosen within 0.2 standard deviations (SD) of the logit of the propensity score (LPS). In this example, both women are between 55 and 59 years old, do not participate in the BC DMP, their first BC diagnosis was documented within 12 months prior to the index date, both had breast-conserving surgery within 6 months prior to the index date, they did not receive radiation therapy, and both had US between surgery and the (MRI) index date but not MG. The intervention subject received an MRI (which is classified as non-conform because the waiting period after surgery was not adhered to and the prior MG is missing) and the control subject did not receive an MRI and, according to reimbursement conditions, should not have. LPS: Logit of the propensity score, SD: standard deviation. Source: own representation.

4.4.2 Estimations

To estimate the impact on *(i)* direct medical cost, *(ii)* medical resource consumption and *(iii)* women's personal well-being, we apply a DiD approach. This allows us to additionally control for differences in outcomes between the non-conform MRI and control group that are due to unobserved heterogeneity.

The DiD model is given by:

$$Outcome_{ijkt} = \alpha_0 + \beta_1 MRI_{ijk} + \beta_2 post_t + \beta_3 change_{ijkt} + \beta_4 X_{ijk} + \epsilon_{ijkt} \qquad (1)$$

where $Outcome_{ijkt}$ refers to the outcomes in categories *(i)* - *(iii)* in Section 4.3.2 for individual i in county j in region k in time-period t. We run the model separately for each outcome. α_0 is a constant. MRI_{ijk} is a dummy variable that takes the value one if individual i in county j in region k received a non-conform MRI and $post_t$ is a dummy variable that takes the value one in the post-period. $change_{ijkt}$ is a dummy variable that takes the value one for all individuals that received a non-conform MRI and where observation took place in the post-period. Hence, coefficient β_3 captures the effect of a non-conform MRI on individuals' outcomes. X_{ijk} is a vector that contains the set of conditioning variables from the matching step and is added in order to reduce variance in the outcome and to make the estimates more precise. ϵ_{ijkt} is an error

term. Due to the serial correlation of our observations, this error term might be correlated (Bertrand et al. 2004). Hence, we collapse the data from the bi-annual risk adjustment as well as the follow-up period by taking averages. We are left with only two time periods: pre- and post-intervention period. To account for remaining correlation, standard errors are clustered at the individual-level. We estimate the model from equation (1) with weighted least squares (WLS). The weights are retrieved from the matching step. Because cost and countable medical resource consumption outcomes have rightly skewed distributions and observations contain zeros, outcomes are transformed by taking the natural logarithm of the outcome plus one $\left(log\left(1 + Outcome_{ijkt}\right)\right)$. Thus, these effects are expressed in percent. To account for cost outliers, cost outcomes are further winsorized at the 99% percentile. The effect on the dichotomous indicators for medical resource consumption and women's personal well-being may be interpreted as changes in the linear probability $Pr\left(Outcome_{ijkt} = 1\right)$. A drawback of linear probability models (LPM) is that they may predict probabilities outside the interval [0,1]. As we are only interested in the DiD coefficient though, this may be neglected. We estimate the models with heteroscedasticity robust standard errors because the LPMs' error term is heteroscedastic by definition,

For the outcomes in category *(iv)*, BC specific procedural outcomes, a pre-post comparison is not expedient. Therefore, we choose a weighted LPM to estimate the effect of non-conform MRI. The LPM model takes the following form:

$$Pr\left(Outcome_{ijk} = 1\right) = \alpha_0 + \beta_1 MRI_{ijk} + \beta_2 X_{ijk} + \epsilon_{ijk} \qquad (2)$$

where $Outcome_{ijk}$ refers to the BC specific procedures listed in Section 4.3.2 for individual i in county j in region k. α_0 is a constant and MRI_{ijk} a dummy variable that takes the value one in case individual i in county j in region k received a non-conform MRI. X_{ijk} is a vector that contains the conditioning variables from the matching step and ϵ_{ijk} is an error term.

Lastly, we perform survival analyses using the Kaplan-Meier method, to test whether a non-conform MRI has an effect on patients' mortality (outcome category *v*).

4.5 Results

In this section, we present the results of our analyses on the effect of a non-conform MRI scan on cost and outcomes. Initially, we will show the main results, including our preferred specification. Subsequently, we will discuss the robustness of our main findings by accounting

for a potential misclassification of non-conform MRI exams, limitations of the risk adjustment method, the common trends assumption, which is crucial for the application of the DiD approach, the follow-up period and for subsamples of our original sample. Finally, we conduct subgroup analyses to identify heterogeneous treatment effects.

4.5.1 Main Results

Estimation results from the DiD and linear probability regressions are presented in Table 4-1. Column 1 displays the results from a basic estimation without risk-adjustment or additional covariates. Column 2 reports the estimates after applying PSM and Colum 3 shows the estimates after additionally including the conditioning variables from the matching step, as specified in equations (1) and (2), respectively. This expanded specification is our preferred one as it yields precise estimates and, thus, serves as foundation for our further analyses. Comparing Columns 2 and 3 shows that the ATTs remain mostly unchanged, as, due to the successful matching algorithm, the additional conditioning variables are mean-independent of the intervention indicator. Yet, adding further conditioning covariates leads to increased variation and smaller standard errors and, therefore, improves the precision of the estimates. Analyzing Column 1 reveals that the estimates on medical cost and resource consumption tend to be smaller than in Column 3, while the effects on BC specific procedures remain mostly unaffected. Explaining factors include the difference in average age between the basic and PSM control group (63.77 vs. 56.55 years). Our main results suggest that the provision of a non-conform MRI on average leads to a significant increase of 26.7% (p<0.01) in the total direct medical cost of care compared to the control group. This corresponds to an average absolute increase of €1,717 per year in the two-year follow-up period.[25] Stratification by care sector reveals that inpatient (+73.8%, p<0.01), outpatient (+25.6%, p<0.01) as well as pharmaceutical cost (+23.6%, p<0.05) contribute to the increase in total cost. We further find that medical resource consumption is more intense in the non-conform MRI group, reflecting the growing cost. While the share of patients with at least one hospitalization (due to BC) grows by 10.5 ppts. (8.8 ppts.), the number of hospital days (due to BC) increases by 16.2% (15.4%). All effects on medical resource consumption in the inpatient sector are significant at the one percent level.

[25] Table 4-9 in the Appendix shows the outcome means for both groups in the pre- and post-period. Women in the non-conform MRI group, for instance, on average caused total direct medical cost of €8,149 per year in the post-period. Computing the absolute change yields an average increase of €1,717 per year. Consequently, over the two-year follow-up period, this sums up to average additional direct medical cost of care of €3,434 for each woman who suffered from breast cancer and received a non-conform MRI scan compared to similar women who did not receive any MRI scan.

Table 4-1: Main Results – Outcome Categories *(i)-(iv)*

	(1) Basic Model		(2) + PSM		(3) + Controls	
(i) Direct medical cost						
Total cost	0.098*	(0.051)	0.267***	(0.080)	0.267***	(0.057)
Inpatient cost	0.194	(0.179)	0.738***	(0.274)	0.738***	(0.217)
Outpatient cost	0.146***	(0.040)	0.256***	(0.063)	0.256***	(0.044)
Pharmaceutical cost	0.145	(0.093)	0.236*	(0.143)	0.236**	(0.097)
(ii) Medical resource consumption						
Hospitalization†	0.039*	(0.023)	0.105***	(0.035)	0.105***	(0.028)
Thereof BC specific†	-0.002	(0.018)	0.088***	(0.026)	0.088***	(0.020)
Hospital days	0.011	(0.046)	0.162**	(0.070)	0.162***	(0.053)
Thereof BC specific	-0.002	(0.034)	0.154***	(0.047)	0.154***	(0.037)
Outpatient consultation†	0.001***	(0.000)	0.002*	(0.001)	0.002*	(0.001)
Thereof BC specific†	0.029	(0.021)	0.036	(0.034)	0.036*	(0.021)
Outpatient consultations	0.067**	(0.027)	0.086**	(0.042)	0.086***	(0.026)
Thereof BC specific	0.075*	(0.043)	0.121*	(0.065)	0.121***	(0.035)
Pharmaceutical prescription†	0.009	(0.007)	0.013	(0.012)	0.013	(0.011)
Thereof BC specific†	0.053**	(0.023)	0.034	(0.034)	0.034**	(0.017)
Pharmaceutical prescriptions	0.059	(0.039)	0.074	(0.063)	0.074**	(0.034)
Thereof BC specific	0.106***	(0.035)	0.067	(0.051)	0.067***	(0.025)
(iii) Well-being						
Depressive episode†	0.018	(0.022)	0.004	(0.034)	0.004	(0.019)
Sickness absence†	0.009	(0.023)	0.003	(0.035)	0.003	(0.012)
Observations	181,890		3,512		3,512	
(iv) BC specific procedures						
Biopsy†	0.037***	(0.006)	0.045***	(0.008)	0.045***	(0.008)
Further imaging†	0.012***	(0.004)	-0.007	(0.008)	-0.009	(0.007)
Breast conserving surgery †	0.082***	(0.009)	0.084***	(0.012)	0.087***	(0.011)
Mastectomy†	0.032***	(0.006)	0.031***	(0.008)	0.030***	(0.007)
Anti HER2 therapy†	0.001	(0.001)	-0.006*	(0.003)	-0.006***	(0.002)
Chemotherapy†	0.011*	(0.006)	0.005	(0.010)	0.013*	(0.008)
Hormone therapy†	0.154***	(0.012)	-0.030	(0.021)	-0.009	(0.015)
Radiation therapy†	0.027***	(0.007)	0.026***	(0.010)	0.024***	(0.008)
Observations	90,945		1,756		1,756	

Note: This table shows the effect of a non-conform MRI on *(i)* the direct medical cost of care, *(ii)* indicators for medical resource consumption and on *(iii)* women's personal well-being over a follow-up period of two years as well as on *(iv)* breast cancer specific procedures over 90 days after the non-conform MRI. For outcomes in categories *(i)-(iii)*, each entry displays the ATT of a separate regression and the respective standard error, clustered at the individual level, in parentheses. For outcomes in category *(iv)*, each entry displays the ATT of a separate regression and the respective heteroscedasticity robust standard error in parentheses. Column 1 presents the results from a basic DiD and a basic LPM model without risk-adjustment or further conditioning variables for outcomes in categories *(i)-(iii)* and in category *(iv)*, respectively. Column 2 reports the DiD (LPM) results after risk-adjustment with PSM and Column 3 presents the DiD (LPM) estimators after additionally including the conditioning variables from the matching step, as specified in equation (1) and in equation (2), respectively. Cost and countable medical resource consumption outcomes are transformed using the natural logarithm of the outcome plus one. Cost are reported in 2018 Euros, adjusted for inflation and winsorized at the 99% percentile. If applicable, regression models are weighted by the weights from the matching step. † The outcome represents a share. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data: Techniker Krankenkasse.

Moreover, the share of patients with at least one outpatient consultation slightly increases (+0.2 ppts., p<0.1) and the number of consultations is significantly higher (+8.6 %, p<0.01). When only considering the share of patients with at least one consultation (+3.6 ppts, p<0.1) and the number of consultations (+12.1%, p<0.01) due to BC, effects further increase in magnitude. A greater share of patients receives at least one pharmaceutical prescription due to BC (+3.4 ppts., p<0.05) and the number of total prescriptions and prescriptions due to BC increase by 7.4 % (p<0.05) and 6.7% (p<0.01), respectively. Effects on the share of patients receiving at least one prescription and on women's personal well-being, by contrast, are found to be small and insignificant.

The shares of patients receiving local diagnostic and treatment procedures such as a biopsy (+4.5 ppts.), breast conserving surgery (+8.7 ppts.), mastectomy (+3.0 ppts.) and radiation therapy (+2.4 ppts.) are higher in the non-conform MRI than in the control group. Effects are significant at the one percent level and explain the elevated resource utilization in the in- and outpatient sector. We find no significant effect on the probability to receive further imaging though. Regarding systemic treatments, estimates show mixed results. While the patient share that receives Anti HER2 therapy decreases by 0.6 ppts. (p<0.01), the share receiving hormone therapy remains unaffected and the share that receives chemotherapy increases by 1.3 ppts. (p<0.1).

Results from the survival analyses are displayed in Table 4-2. Column I reports the Kaplan-Meier survival estimates without risk adjustment, whereas Column II shows the estimates after applying PSM. The difference in survival probabilities between the non-conform MRI and the basic control group is larger and more significant than between the non-conform MRI and the PSM control group. Again, the difference in age between the groups may serve as an explanation. Our main results show that for women in the non-conform MRI group, the two-year all-cause survival probability is 1.3 ppts. higher than for women in the control group. The stratified Log-Rank test for equality of survival functions rejects the null hypothesis and indicates a significant difference in survival functions (p=0.062). The corresponding Kaplan-Meier survival curves are depicted in Figure 4-5 in the Appendix. Relating the findings in the different categories reveals that while health care utilization is more intense, the effect on patients' all-cause survival is beneficial.

Table 4-2: Main Results – Outcome Category *(v)* Survival

	(I) No risk-adjustment		(II) Main specification	
Non-conform MRI	0.986	(0.004)	0.990	(0.004)
Control	0.958	(0.001)	0.977	(0.005)
Difference	0.030***		0.013*	
Log-Rank test	$p<0.001$		$p=0.062$	
Observations	90,945		1,756	

Note: This table displays the Kaplan-Meier survival estimates for the non-conform MRI and the control group. The corresponding survival standard errors are presented in parentheses. The effect of a non-conform MRI on all-cause mortality is assessed over a follow-up period of two years. Column I reports the results without risk adjustment and Column II shows the estimates after applying PSM. If applicable, regression models are weighted by the weights from the matching step. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data: Techniker Krankenkasse.

4.5.2 Robustness Checks

Results from several robustness checks are presented in Table 4-3 and in Table 4-4.[26] Firstly, we check whether a misclassification of non-conform MRI scans affects our results. We particularly focus on the coding of lymph node metastases diagnoses as the latter are not relevant to billing and, therefore, the attending physicians' documentation may not be reliable. While, in the worst case, this would lead to misclassification of 238 non-conform MRI scans (compare Figure 4-1), our results remain robust when excluding the respective scans from the analyses (see Column 4 in Table 4-3 and Column III in Table 4-4).

Secondly, to test the robustness of the risk adjustment method, we apply Entropy Balancing (EB) as an alternative to PSM. EB is a reweighting mechanism to balance the observed covariates of observational data (Hainmueller 2012). We use EB to reweight observations from the control group until appropriate balance is achieved between non-conform MRI and control group regarding the mean and the variance of the set of covariates that we accounted for by PSM approach (see Section 4.4.1).[27] By contrast to PSM, no observations are discarded. We find that the results are robust to this alternative risk adjustment method (see Column 5 in Table 4-3 and Column IV in Table 4-4). Only the effect on the share of patients receiving hormone therapy is notably larger and significant. This may be driven by differences in sample size and would suggest that we underestimate the effect in our original specification.

[26] In the light of the large number of 27 outcomes and widely varying sample sizes, we consider the overall picture of the coefficients when assessing whether the results are robust. While in all our robustness analyses the signs of the coefficients remain unchanged (expect for very few exemptions with (insignificant) coefficients close to zero), the significance level changes in a few cases. However, in case the significance changes, coefficients are very similar in terms of size. Thus, the strict criteria of both unchanged signs and significance is not always applicable.
[27] After the application of EB, the standardized mean difference was zero for all covariates.

Table 4-3: Robustness – Outcome Categories *(i)-(iv)*

	(3) Main results		(4) Lymph node metastases coding		(5) Entropy Balancing		(6) Placebo regressions		(7) Categories *(i)-(iii)*: simple WLS / category *(iv)*: 91-181 days follow-up		(8) Without the deceased		(9) Without breast cancer DMP	
(i) Direct medical cost														
Total cost	0.267***	(0.057)	0.251***	(0.082)	0.265***	(0.046)	0.021	(0.059)	0.267***	(0.044)	0.278***	(0.057)	0.251***	(0.058)
Inpatient cost	0.738***	(0.217)	0.631**	(0.283)	0.837***	(0.179)	-0.418**	(0.194)	0.572***	(0.173)	0.764***	(0.218)	0.637***	(0.222)
Outpatient cost	0.256***	(0.044)	0.258***	(0.064)	0.245***	(0.036)	0.098*	(0.050)	0.231***	(0.034)	0.263***	(0.044)	0.261***	(0.045)
Pharmaceutical cost	0.236**	(0.097)	0.206	(0.148)	0.254***	(0.084)	-0.139	(0.093)	0.329***	(0.077)	0.280***	(0.096)	0.207**	(0.100)
(ii) Medical resource consumption														
Hospitalization†	0.105***	(0.028)	0.094**	(0.037)	0.110***	(0.024)	0.028	(0.023)	0.065***	(0.023)	0.107***	(0.029)	0.091***	(0.029)
Thereof BC specific†	0.088***	(0.020)	0.079***	(0.025)	0.091***	(0.017)	0.026	(0.016)	0.079***	(0.014)	0.093***	(0.020)	0.089***	(0.020)
Hospital days	0.162***	(0.053)	0.131*	(0.073)	0.165***	(0.047)	0.086*	(0.049)	0.128***	(0.043)	0.177***	(0.053)	0.142***	(0.054)
Thereof BC specific	0.154***	(0.037)	0.131***	(0.046)	0.133***	(0.036)	0.056*	(0.033)	0.131***	(0.028)	0.172***	(0.036)	0.152***	(0.037)
Outpatient consultation†	0.002*	(0.001)	0.002*	(0.001)	0.001***	(0.000)	-0.001	(0.002)	0.002*	(0.001)	0.002	(0.001)	0.002*	(0.001)
Thereof BC specific†	0.036*	(0.021)	0.027	(0.036)	0.027*	(0.016)	0.051**	(0.020)	0.027	(0.017)	0.032	(0.021)	0.036	(0.022)
Outpatient consultations	0.086***	(0.026)	0.084*	(0.044)	0.088***	(0.020)	0.019	(0.027)	0.075***	(0.022)	0.083***	(0.025)	0.086***	(0.026)
Thereof BC specific	0.121***	(0.035)	0.105	(0.066)	0.085***	(0.029)	0.131***	(0.039)	0.108***	(0.030)	0.115**	(0.035)	0.123***	(0.036)
Pharmaceutical prescription†	0.013	(0.011)	0.015	(0.013)	0.014**	(0.008)	0.011	(0.013)	0.026***	(0.008)	0.015	(0.011)	0.014	(0.011)
Thereof BC specific†	0.034**	(0.017)	0.029**	(0.034)	0.032**	(0.013)	0.023	(0.015)	0.047***	(0.014)	0.036**	(0.017)	0.034**	(0.018)
Pharmaceutical prescriptions	0.074**	(0.034)	0.065	(0.066)	0.081***	(0.029)	-0.039	(0.034)	0.090***	(0.029)	0.090***	(0.033)	0.062*	(0.034)
Thereof BC specific	0.067***	(0.025)	0.064	(0.052)	0.054**	(0.021)	0.039	(0.025)	0.084***	(0.022)	0.066***	(0.026)	0.069***	(0.026)
(iii) Well-being														
Depressive episode†	0.004	(0.019)	0.007	(0.036)	0.013	(0.016)	-0.008	(0.015)	-0.008	(0.017)	0.002	(0.019)	0.006	(0.020)
Sickness absence†	0.003	(0.012)	0.002	(0.037)	0.003	(0.016)	0.008	(0.018)	0.008	(0.016)	0.002	(0.020)	0.002	(0.020)
Observations	3,512		3,230		181,890		4,422		1,756		3,446		3,380	
(iv) BC specific procedures														
Biopsy†	0.045***	(0.008)	0.044***	(0.009)	0.037***	(0.006)	0.007	(0.006)	0.004	(0.004)	0.045***	(0.008)	0.047***	(0.008)
Further imaging†	-0.009	(0.007)	-0.011	(0.008)	-0.017**	(0.007)	0.000[a]	(0.000)	0.004	(0.004)	-0.010	(0.007)	-0.009	(0.008)
Breast conserving surgery †	0.087***	(0.011)	0.074***	(0.012)	0.072***	(0.008)	0.018*	(0.009)	0.001	(0.005)	0.086***	(0.011)	0.089***	(0.011)
Mastectomy†	0.030***	(0.007)	0.026***	(0.007)	0.031***	(0.005)	0.001	(0.005)	0.007**	(0.003)	0.030***	(0.007)	0.032***	(0.007)
Anti HER2 therapy†	-0.006***	(0.002)	-0.005*	(0.003)	0.001	(0.001)	0.003	(0.002)	-0.004	(0.002)	-0.003*	(0.002)	-0.006***	(0.002)
Chemotherapy†	0.013*	(0.008)	0.001	(0.010)	0.007	(0.006)	-0.012	(0.007)	0.012	(0.008)	0.015**	(0.007)	0.013	(0.008)
Hormone therapy†	-0.009	(0.015)	-0.002	(0.021)	0.062***	(0.014)	0.010	(0.012)	0.039***	(0.014)	-0.009	(0.015)	-0.006	(0.015)
Radiation therapy†	0.024***	(0.008)	0.028***	(0.009)	0.021***	(0.006)	0.017**	(0.007)	0.015**	(0.006)	0.024***	(0.008)	0.024***	(0.008)
Observations	1,756		1,615		90,945		2,211		1,756		1,723		1,690	

Note: The table shows the results for six robustness checks. Specifications for all Columns, except for Column 6, are based on our preferred DiD and LPM models from equation (1) and equation (2), respectively. Samples and the set of covariates are altered as described in Section 4.5.2. Estimates in Column 3 coincide with the estimates in Column 3 of Table 4-1 and report the results from our preferred specification. Column 4 depicts the results when excluding non-conform MRIs scans, which would be conform to reimbursement conditions if the attending physician had coded a lymph node metastases diagnosis. Column 5 reports the estimates when Entropy Balancing is applied as an alternative risk adjustment method to PSM. Column 6 shows the results from placebo regressions. We conducted the placebo regressions with newly matched control subjects, whereby we matched on the one-year pre-period only. Apart from that, we used the same set of covariates as in our main analyses. After matching, the placebo sample is well-balanced over all covariates. For outcomes in categories *(i)* - *(iii)*, Column 7 reports the estimates from a simple WLS regression without DiD and post indicators, while for outcomes in category *(iv)*, Column 7 reports the estimates for an alternative follow-up period of 91-180 days after the intervention. Columns 8 and 9 show the results for subsamples of our original sample without the deceased and without DMP participants, respectively. For outcomes in categories *(i)-(iii)* and in category *(iv)*, the reported standard errors in parentheses are clustered at individual level and are heteroscedasticity robust, respectively. Cost and countable medical resource consumption outcomes are transformed using the natural logarithm of the outcome plus one. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. Regression models are weighted by the weights from the risk adjustment step. † The outcome represents a share. [a] There is no variation, as, by definition of the study inclusion criteria, women received no other MRI scan within 12 months prior to the index date. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Data: Techniker Krankenkasse.

Table 4-4: Robustness – Outcome Category *(v)* Survival

	(II) Main specification		(III) Lymph node metastases coding		(IV) Entropy Balancing		(V) Without BC DMP	
Non-conform MRI	0.990	(0.004)	0.989	(0.004)	0.986	(0.004)	0.991	(0.004)
Control	0.977	(0.005)	0.978	(0.005)	0.969	(0.005)	0.977	(0.005)
Difference	0.013*		0.011*		0.017**		0.014**	
Log-Rank test	p=0.062		p=0.085		p=0.048		p=0.041	
Observations	1,756		1,615		90,945		1,690	

Note: This table shows the survival estimates for three robustness checks. The respective survival standard errors are reported in parentheses. Samples are altered as described in Section 4.5.2. Estimates in Column II coincide with the estimates in Column II of Table 4-2 and report the results from our preferred specification. Column III depicts the results when excluding non-conform MRIs scans, which would be conform to reimbursement conditions if the attending physician had coded a lymph node metastases diagnosis. Column IV reports the estimates when Entropy Balancing is applied as an alternative risk adjustment method to PSM. Column V shows the results for a subsample of our original sample without DMP participants. Regression models are weighted by the weights from the risk adjustment step. *** p<0.01, ** p<0.05, * p<0.1. Data: Techniker Krankenkasse.

Our analyses further rely on the crucial assumption that all relevant covariates are included in the risk adjustment step such that heterogeneity due to observable factors is removed. Compared to clinical data, administrative data only provide limited information on a patient's clinical condition, which may affect our outcomes. We could alleviate this limitation by additionally controlling for tumor characteristics that are documented for BC DMP participants and are frequently used for risk adjustment in related clinical studies (i.a. Debald et al. 2015; Fortune-Greeley et al. 2014). We cannot add these characteristics to the set of covariates because only part of our sample (n=379 [non-conform MRI=148, control=231]) consists of present or past DMP participants. Instead, for the subsample of DMP participants, we test whether average differences in the tumor characteristics between non-conform MRI and control group differ significantly from zero. Applying Fisher's exact test, we find that there are no significant differences in the spread and size of the primary tumor (p≥0.67), regional lymph node involvement (p≥0.29), pathological tumor grading (p≥0.60), hormone receptor status (p≥1.00), manifestation of distant metastases (p≥0.44) and manifestation of local recurrences (p≥0.50). See Table 4-10 in the Appendix for an overview.

To establish a causal relation in a DiD approach, the common trends assumption must hold. The assumption states that, in absence of the intervention, the intervention group would have followed the same trend in the outcome as the control group (Angrist and Pischke 2009). The common trends assumption is visually inspected, and plots are depicted in Figure 4-6 in the Appendix. As a more formal robustness check, we additionally run placebo tests by testing for an effect of the intervention one year earlier, while the post-period is limited to one year.

Column 6 in Table 4-3 shows that almost all placebo estimates are closer to zero and less significant than our main results. Solely the estimates on inpatient cost and BC specific outpatient consultation(s) stand out. However, this may be relativized by the fact that the placebo sample contains more observations and a larger share of zeros for the mentioned outcomes than the original sample. The latter makes hypotheses testing less valid. Moreover, it is mandatory to consult an outpatient physician for referral purposes before receiving an MRI. This may explain the increased BC specific consumption in the outpatient sector.

Further, we particularly address the estimation of *(iv)* the BC specific outcomes. A potential concern may be that a follow-up period of 90 days is not sufficiently long to capture the effect of an MRI. Treatment may be delayed because of comorbidities, patient preferences or access restrictions (Bleicher et al. 2016). Column 7 in Table 4-3 presents the results for an alternative follow-up period of 91-180 days. Except for the share of patients receiving hormone therapy, estimates are closer to zero and less significant than our original findings. Hormone therapy is usually only administered after surgery or completion of chemotherapy (Guideline Programs Oncology 2017), thus the effect may set in with a delay. Moreover, the DiD approach cannot be applied for the estimation of the BC specific outcomes and, therefore, time-invariant unobserved heterogeneity cannot be accounted for. This may distort the estimates.

Yet, conducting simple WLS instead of DiD regressions for *(i)* the cost, *(ii)* medical resource consumption and *(iii)* personal well-being outcomes reveals that the results remain virtually unchanged (see Column 7 in Table 4-3). Thus, we conjecture that the distortion of the BC specific outcomes is also negligible. In addition, it could be argued that, instead of a linear probability model, a logit model should be estimated. However, running the alternative logit regressions and comparing the results confirms the signs and significance levels of our original estimates.

Lastly, we conduct sensitivity analyses by considering subsamples of our original sample. Firstly, we omit all women who died in the post-period. The rationale is that at the end of life, patients cause exceptionally high medical expenses, which may distort the estimates (Riley and Lubitz 2010). Running the analyses without the deceased (n=43 [non-conform MRI=17, control=26]) yields that estimates remain virtually unchanged (see Column 8 in Table 4-3). Secondly, we exclude women who participated in the BC DMP at the time of the index date (n=66 [non-conform MRI=27, control=39]). The DMP aims at establishing structured and coordinated treatment pathways as well as increasing patient participation and adherence (Federal Joint Committee 2020b). Hence, the MRI may have a particular effect on women who

participate in the DMP, which, in turn, would affect our results. Column 9 in Table 4-3 as well as Column V in Table 4-4 present the estimates for the subsample of non-participants. Magnitudes and significance levels remain stable. Thirdly, we conducted the analyses without MRI scans, which were provided by members of the auditing ASHIPs. The prospect of possibly being audited may influence the physicians' selection of patients, for whom they provide (non-conform) MRIs. This, in turn, may affect our results. Yet, the estimates confirm our main results, and the reader is referred to Table 4-11 in the Appendix.

4.5.3 Heterogeneous Effects

We perform three subgroup analyses. Firstly, we stratify our sample by purpose of the MRI in the course of BC treatment, secondly, by reason of non-conformity and, thirdly, by settlement structure in women's residential area. Results are presented in Table 4-5 and in Table 4-6.

In the course of BC treatment, MRIs may be provided for two different purposes: for primary diagnosis or for surveillance after treatment of primary BC.[28] Depending on the purpose, effects on the future treatment course may differ. Results indicate that the effect of a non-conform MRI on cost, medical resource consumption and BC specific procedures in the primary diagnosis subgroup (n=177 [non-conform MRI=63, control=114]) tends to be stronger than in the surveillance subgroup (n=1,579 [non-conform MRI=604, control=975). For the former subgroup, almost all coefficients are larger in magnitude than for the latter. However, some of those coefficients are less significant or even insignificant. This should be seen in the light of the considerably smaller sample size though. Findings may be explained by differences in pre-period levels between subgroups and by the fact that, in the beginning of a treatment course, a non-conform MRI may have a greater impact on cost and outcomes than at later stages because costly invasive procedures and medication are usually administered upfront.[29] The difference in survival probabilities in the primary diagnosis subgroup is insignificant, whereas women who received a non-conform MRI for the purpose of surveillance on average benefit from an increased survival probability of 1.2 ppts. (p=0.080) compared to the respective control group.

[28] We consider an MRI to be applied for primary diagnosis, if the first BC diagnosis was documented within one year in advance of the (MRI) index date and neither breast conserving surgery nor radiation therapy preceded the MRI. An MRI is defined to be provided for surveillance if the earliest BC diagnosis was documented more than one year in advance of the (MRI) index date, or if the earliest BC diagnosis was documented within one year in advance of the index date but breast conserving surgery or radiation therapy preceded the MRI.
[29] Descriptive statistics for the subgroup analyses are reported Tables 4-12 to 4-14 in the Appendix.

Table 4-5: Heterogeneous Effects – Outcome Categories *(i)-(iv)*

	(10) MRI for primary diagnosis		(11) MRI for surveillance		(12) MRI premature		(13) MRI as replacement		(14) Urban area residents		(15) Rural area residents	
(i) Direct medical cost												
Total cost	0.420	(0.278)	0.251***	(0.083)	0.488**	(0.228)	0.261***	(0.082)	0.216**	(0.092)	0.449***	(0.164)
Inpatient cost	1.846**	(0.824)	0.623**	(0.287)	1.891**	(0.775)	0.635**	(0.288)	0.495	(0.310)	1.66***	(0.584)
Outpatient cost	0.322	(0.243)	0.249***	(0.064)	0.204	(0.159)	0.277***	(0.063)	0.215***	(0.072)	0.395***	(0.128)
Pharmaceutical cost	0.726	(0.546)	0.185	(0.146)	0.521	(0.500)	0.196	(0.149)	0.199	(0.164)	0.361	(0.226)
(ii) Medical resource consumption												
Hospitalization†	0.246**	(0.105)	0.090**	(0.037)	0.226**	(0.099)	0.094**	(0.038)	0.076*	(0.04)	0.214***	(0.076)
Thereof BC specific†	0.310***	(0.089)	0.065**	(0.026)	0.095	(0.077)	0.093***	(0.023)	0.084***	(0.029)	0.105*	(0.556)
Hospital days	0.423**	(0.211)	0.135*	(0.073)	0.543**	(0.214)	0.127*	(0.073)	0.107	(0.078)	0.370**	(0.154)
Thereof BC specific	0.560***	(0.165)	0.112**	(0.048)	0.316*	(0.174)	0.154***	(0.042)	0.144***	(0.052)	0.192*	(0.107)
Outpatient consultation†	0.008	(0.008)	0.002	(0.001)	0.000 [a]	(0.000)	0.003*	(0.002)	0.003*	(0.002)	0.000[a]	(0.000)
Thereof BC specific†	0.183*	(0.107)	0.021	(0.035)	0.119	(0.084)	0.026	(0.037)	0.058	(0.038)	-0.041	(0.071)
Outpatient consultations	0.173	(0.164)	0.077*	(0.043)	0.103	(0.115)	0.086*	(0.045)	0.069	(0.049)	0.152*	(0.085)
Thereof BC specific	0.451**	(0.180)	0.086	(0.067)	0.179	(0.231)	0.106	(0.067)	0.168**	(0.073)	-0.047	(0.136)
Pharmaceutical prescription†	0.056	(0.056)	0.008	(0.012)	-0.000	(0.017)	0.014	(0.014)	0.018	(0.014)	-0.009	(0.024)
Thereof BC specific†	0.286***	(0.074)	0.007	(0.036)	0.024	(0.112)	0.036	(0.035)	0.034	(0.038)	0.035	(0.073)
Pharmaceutical prescriptions	0.169	(0.227)	0.064	(0.065)	0.168	(0.181)	0.068	(0.068)	0.058	(0.072)	0.128	(0.128)
Thereof BC specific	0.444***	(0.116)	0.028	(0.055)	0.068	(0.175)	0.078	(0.054)	0.069	(0.058)	0.062	(0.110)
(iii) Well-being												
Depressive episode†	0.024	(0.100)	0.002	(0.036)	-0.060	(0.136)	0.014	(0.037)	0.020	(0.039)	-0.058	(0.073)
Sickness absence†	-0.008	(0.110)	0.004	(0.037)	0.012	(0.139)	0.005	(0.038)	0.001	(0.04)	0.011	(0.075)
Observations	354		3,158		218		3,004		2,768		744	
(iv) BC specific procedures												
Biopsy†	0.071**	(0.035)	0.042***	(0.008)	0.071*	(0.040)	0.046***	(0.009)	0.051***	(0.010)	0.026**	(0.013)
Further imaging†	0.127***	(0.045)	-0.021***	(0.007)	0.024	(0.024)	-0.011	(0.009)	-0.004	(0.009)	-0.017	(0.017)
Breast conserving surgery †	0.357***	(0.068)	0.055***	(0.011)	0.119**	(0.050)	0.076***	(0.013)	0.093***	(0.015)	0.054**	(0.021)
Mastectomy†	0.071**	(0.035)	0.026***	(0.007)	0.048	(0.033)	0.026***	(0.008)	0.026***	(0.008)	0.048***	(0.018)
Anti HER2 therapy†	0.008	(0.018)	-0.007**	(0.003)	-0.024	(0.023)	-0.005	(0.003)	-0.006*	(0.004)	-0.004	(0.004)
Chemotherapy†	0.000	(0.057)	0.006	(0.009)	0.048	(0.033)	0.003	(0.010)	0.009	(0.011)	-0.010	(0.024)
Hormone therapy†	0.103*	(0.054)	-0.014	(0.023)	-0.048	(0.095)	-0.002	(0.021)	-0.008	(0.024)	0.018	(0.046)
Radiation therapy†	0.198***	(0.057)	0.007	(0.008)	0.024	(0.024)	0.029***	(0.009)	0.033***	(0.011)	-0.001	(0.017)
Observations	177		1,579		109		1,502		1,384		372	

Note: The table shows the results for three subgroup analyses. We stratify the sample by the purpose of the MRI in the course of BC treatment, by the reason of non-conformity of the MRI and by the settlement structure in women's residential area, respectively. Samples are altered as described in Section 4.5.3. By forming the subgroups, we split up the initially matched sample, which in turn may result in biased estimates. However, the criteria, by which we defined the primary diagnosis, surveillance, premature and replacement subgroups, were also included as exact matching criteria. Thus, we do not split-up those matched pairs. We further checked the standardized mean differences for all subgroups and the differences were smaller than <0.2, with a few exceptions merely being smaller than <0.3. Hence, the patient cohorts are still well-balanced, while we do not lose any observations. As a further robustness check, we also conducted the analyses with newly matched subgroups and the results are consistent with the findings presented in this table, yet sample sizes are smaller resulting in increased variance. Generally, specifications are based on our preferred regression model from equation (1) and equation (2), respectively, with the regression results provided in Column 3 of Table 4-1. Yet, due to the uneven distribution of observations in the subgroups and the resulting sample sizes, outcomes are estimated without covariates X_{ijk}. For outcomes in categories *(i)-(iii)* we report standard errors clustered at the individual-level, and for outcomes in category *(iv)* we report heteroscedasticity robust standard errors. Cost and countable medical resource consumption outcomes are transformed using the natural logarithm of the outcome plus one. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. Regression models are weighted by the weights from the matching step. † The outcome represents a share. [a] There is no variation because all patients have at least one outpatient consultation during the pre- and during the post-period. *** p<0.01, ** p<0.05, * p<0.1. Data: Techniker Krankenkasse.

Table 4-6: Heterogeneous Effects – Outcome Category *(v)* Survival

	(VI) MRI for primary diagnosis	(VII) MRI for surveillance	(VIII) MRI premature	(IX) MRI as replacement	(X) Urban area residents	(XI) Rural area residents
Non-conform MRI	0.968 (0.022)	0.992 (0.004)	1.000 (0.000)	0.988 (0.005)	0.992 (0.004)	0.981 (0.011)
Control	0.944 (0.022)	0.980 (0.005)	0.952 (0.028)	0.978 (0.005)	0.977 (0.005)	0.977 (0.011)
Difference	0.024	0.012*	0.048*	0.010	0.015**	0.004
Log-Rank test	p=0.470	p=0.080	p=0.091	p=0.198	p=0.037	p=0.836
Observations	177	1,579	109	1,502	1,384	372

Note: The table shows the Kaplan-Meier survival estimates for three subgroup analyses. We stratify the sample by the purpose of the MRI in the course of BC treatment, by the reason of non-conformity of the MRI and by the settlement structure in women's residential area, respectively. Samples are altered as described in Section 4.5.3. The respective survival standard errors are reported in parentheses. Regression models are weighted by the weights from the matching step. *** p<0.01, ** p<0.05, * p<0.1. Data: Techniker Krankenkasse.

Secondly, we stratify the analyses by reason of non-conformity of the MRI. A breast MRI may be classified as non-conform for two reasons: it was provided prematurely after breast conserving or reconstruction therapy and/or in case it served as a replacement for conventional imaging (compare Section 4.3.1.1.).[30] Depending on the reason of non-conformity, the effects of the MRI may be heterogeneous. Running the analyses separately for both subgroups suggests that, compared to the replacement subgroup (n=1,502 [non-conform MRI=571, control=931]), in the premature subgroup (n=109 [non-conform MRI=42, control=67]) a non-conform MRI has a greater effect on cost and resource consumption in the inpatient sector as well as on invasive BC specific procedures. This also applies to the effect on total cost. For the outpatient sector a reverse picture emerges as resource consumption and cost more strongly increase for the replacement subgroup. Most notably, women who received the MRI prematurely benefit from an increased all-cause survival probability of 4.8 ppts. (p=0.091) compared to the control group, whereas we find no significant effect for the replacement subgroup. Apparently, in some cases, physicians rightly presume that a premature MRI is medically indicated and, therefore, reimbursement conditions are disregarded. This may explain the increased utilization in the inpatient sector shortly after the premature MRI and the increased survival probability compared to the replacement subgroup.

Thirdly, we stratify the sample by the settlement structure in women's residential area.[31] Women who live in urban areas may have better access to healthcare services than women who

[30] In 54 cases, the MRI is classified as non-conform because it was performed prematurely and as a replacement. We exclude these cases from the subgroup analysis as we cannot assign them unambiguously.

[31] We use the county classification based on settlement structure of the German Federal Office for Building and Regional Planning. Sparsely and moderately populated counties are defined as rural areas and urban counties and big cities are defined as urban areas.

live in rural areas, which, in turn, may influence the estimates. Results show that in the rural subgroup (n=372 [non-conform MRI=154, control=218]), a non-conform MRI causes a larger rise in cost than in the urban subgroup (n=1,384 [non-conform MRI=513, control=871]), particularly in the inpatient sector. This is also reflected by the effects on medical resource consumption. By contrast, outpatient consultations due to BC more strongly increase in the urban subgroup. These findings may be traced back to differences in pre-period levels between the subgroups as well as to regional variation in BC treatment patterns (Xu et al. 2016).[32] Results further suggest that, in the urban subgroup, a non-conform MRI leads to a sharper increase in the shares of women receiving biopsy, breast conserving surgery and radiation therapy than in the rural subgroup. On the contrary, the effect on the share of women receiving mastectomy is weaker. The higher share of women receiving radiation therapy may be explained by the greater density of radiology practices in urban areas and the share of women receiving breast conserving surgery, which often entails supplemental radiation therapy (Feinstein et al. 2013). We deduce that, for women residing in urban areas, treatment patterns after a non-conform MRI seem to be more firmly established, less aggressive and rather shifted to the outpatient sector compared to women living in rural areas. While for the rural subgroup we do not find a significant effect on mortality, for the urban subgroup, findings coincide with our main results.

4.6 Discussion

This study makes several contributions to the literature. Firstly, to the best of our knowledge, this paper constitutes the first study that investigates whether outpatient breast MRI scans in Germany are provided compliant with the reimbursement conditions of the SHI. Using claims data, we compare actual clinical practice to the reimbursement conditions of the SHI and find that 34.3% of the MRIs that were billed in 2016 may be classified as non-conform. Secondly, we comprehensively investigate the effect of a non-conform MRI on cost and outcomes. Our main results suggest that women in the non-conform MRI group on average cause significantly higher direct medical cost of care (+26.7%) in the two-year follow-up period than women in the control group. Mirroring the elevated cost, (BC specific) medical resource consumption increases. Women in the non-conform MRI group are more likely to receive additional invasive diagnostic and treatment procedures within a follow-up period of 90 days after the intervention.

[32] Differences in age may also serve as a potential explanation. The average age in the subgroups is similar though.

Concomitantly, women in the non-conform MRI group benefit from an increased survival probability of 1.3 ppts. compared to women in the control group.

The classification of the non-conform MRI scans is made conservatively: According to reimbursement conditions, a necessary condition is that prior MG and US could not sufficiently describe the tumor. Whether MG and US were indeed insufficient cannot be verified with claims data though. Thus, we always assume that, if performed prior to the MRI, MG and US were insufficient and the adjuvant MRI, therefore, conform. Moreover, we always consider the first date of a sequence of radiation therapy when checking whether the mandatory waiting period after radiation therapy was complied with. A sequence on average takes 4-6 weeks. Further, we cannot observe whether women received breast-conserving therapy prior to our study period. Hence, we classified MRIs, which were preceded by MG and US but without documentation of breast conserving therapy, as ambiguous. In case no breast conserving therapy preceded, the MRI would also be non-conform (compare Figure 4-1). By contrast, some MRIs may be falsely classified as non-conform. Besides the argument that lymph node metastases may not be coded reliably, we cannot preclude that patients received prior MG or US as individual health services (IGeL). IGeL are not covered by the SHI and hence not documented in our data. Nevertheless, we conjecture that this is negligible as IGeL are predominantly provided in the scope of screening and not for diagnosis or treatment (Medical Service of German Statutory Health Insurance providers 2018). Finally, our quota of non-conform MRIs is consistent with the findings from, i.a. Parmar et al. (2013) who analyze Medicare data and conclude that 35% of all breast MRI scans provided for surveillance are non-adherent to U.S. guidelines.

By considering a total of 27 outcomes across various outcome categories, this study comprehensively investigates the effect of a non-conform MRI. By contrast, most of the related studies only consider outcomes from a single category. Regarding the effect on cost and medical resources consumption outcomes, our results are widely congruent with the findings from other international studies, which examined the effects of breast MRI scans in the course of breast cancer treatment in general (see, i.a., Bedrosian et al. 2016; Brennan et al. 2010; Houssami et al. 2017; Lee and Houssami 2016; Padia et al. 2016; Wang et al. 2016). While we find no significant effect on women's personal well-being, the evidence on this relation is mixed (Brédart et al. 2012; Spiegel et al. 2011). It should be kept in mind though that the determinability of well-being with administrative data is rather limited. Most notably, our results imply that women in the non-conform MRI group benefit from an increased survival probability of 1.3 ppts. compared to women in the control group. In view of the two-year

survival rate from the German life tables, which was 99.1% for women at age 58 for the years 2016-2018, the effect size is notable (German Federal Statistical Office 2019). Inspecting the trends in the survival curves further indicates that extending the follow-up period might be expedient. Clearly, it would be even more conclusive to examine the effect of a non-conform MRI on disease-specific survival. Unfortunately, due to data limitations, we cannot observe the cause of death. Hence, from our results, the effect of a non-conform MRI on disease-specific survival is unclear.[33] Other observational studies find no significant effect of breast MRI on disease-specific survival (i.a. Onega et al. 2018).

A further strength of our paper is that we employ a two-stage risk adjustment procedure for the analysis of the effect of a non-conform MRI on cost and medical resource consumption outcomes. Applying both PSM and DiD estimation, we create a quasi-experimental setting using administrative data. While related retrospective studies based on clinical data (see, for instance, Houssami et al. 2017; Lee and Houssami 2016; Padia et al. 2016) are typically associated with selection bias and baseline differences in the study population due to non-randomized MRI assignment, we attempt to alleviate these weaknesses with our approach.[34] Even though for the estimation of the BC specific outcomes we only apply PSM, we still account for a more comprehensive set of observed covariates than related studies (compare Fortune-Greeley et al. 2014). Nevertheless, a limitation of our risk adjustment approach is that, by contrast to clinical data, administrative data only provide limited information on a patients' clinical conditions. Physicians' decision to refer a patient for a (non-conform) MRI is particularly influenced by patients' tumor characteristics and breast density, as these factors are found to be predictive of the informational value of an MRI adjunct to conventional imaging (see Lehman et al. 2009 for a meta-analysis). While we showed above that there are no significant differences in the tumor characteristics that are documented for DMP participants, we cannot control for the entire set of relevant characteristics. Breast density is related to age, postmenopausal status, obesity and hormone intake (Boyd et al. 2010). We controlled for age and obesity in the PSM step, yet postmenopausal status and hormone intake cannot be directly

[33] German claims data do not document the cause of death. Nevertheless, in case a patient deceased in hospital, the diagnosis in primary position may provide an indication. We find that in the non-conform MRI group (control group), 71% (50%) of the deceased died in hospital and 40% (70%) of these had a BC diagnosis in primary position. Thus, at least 28% (35%) of the deceased in the non-conform MRI group (control group) likely died of breast cancer.

[34] Using a retrospective cohort study design with administrative data, Achelrod et al. (2017), Haas et al. (2012) and Stroka (2016), for instance, choose a similar approach to overcome selection bias and heterogeneity. They examine the effect of home telemonitoring for COPD patients on cost and outcomes, compare cost and postoperative complications of laparoscopic versus open appendectomy and estimate the effect of institutionalization on the drug intake of the elderly, respectively.

observed.[35] If we assumed that, despite risk adjustment, a selection bias remains, we would overestimate the effect on cost, medical resource consumption and BC specific procedures. Furthermore, we observe variation on referring physician and on radiologist level regarding the referral for and provision of non-conform, ambiguous and conform MRI scans. Thus, we surmise that patient preferences may also have an influence on the decision for an MRI, and results may further be affected by unobserved heterogeneity, such as women's personality traits. Yet, by applying DiD estimation for the majority of the outcomes, we are able to overcome the latter limitation and minimize confounding due to unobserved factors.

Certainly, the results are highly dependent on the choice of the control subjects. As control group for women who suffer from breast cancer and received a non-conform MRI, we choose similar women who did not receive any MRI and, according to reimbursement conditions, should not have at that point in time. It could also be argued that women who received a conform MRI should serve as control group. The rationale is that they would be more suitable with respect to the clinical conditions because clinical conditions supposedly influence the decision for (diagnostic) procedures. However, we aim at examining the effect of a non-conform MRI ceteris paribus, and, according to reimbursement conditions of the SHI, that would be no MRI. Besides, we only identified 1,033 MRI scans as conform and the resulting small sample size would lead to statistical problems (compare Figure 4-1).

Finally, we perform several robustness checks and demonstrate that the results remain stable when we account for a possible misclassification of non-conform MRI scans, when, instead of PSM, EB is used for risk adjustment, when we restrict the study sample and when an alternative follow-period is chosen (for the BC specific outcomes). Placebo tests also hold. Furthermore, we conduct subgroup analyses to identify heterogeneous effects. Results indicate that the subgroup of patients who received the MRI prematurely may comparatively benefit from the MRI because the survival probability is 4.8 ppts. higher in the non-conform MRI than in the control group. A potential limitation is the sample size though, especially when performing subgroup analyses. Yet, it should be emphasized, that this is the largest sample of administrative data available in Germany.

Judged by the reimbursement conditions of the SHI, it may well be concluded that more than one third of the breast MRI scans, which are provided for women suffering from breast cancer,

[35] Obesity is controlled for by an Elixhauser comorbidity group. We cannot directly observe women's post-menopausal status, however, women are 58 years old on average and we matched exactly on 5-year-age bins. Also, hormone intake cannot be fully observed. We can only control for part of the relevant substances, which are included in the set of substances we considered for the BC procedure "hormone therapy".

are subject to overprovision. As already argued, physicians may overprovide the scans for a variety of (clinical) reasons. The clinical reasons, in particular, either cannot be represented at all or can only be represented in part with administrative data. To reduce the non-conform provision of breast MRI scans, raising awareness among referring physicians, radiologists and patients may constitute a first fundamental measure. Complementing measures such as intensified monitoring with the help of claims data and regress claims would create even stronger (financial) incentives for physicians to forgo non-conform MRI scans.

Our analyses on the effect of a non-conform MRI rely on a number of assumptions and, ideally, RCTs should be conducted. Nevertheless, relating our findings in the different outcome categories yields that, for women who received a non-conform MRI, health care utilization is more intense, while the effect on women's survival is positive. Hence, considering cost and outcomes, it is not straightforward whether the overprovision of MRI scans is indeed ineffective and, therefore, should be targeted. Heterogeneous treatment effects further impede the judgement. Especially the subgroup of women who received the MRI prematurely may comparatively benefit from an increased survival probability and the decision prerogative of the physician. Therefore, we conclude that further research is needed to produce even stronger evidence on the effectiveness of the provision of breast MRI scans, which are non-conform to guideline-based reimbursement conditions and to form more concrete policy implications. Detailed (subgroup) analyses, merging administrative data with clinical data, cancer registries or cause of death statistics and essentially complementing the findings with cost-effectiveness-analyses may provide valuable insights.

References

Abraha, Iosief; Montedori, Alessandro; Serraino, Diego; Orso, Massimiliano; Giovannini, Gianni; Scotti, Valeria et al. (2018). Accuracy of administrative databases in detecting primary breast cancer diagnoses: a systematic review. BMJ Open 8 (7). DOI: 10.1136/bmjopen-2017-019264.

Achelrod, Dmitrij; Schreyögg, Jonas; Stargardt, Tom (2017). Health-economic evaluation of home telemonitoring for COPD in Germany: evidence from a large population-based cohort. The European Journal of Health Economics 18 (7), pp. 869–882. DOI: 10.1007/s10198-016-0834-x.

Achelrod, Dmitrij; Welte, Tobias; Schreyögg, Jonas; Stargardt, Tom (2016). Costs and outcomes of the German disease management programme (DMP) for chronic obstructive pulmonary disease (COPD)—A large population-based cohort study. Health Policy 120 (9), pp. 1029–1039. DOI: 10.1016/j.healthpol.2016.08.002.

Angrist, Joshua David; Pischke, Jörn-Steffen (2009). Mostly harmless econometrics. An empiricist's companion. Princeton: Princeton University Press.

Associazione Italiana di Oncologia Medica (2018). Neoplasie Della Mammella. Linee guida. Available online at https://www.aiom.it/wp-content/uploads/2018/11/2018_LG_AIOM_Mammella.pdf, checked on 2/28/2021.

Austin, Peter C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine 28 (25), pp. 3083–3107. DOI: 10.1002/sim.3697.

Austin, Peter C. (2010). Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. American Journal of Epidemiology 172 (9), pp. 1092–1097. DOI: 10.1093/aje/kwq224.

Austin, Peter C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical Statistics 10 (2), pp. 150–161. DOI: 10.1002/pst.433.

Baicker, Katherine; Fisher, Elliott S.; Chandra, Amitabh (2007). Malpractice liability costs and the practice of medicine in the Medicare program. Health Affairs (Project Hope) 26 (3), pp. 841–852. DOI: 10.1377/hlthaff.26.3.841.

Baxi, Shrujal S.; Kale, Minal; Keyhani, Salomeh; Roman, Benjamin R.; Yang, Annie; Derosa, Antonio P.; Korenstein, Deborah (2017). Overuse of Health Care Services in the Management of Cancer: A Systematic Review. Medical Care 55 (7), pp. 723–733. DOI: 10.1097/MLR.0000000000000734.

Bedrosian, Isabelle; Xing, Yan; Abdel Rahman, Shereen; Allen, Lisa; Le-Petross, Huong; Whitman, Gary J. et al. (2016). A Cost Analysis of Preoperative Breast MRI Use for Patients with Invasive Lobular Cancer. Annals of Surgical Oncology 23 (1), pp. 23–29. DOI: 10.1245/s10434-015-4710-1.

Bertrand, Marianne; Duflo, Esther; Mullainathan, Sendhil (2004). How Much Should We Trust Differences-In-Differences Estimates? The Quarterly Journal of Economics 119 (1), pp. 249–275. DOI: 10.1162/003355304772839588.

Bleicher, Richard J.; Ruth, Karen; Sigurdson, Elin R.; Beck, J. Robert; Ross, Eric; Wong, Yu-Ning et al. (2016). Time to Surgery and Breast Cancer Survival in the United States. JAMA Oncology 2 (3), pp. 330–339. DOI: 10.1001/jamaoncol.2015.4508.

Boyd, Norman F.; Martin, Lisa J.; Bronskill, Michael; Yaffe, Martin J.; Duric, Neb; Minkin, Salomon (2010). Breast tissue composition and susceptibility to breast cancer. Journal of the National Cancer Institute 102 (16), pp. 1224–1237. DOI: 10.1093/jnci/djq239.

Brédart, Anne; Kop, Jean-Luc; Fall, Mouhamadou; Pelissier, Sandra; Simondi, Cécile; Dolbeault, Sylvie et al. (2012). Anxiety and specific distress in women at intermediate and high risk of breast cancer before and after surveillance by magnetic resonance imaging and mammography versus standard mammography. Psycho-Oncology 21 (11), pp. 1185–1194. DOI: 10.1002/pon.2025.

Brennan, Sandra; Liberman, Laura; Dershaw, D. David; Morris, Elizabeth (2010). Breast MRI screening of women with a personal history of breast cancer. American Journal of Roentgenology 195 (2), pp. 510–516. DOI: 10.2214/AJR.09.3573.

Chandra, Amitabh; Staiger, Douglas O. (2007). Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks. Journal of Political Economy 115 (1), pp. 103–140. DOI: 10.1086/512249.

Debald, Manuel; Abramian, Alina; Nemes, Lisa; Döbler, Michael; Kaiser, Christina; Keyver-Paik, Mignon-Denise et al. (2015). Who may benefit from preoperative breast MRI? A single-center analysis of 1102 consecutive patients with primary breast cancer. Breast Cancer Research and Treatment 153 (3), pp. 531–537. DOI: 10.1007/s10549-015-3556-3.

Eaker, Sonja; Wigertz, Annette; Lambert, Paul C.; Bergkvist, Leif; Ahlgren, Johan; Lambe, Mats (2011). Breast cancer, sickness absence, income and marital status. A study on life situation 1 year prior diagnosis compared to 3 and 5 years after diagnosis. PloS One 6 (3), e18040. DOI: 10.1371/journal.pone.0018040.

European Central Bank (2019). Annual consumer price inflation (HICP).

Federal Association of Statutory Health Insurance Physicians (2017). Qualitätsbericht 2017. Berichtsjahr 2016. Berlin, p. 82. Available online at https://www.kbv.de/media/sp/KBV_Qualitaetsbericht_2017.pdf, checked on 2/28/2021.

Federal Association of Statutory Health Insurance Physicians (2021). Online Version of the EBM 2021. Berlin. Available online at https://www.kbv.de/html/online-ebm.php, checked on 1/28/2021.

Federal Joint Committee (2015). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Qualitätsbeurteilungs-Richtlinien für die Kernspintomographie: Ergänzung einer Übergangsregelung. Berlin. Available online at https://www.g-ba.de/downloads/40-268-3365/2015-09-17_Qb-RL-Kernspintomographie_%C3%84nderung_TrG.pdf, checked on 2/28/2021.

Federal Joint Committee (2020a). Richtlinie des Gemeinsamen Bundesausschusses zu Untersuchungs- und Behandlungsmethoden der vertragsärztlichen Versorgung. Berlin. Available online at https://www.g-ba.de/downloads/62-492-2211/MVV-RL_2020-06-18_iK-2020-07-22.pdf, checked on 9/24/2020.

Federal Joint Committee (2020b). Richtlinie des Gemeinsamen Bundesausschusses zur Zusammenführung der Anforderungen an strukturierte Behandlungsprogramme nach § 137f Absatz 2 SGB V. Available online at https://www.g-ba.de/downloads/62-492-2350/DMP-A-RL_2020-12-17_iK-2021-01-01.pdf, checked on 2/28/2021.

Feinstein, Aaron J.; Soulos, Pamela R.; Long, Jessica B.; Herrin, Jeph; Roberts, Kenneth B.; Yu, James B.; Gross, Cary P. (2013). Variation in receipt of radiation therapy after breast-conserving surgery: assessing the impact of physicians and geographic regions. Medical Care 51 (4), pp. 330–338. DOI: 10.1097/MLR.0b013e31827631b0.

Fortune-Greeley, Alice K.; Wheeler, Stephanie B.; Meyer, Anne-Marie; Reeder-Hayes, Katherine E.; Biddle, Andrea K.; Muss, Hyman B.; Carpenter, William R. (2014). Preoperative breast MRI and surgical outcomes in elderly women with invasive ductal

and lobular carcinoma: a population-based study. Breast Cancer Research and Treatment 143 (1), pp. 203–212. DOI: 10.1007/s10549-013-2787-4.

German Federal Statistical Office (2019). Sterbetafeln 2016/2018. Ergebnisse aus der laufenden Berechnung von Periodensterbetafeln für Deutschland und die Bundesländer. Available online at https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/Publikationen/Downloads-Sterbefaelle/periodensterbetafeln-bundeslaender-5126204187004.pdf;jsessionid=030D733B2493EFF9D82A55274F6A0D85.internet8742?__blob=publicationFile, checked on 10/2/2020.

German Institute of Medical Documentation and Information (2019a). ICD-10-GM 2019. Cologne.

German Institute of Medical Documentation and Information (2019b). OPS Version 2019. Cologne.

Gosden, T.; Forland, F.; Kristiansen, I. S.; Sutton, M.; Leese, B.; Giuffrida, A. et al. (2001). Impact of payment method on behaviour of primary care physicians: a systematic review. Journal of Health Services Research & Policy 6 (1), pp. 44–55. DOI: 10.1258/1355819011927198.

Guideline Programs Oncology (2017). Interdisziplinäre S3-Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms. Langversion 4.0. Berlin. Available online at https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Mammakarzinom_4_0/Version_4.0/LL_Mammakarzinom_Langversion_4.0_01.pdf, checked on 2/28/2021.

Haas, Laura; Stargardt, Tom; Schreyoegg, Jonas (2012). Cost-effectiveness of open versus laparoscopic appendectomy: a multilevel approach with propensity score matching. The European Journal of Health Economics 13 (5), pp. 549–560. DOI: 10.1007/s10198-011-0355-6.

Hainmueller, Jens (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Polit. anal. 20 (1), pp. 25–46. DOI: 10.1093/pan/mpr025.

Hayes, Lisa M.; Frebault, Julia S.; Landercasper, Jeffrey; Borgert, Andrew J.; Vang, Choua A.; Johnson, Jeanne M.; Linebarger, Jared H. (2016). Extramammary findings in diagnostic breast magnetic resonance imaging among patients with known breast cancer:

incidence and cost analysis. American Journal of Surgery 212 (6), pp. 1194–1200. DOI: 10.1016/j.amjsurg.2016.09.020.

Helsedirektoratet (2018). Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft (11). Available online at https://nbcgblog.files.wordpress.com/2018/08/is-2736_brystkrefthandlingsprogram-11-utgave.pdf, checked on 2/28/2021.

Hong, Chi-Chen; Ambrosone, Christine B.; Goodwin, Pamela J. (2015). Comorbidities and Their Management: Potential Impact on Breast Cancer Outcomes. Advances in Experimental Medicine and Biology 862, pp. 155–175. DOI: 10.1007/978-3-319-16366-6_11.

Houssami, Nehmat; Turner, Robin M.; Morrow, Monica (2017). Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer. Breast Cancer Research and Treatment 165 (2), pp. 273–283. DOI: 10.1007/s10549-017-4324-3.

Kuhl, Christiane (2007). The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. Radiology 244 (2), pp. 356–378. DOI: 10.1148/radiol.2442051620.

Kuhl, Christiane; Weigel, Stefanie; Schrading, Simone; Arand, Birke; Bieling, Heribert; König, Roy et al. (2010). Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. Journal of Clinical Oncology 28 (9), pp. 1450–1457. DOI: 10.1200/JCO.2009.23.0839.

Kuo, Raymond N.; Dong, Yaa-Hui; Liu, Jen-Pei; Chang, Chia-Hsuin; Shau, Wen-Yi; Lai, Mei-Shu (2011). Predicting healthcare utilization using a pharmacy-based metric with the WHO's Anatomic Therapeutic Chemical algorithm. Medical Care 49 (11), pp. 1031–1039. DOI: 10.1097/MLR.0b013e31822ebe11.

Lee, Janie M.; Houssami, Nehmat (2016). Imaging Surveillance of Women With a Personal History of Breast Cancer. In : Breast Cancer Screening: An Examination of Scientific Evidence: Elsevier, pp. 299–322. DOI: 10.1016/B978-0-12-802209-2.00012-7.

Lehman, Constance D.; DeMartini, Wendy; Anderson, Benjamin O.; Edge, Stephen B. (2009). Indications for breast MRI in the patient with newly diagnosed breast cancer. Journal of the National Comprehensive Cancer Network 7 (2), pp. 193–201. DOI: 10.6004/jnccn.2009.0013.

Little, Paul; Dorward, Martina; Warner, Greg; Stephens, Katharine; Senior, Jane; Moore, Michael (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. BMJ (Clinical research ed.) 328 (7437), p. 444. DOI: 10.1136/bmj.38013.644086.7C.

Mausbach, Brent T.; Irwin, Scott A. (2017). Depression and healthcare service utilization in patients with cancer. Psycho-Oncology 26 (8), pp. 1133–1139. DOI: 10.1002/pon.4133.

Medical Service of German Statutory Health Insurance providers (2018). IGeL-Report 2018. Available online at https://www.igel-monitor.de/fileadmin/Downloads/Presse/2018_05_03_PK_IGEL-MONITOR/18_05_03_IGeL-Report_2018_Ausfuehrlicher_Bericht.pdf, checked on 2/28/2021.

NICE (2018). Early and locally advanced breast cancer: diagnosis and management. NICE guideline. Available online at https://www.nice.org.uk/guidance/ng101/resources/early-and-locally-advanced-breast-cancer-diagnosis-and-management-pdf-66141532913605, checked on 2/28/2021.

Onega, Tracy L.; Zhu, W.; Weiss, J. E.; Goodrich, M.; Tosteson, A. N. A.; DeMartini, W. et al. (2018). Preoperative breast MRI and mortality in older women with breast cancer. Breast Cancer Research and Treatment 170 (1), pp. 149–157. DOI: 10.1007/s10549-018-4732-z.

Padia, Shilpa A.; Freyvogel, Mary; Dietz, Jill; Valente, Stephanie; O'Rourke, Colin; Grobmyer, Stephen R. (2016). False-positive Extra-Mammary Findings in Breast MRI: Another Cause for Concern. The Breast Journal 22 (1), pp. 90–95. DOI: 10.1111/tbj.12524.

Parmar, Abhishek D.; Sheffield, Kristin M.; Vargas, Gabriela M.; Han, Yimei; Chao, Celia; Riall, Taylor S. (2013). Quality of post-treatment surveillance of early stage breast cancer in Texas. Surgery 154 (2), pp. 214–225. DOI: 10.1016/j.surg.2013.04.004.

Parsyan, Armen; Alqahtani, Awadh; Mesurolle, Benoit; Meterissian, Sarkis (2013). Impact of preoperative breast MRI on surgical decision making and clinical outcomes: a systematic review. World Journal of Surgery 37 (9), pp. 2134–2139. DOI: 10.1007/s00268-013-2077-7.

Quan, Hude; Sundararajan, Vijaya; Halfon, Patricia; Fong, Andrew; Burnand, Bernard; Luthi, Jean-Christophe et al. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical Care 43 (11), pp. 1130–1139. DOI: 10.1097/01.mlr.0000182534.19832.83.

Riley, Gerald F.; Lubitz, James D. (2010). Long-term trends in Medicare payments in the last year of life. Health Services Research 45 (2), pp. 565–576. DOI: 10.1111/j.1475-6773.2010.01082.x.

Rosenbaum, Paul R.; Rubin, Donald B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70 (1), p. 41. DOI: 10.2307/2335942.

Ryan, Andrew M.; Burgess, James F.; Dimick, Justin B. (2015). Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. Health Services Research 50 (4), pp. 1211–1235. DOI: 10.1111/1475-6773.12270.

Spiegel, Tirtza N.; Esplen, Mary Jane; Hill, Kimberley A.; Wong, Jiahui; Causer, Petrina A.; Warner, Ellen (2011). Psychological impact of recall on women with BRCA mutations undergoing MRI surveillance. The Breast 20 (5), pp. 424–430. DOI: 10.1016/j.breast.2011.04.004.

Stroka, Magdalena A. (2016). Drug overprescription in nursing homes: an empirical evaluation of administrative data. The European Journal of Health Economics 17 (3), pp. 257–267. DOI: 10.1007/s10198-015-0676-y.

Stuart, Elizabeth A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science 25 (1), pp. 1–21. DOI: 10.1214/09-STS313.

Techniker Krankenkasse (2021). Die TK in Zahlen. Eine Übersicht über das Jahr 2020. Hamburg. Available online at https://www.tk.de/resource/blob/2053672/a0e7a33de4a2b49b3b1f3a87d1734846/tk-in-zahlen-pdf-data.pdf, checked on 2/25/2021.

The American Society of Breast Surgeons (2018). Consensus Guideline on Diagnostic and Screening Magnetic Resonance Imaging of the Breast. Available online at https://www.breastsurgeons.org/docs/statements/Consensus-Guideline-on-Diagnostic-and-Screening-Magnetic-Resonance-Imaging-of-the-Breast.pdf, checked on 2/28/2021.

Waks, Adrienne G.; Winer, Eric P. (2019). Breast Cancer Treatment: A Review. JAMA 321 (3), pp. 288–300. DOI: 10.1001/jama.2018.19323.

Wang, Shi-Yi; Long, Jessica B.; Killelea, Brigid K.; Evans, Suzanne B.; Roberts, Kenneth B.; Silber, Andrea; Gross, Cary P. (2016). Preoperative Breast Magnetic Resonance Imaging and Contralateral Breast Cancer Occurrence Among Older Women With Breast Cancer. JCO 34 (4), pp. 321–328. DOI: 10.1200/JCO.2015.62.9741.

Wang, Shi-Yi; Long, Jessica B.; Killelea, Brigid K.; Evans, Suzanne B.; Roberts, Kenneth B.; Silber, Andrea L. et al. (2018). Associations of preoperative breast magnetic resonance imaging with subsequent mastectomy and breast cancer mortality. Breast Cancer Research and Treatment 172 (2), pp. 453–461. DOI: 10.1007/s10549-018-4919-3.

Xu, Xiao; Herrin, Jeph; Soulos, Pamela R.; Saraf, Avantika; Roberts, Kenneth B.; Killelea, Brigid K. et al. (2016). The Role of Patient Factors, Cancer Characteristics, and Treatment Patterns in the Cost of Care for Medicare Beneficiaries with Breast Cancer. Health Services Research 51 (1), pp. 167–186. DOI: 10.1111/1475-6773.12328.

Appendix

Table 4-7: Overview of the Conditioning Variables

Category	Variable(s)	Definition
Socio-demographic Indicators at the time of the index date	Age	In years
	Age squared	In years
	Person group	6 categories (employee, self-employed, student, unemployed, welfare recipient, pensionary)
	Insurance status	1 = covered by family member, 0 otherwise
	Insurance category	1 = voluntarily insured, 0 otherwise
	High income	Interaction between employee and insurance category
Comorbidities/ Well-being	DMP other than breast cancer	1 = participation at time of index date, 0 otherwise
	Elixhauser comorbidity groups [a]	30 categories
	Pharmacy-based classes	32 categories
	Sickness absence 0-3 months pre	1= at least 1 day of sickness absence prior to index date, 0 otherwise
	Sickness absence 4-12 months pre	1= at least 1 day of sickness absence prior to index date, 0 otherwise
	Sickness absence 13-24 months pre	1= at least 1 day of sickness absence prior to index date, 0 otherwise
	Depressive episode 0-3 months pre	1= at least 1 depressive episode prior to index date, 0 otherwise
	Depressive episode 4-12 months pre	1= at least 1 depressive episode prior to index date, 0 otherwise
	Depressive episode 13-24 months pre	1= at least 1 depressive episode prior to index date, 0 otherwise
Direct medical cost of care	Total cost 0-3 months pre	In 2018 Euros, prior to index date
	Total cost 4-12 months pre	In 2018 Euros, prior to index date
	Total cost 13-24 months pre	In 2018 Euros, prior to index date
	Inpatient cost 0-3 months pre	In 2018 Euros, prior to index date
	Inpatient cost 4-12 months pre	In 2018 Euros, prior to index date
	Inpatient cost 13-24 months pre	In 2018 Euros, prior to index date
	Outpatient cost 0-3 months pre	In 2018 Euros, prior to index date
	Outpatient cost 0-3 months pre squared	In 2018 Euros, prior to index date
	Outpatient cost 4-12 months pre	In 2018 Euros, prior to index date
	Outpatient cost 13-24 months pre	In 2018 Euros, prior to index date
	Pharmaceutical cost 0-3 months pre	In 2018 Euros, prior to index date
	Pharmaceutical cost 4-12 months pre	In 2018 Euros, prior to index date
	Pharmaceutical cost 13-24 months pre	In 2018 Euros, prior to index date
Breast cancer specific indicators	DMP breast cancer present	1 = participation at index date, 0 otherwise
	DMP breast cancer past	1 = participation prior to index date, 0 otherwise
	ICD C50 0-12 months pre	1 = diagnosed prior to index date, 0 otherwise
	ICD C50 13-24 months pre	1 = diagnosed prior to index date, 0 otherwise
	ICD C50.8 0-12 months pre	1 = diagnosed prior to index date, 0 otherwise
	ICD C50.8 13-24 months pre	1 = diagnosed prior to index date, 0 otherwise
	Timing of 1st breast cancer diagnosis	6 categories (within 1, 2, 3, 4, 5 or >5 year(s) prior to index date)
	Lymph node metastases 0-12 m. pre	1 = diagnosed prior to index date, 0 otherwise
	Lymph node metastases 13-24 m. pre	1 = diagnosed prior to index date, 0 otherwise
	Mammography	1 = received prior to index date, 0 otherwise
	Ultrasound	1 = received prior to index date, 0 otherwise
	Mammography*Ultrasound	Interaction between Mammography & Ultrasound
	Inpatient MRI 13-24 months pre	1 = received prior to index date, 0 otherwise
	Outpatient MRI 13-24 months pre	1 = received prior to index dare, 0 otherwise
	Biopsy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Biopsy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Biopsy 13-24 months pre	1 = received prior to index date, 0 otherwise
	Conserving surgery 0-3 months pre	1 = received prior to index date, 0 otherwise
	Conserving surgery 4-12 months pre	1 = received prior to index date, 0 otherwise
	Conserving surgery 13-24 months pre	1 = received prior to index date, 0 otherwise
	Mastectomy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Mastectomy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Mastectomy 13-24 months pre	1 = received prior to index date, 0 otherwise
	Anti HER2 therapy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Anti HER2 therapy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Anti HER2 therapy 13-24 months pre	1 = received prior to index date, 0 otherwise
	Chemotherapy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Chemotherapy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Chemotherapy 13-24 months pre	1 = received prior to index date, 0 otherwise
	Hormone therapy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Hormone therapy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Hormone therapy 13-24 months pre	1 = received prior to index date, 0 otherwise
	Radiation therapy 0-3 months pre	1 = received prior to index date, 0 otherwise
	Radiation therapy 4-12 months pre	1 = received prior to index date, 0 otherwise
	Radiation therapy 13-24 months pre	1 = received prior to index date, 0 otherwise

Table 4-7 continued

Category	Variable(s)	Definition
Breast cancer specific indicators	Waiting period surgery	1 = surgery < 6 months prior to index date, 0 otherwise
	Waiting period radiation therapy	1 = radiation therapy < 12 months prior to index date, 0 otherwise
	Hospitalization 0-3 months pre	1 = at least 1 hospital day prior to index date, 0 otherwise
	Hospitalization 4-12 months pre	1 = at least 1 hospital day prior to index date, 0 otherwise
	Hospitalization 13-24 months pre	1 = at least 1 hospital day prior to index date, 0 otherwise
	Hospitalization BC 0-3 months pre	1 = at least 1 hospital day due to BC prior to index date, 0 otherwise
	Hospitalization BC 4-12 months pre	1 = at least 1 hospital day due to BC prior to index date, 0 otherwise
	Hospitalization BC 13-24 months pre	1 = at least 1 hospital day due to BC prior to index date, 0 otherwise
	Hospital days 0-3 months pre	Number of hospital days prior to index date
	Hospital days 4-12 months pre	Number of hospital days prior to index date
	Hospital days 13-24 months pre	Number of hospital days prior to index date
	Hospital days BC 0-3 months pre	Number of hospital days due to BC prior to index date
	Hospital days BC 4-12 months pre	Number of hospital days due to BC prior to index date
	Hospital days BC 13-24 months pre	Number of hospital days due to BC prior to index date
	Outpat. consultation 0-3 months pre	1 = at least 1 outpatient consultation prior to index date, 0 otherwise
	Outpat. consultation 4-12 months pre	1 = at least 1 outpatient consultation prior to index date, 0 otherwise
	Outpat. consultation 13-24 months pre	1 = at least 1 outpatient consultation prior to index date, 0 otherwise
	Outpat. consultation BC 0-3 m. pre	1 = at least 1 outpat. consult. due to BC prior to index date, 0 otherwise
	Outpat. consultation BC 4-12 m. pre	1 = at least 1 outpat. consult. due to BC prior to index date, 0 otherwise
	Outpat. consultation BC 13-24 m. pre	1 = at least 1 outpat. consult. due to BC prior to index date, 0 otherwise
	Outpat. consultations 0-3 months pre	Number of outpatient consultations prior to index date
	Outpat. consultations 4-12 months pre	Number of outpatient consultations prior to index date
	Outpat. consultations 13-24 months pre	Number of outpatient consultations prior to index date
	Outpat. consultations BC 0-3 m. pre	Number of outpatient consultations due to BC prior to index date
	Outpat. consultations BC 4-12 m. pre	Number of outpatient consultations due to BC prior to index date
	Outpat. consultations BC 13-24 m. pre	Number of outpatient consultations due to BC prior to index date
	Outpat. gyn*onco consultation 3 m. pre	1 = at least 1 gynecologist & at least 1 oncologist outpatient consultation prior to index date, 0 otherwise
	Prescription 0-3 months pre	1 = at least 1 pharmaceut. prescription prior to index date, 0 otherwise
	Prescription 4-12 months pre	1 = at least 1 pharmaceut. prescription prior to index date, 0 otherwise
	Prescription 13-24 months pre	1 = at least 1 pharmaceut. prescription prior to index date, 0 otherwise
	Prescription BC 0-3 months pre	1 = at least 1 pharmaceut. pr. due to BC prior to index date, 0 otherwise
	Prescription BC 4-12 months pre	1 = at least 1 pharmaceut. pr. due to BC prior to index date, 0 otherwise
	Prescription BC 13-24 months pre	1 = at least 1 pharmaceut. pr. due to BC prior to index date, 0 otherwise
	Prescriptions 0-3 months pre	Number of pharmaceutical prescriptions prior to index date
	Prescriptions 4-12 months pre	Number of pharmaceutical prescriptions prior to index date
	Prescriptions 13-24 months pre	Number of pharmaceutical prescriptions prior to index date
	Prescriptions BC 0-3 months pre	Number of pharmaceutical prescriptions due to BC prior to index date
	Prescriptions BC 4-12 months pre	Number of pharmaceutical prescriptions due to BC prior to index date
	Prescriptions BC 13-24 months pre	Number of pharmaceutical prescriptions due to BC prior to index date
County-level information in 2016	Average income	Average household income in 2016 Euros
	Academic degree	Share of inhabitants with academic degree
	Hospital beds	Number of hospital beds/ 1,000 inhabitants
	Gynecologists/100K inhabitants	Number of full-time equivalent gynecologists/ 100,000 inhabitants
	Average age of gynecologists	In years
	Salaried gynecologists	Share of salaried gynecologists
	Self-employed gynecologists	Share of self-employed gynecologists
	Settlement structure	4 categories (rural sparsely populated, rural moderately populated, urban, large city)
Region-level information in 2016	Radiologists/100K inhabitants	Number of full-time equivalent radiologists/ 100,000 inhabitants
	Average age of radiologists	In years
	Salaried radiologists	Share of salaried radiologists
	Self-employed radiologists	Share of self-employed radiologists

Note: This table displays the variables that are entered as covariates in the logistic regression model to estimate the propensity scores. [a] Excluding Elixhauser group "Depression" and excluding breast cancer diagnoses (ICD C50) in the Elixhauser group "Solid tumor without metastases". Data: Techniker Krankenkasse.

Table 4-8: Descriptive Statistics of the Conditioning Variables – Means and D-Statistics

Variable	Intervention		Controls		D-statistic	
	Raw	PSM	Raw	PSM	Raw	PSM
Socio-demographic indicators						
Age	56.55	57.77	63.77	57.78	-0.6278	-0.0003
Age squared	3,326	3,455	4,202	3,452	-0.6373	0.0025
Employee†	0.55	0.53	0.35	0.51	0.4079	0.0431
Self-employed†	0.05	0.04	0.02	0.05	0.1145	-0.0244
Student†	0.00	0.00	0.00	0.00	0.0112	-0.0513
Unemployed†	0.03	0.03	0.02	0.02	0.0223	0.0394
Welfare recipient†	0.00	0.00	0.00	0.00	-0.0553	0.0309
Pensionary†	0.33	0.36	0.57	0.35	-0.5155	0.0078
Covered by family member†	0.10	0.11	0.10	0.11	-0.0032	-0.0050
Voluntarily insured†	0.20	0.19	0.13	0.21	0.1755	-0.0646
High-income†	0.10	0.10	0.06	0.10	0.1446	-0.0083
Comorbidities/Well-being						
DMP other than breast cancer†	0.06	0.06	0.11	0.06	-0.1964	-0.0246
Sickness absence 0-3 months pre†	0.25	0.22	0.14	0.21	0.2760	0.0285
Sickness absence 4-12 months pre†	0.39	0.35	0.23	0.33	0.3510	0.0362
Sickness absence 13-24 months pre†	0.41	0.35	0.26	0.37	0.3104	-0.0241
Depressive episode 0-3 months pre†	0.28	0.26	0.25	0.26	0.0625	0.0017
Depressive episode 4-12 months pre†	0.33	0.31	0.30	0.30	0.0680	0.0194
Depressive episode 13-24 months pre†	0.30	0.28	0.28	0.30	0.0339	-0.0432
Elixhauser comorbidity groups†						
Raw	12 out of 30 groups are significantly different $(d > 0.1)$					
PSM	0 out of 30 groups are significantly different $(d > 0.1)$					
Pharmacy-based classes†						
Raw	9 out of 32 classes are significantly different $(d > 0.1)$					
PSM	0 out of 32 classes are significantly different $(d > 0.1)$					
Direct medical cost of care						
Total cost 0-3 months pre	€2,043	€1,725	€2,113	€1,698	-0.0152	0.0059
Total cost 4-12 months pre	€8,384	€6,795	€6,297	€6,423	0.1577	0.0281
Total cost 13-24 months pre	€8,932	€7,117	€7,620	€7,070	0.0907	0.0033
Inpatient cost 0-3 months pre	€374	€281	€482	€316	-0.0535	-0.0174
Inpatient cost 4-12 months pre	€1,533	€1,227	€1,455	€1,045	0.0173	0.0403
Inpatient cost 13-24 months pre	€2,162	€1,682	€1,751	€1,718	0.0792	-0.0070
Outpatient cost 0-3 months pre	€641	€540	€513	€557	0.1221	-0.0163
Outpatient cost 0-3 months pre squared	€1,174,246	€711,316	€1,708,350	€884,100	-0.0440	-0.0142
Outpatient cost 4-12 months pre	€2,690	€2,243	€1,604	€2,304	0.3041	-0.0170
Outpatient cost 13-24 months pre	€2,127	€1,722	€2,016	€1,731	0.0397	-0.0030
Pharmaceutical cost 0-3 months pre	€533	€480	€591	€369	-0.0232	0.0452
Pharmaceutical cost 4-12 months pre	€2,167	€1,670	€1,673	€1,501	0.0666	0.0228
Pharmaceutical cost 13-24 months pre	€2,022	€1,615	€1,895	€1,324	0.0168	0.0385
Breast cancer specific indicators						
DMP breast cancer present†	0.11	0.04	0.10	0.04	0.0456	0.0000
DMP breast cancer past†	0.22	0.18	0.20	0.18	0.0486	0.0000
ICD C50 0-12 months pre †	0.97	0.96	0.94	0.96	0.1219	0.0036
ICD C50 13-24 months pre †	0.85	0.84	0.88	0.84	-0.1083	0.0022
ICD C50.8 0-12months pre †	0.11	0.08	0.07	0.10	0.1376	-0.0679
ICD C50.8 13-24 months pre †	0.11	0.07	0.07	0.09	0.1436	-0.0588
1st BC diagnosis 1 year pre†	0.15	0.15	0.11	0.15	0.1132	0.0000
1st BC diagnosis 2 years pre†	0.60	0.66	0.64	0.66	-0.0818	0.0000
1st BC diagnosis 3 years pre†	0.02	0.00	0.02	0.00	-0.0210	0.0000
1st BC diagnosis 4 years pre†	0.03	0.01	0.02	0.01	0.0192	0.0000
1st BC diagnosis 5 years pre†	0.03	0.01	0.02	0.01	0.0370	0.0000
1st BC diagnosis >5 years pre†	0.18	0.17	0.18	0.17	-0.0058	0.0000
Lymph node metastases 0-12 m. pre†	0.06	0.04	0.05	0.05	0.0575	-0.0360
Lymph node metastases 13-24 m. pre†	0.06	0.04	0.04	0.04	0.0747	-0.0403
Mammography pre†	0.17	0.16	0.62	0.16	-1.0320	0.0000
Ultrasound pre†	0.70	0.69	0.65	0.69	0.1089	0.0000
Mammography*Ultrasound pre†	0.10	0.08	0.53	0.08	-1.0643	0.0000
Inpatient MRI 13-24 months pre†	0.04	0.03	0.01	0.02	0.1631	0.0398
Outpatient MRI 13-24 months pre†	0.45	0.38	0.03	0.35	1.1421	0.0730
Biopsy 0-3 months pre †	0.03	0.03	0.01	0.03	0.1420	-0.0276
Biopsy 4-12 months pre †	0.04	0.03	0.03	0.03	0.0285	0.0168
Biopsy 13-24 months pre †	0.07	0.05	0.04	0.07	0.1304	-0.0610
Breast conserving surgery 0-3 months pre †	0.03	0.02	0.02	0.02	0.0973	0.0000
Breast conserving surgery 4-12 months pre †	0.13	0.08	0.06	0.08	0.2685	0.0000
Breast conserving surgery 13-24 months pre †	0.15	0.11	0.08	0.11	0.2360	0.0000
Mastectomy 0-3 months pre †	0.01	0.00	0.00	0.01	0.0165	-0.0558
Mastectomy 4-12 months pre †	0.02	0.01	0.01	0.02	0.0583	-0.0400
Mastectomy 13-24 months pre †	0.04	0.02	0.02	0.02	0.1086	0.0000
Anti HER2 therapy 0-3 months pre †	0.00	0.00	0.00	0.01	0.0395	-0.0594
Anti HER2 therapy 4-12 months pre †	0.00	0.00	0.00	0.00	0.0388	-0.0634
Anti HER2 therapy 13-24 months pre †	0.00	0.00	0.00	0.00	0.0429	0.0000
Chemotherapy 0-3 months pre †	0.03	0.02	0.03	0.03	-0.0367	-0.0487

Table 4-8 continued

Variable	Intervention		Controls		D-statistic	
	Raw	PSM	Raw	PSM	Raw	PSM
Chemotherapy 4-12 months pre †	0.07	0.04	0.05	0.06	0.0893	-0.0766
Chemotherapy 13-24 months pre †	0.08	0.05	0.05	0.06	0.1094	-0.0579
Hormone therapy 0-3 months pre †	0.13	0.12	0.00	0.12	0.5430	0.0185
Hormone therapy 4-12 months pre †	0.17	0.15	0.00	0.16	0.6154	-0.0169
Hormone therapy 13-24 months pre †	0.14	0.15	0.00	0.13	0.5641	0.0569
Radiation therapy 0-3 months pre †	0.02	0.00	0.02	0.00	-0.0134	0.0000
Radiation therapy 4-12 months pre †	0.16	0.11	0.06	0.11	0.3113	0.0000
Radiation therapy 13-24 months pre †	0.03	0.01	0.08	0.01	-0.1801	0.0000
Waiting period surgery†	0.04	0.01	0.02	0.01	0.1201	0.0000
Waiting period radiation therapy†	0.15	0.11	0.05	0.11	0.3372	0.0000
Hospitalization 0-3 months pre†	0.10	0.08	0.11	0.09	-0.0335	-0.0146
Hospitalization 4-12 months pre†	0.29	0.25	0.25	0.25	0.0943	0.0084
Hospitalization 13-24 months pre†	0.35	0.30	0.30	0.32	0.1020	-0.0530
Hospitalization BC 0-3 months pre†	0.03	0.03	0.03	0.03	0.0156	0.0000
Hospitalization BC 4-12 months pre†	0.11	0.08	0.07	0.08	0.1687	-0.0026
Hospitalization BC 13-24 months pre†	0.17	0.12	0.08	0.13	0.2479	-0.0320
Hospital days 0-3 months pre	0.68	0.59	1.09	0.64	-0.0893	-0.0107
Hospital days 4-12 months pre	3.01	2.39	3.09	2.36	-0.0074	0.0026
Hospital days 13-24 months pre	4.25	3.24	3.71	3.47	0.0374	-0.0161
Hospital days BC 0-3 months pre	0.16	0.14	0.34	0.24	-0.0636	-0.0357
Hospital days BC 4-12 months pre	1.30	0.67	0.98	1.07	0.0429	-0.0470
Hospital days BC 13-24 months pre	1.92	0.90	1.11	1.41	0.0805	-0.0501
Outpat. consultation 0-3 months pre†	1.00	1.00	0.96	1.00	0.2576	0.0000
Outpat. consultation 4-12 months pre†	0.99	0.99	0.99	0.99	-0.0136	-0.0301
Outpat. consultation 13-24 months pre†	1.00	1.00	1.00	1.00	0.0583	0.0000
Outpat. consultation BC 0-3 m. pre†	0.53	0.45	0.37	0.44	0.3364	0.0367
Outpat. consultation BC 4-12 m. pre†	0.61	0.52	0.49	0.53	0.2311	-0.0303
Outpat. consultation BC 13-24 m. pre†	0.58	0.52	0.48	0.52	0.1921	-0.0151
Outpat. consultations 0-3 months pre	10.69	9.87	8.96	10.03	0.2425	-0.0234
Outpat. consultations 4-12 months pre	29.68	26.47	25.38	26.61	0.1978	-0.0065
Outpat. consultations 13-24 months pre	33.77	31.01	32.65	30.79	0.0497	0.0100
Outpat. consultations BC 0-3 months pre	1.67	1.36	1.13	1.35	0.2697	0.0053
Outpat. consultations BC 4-12 months pre	3.92	2.82	2.76	2.89	0.2230	-0.0117
Outpat. consultations BC 13-24 months pre	3.92	3.17	3.23	2.93	0.1253	0.0444
Outpat. gyn*onco consultation 3 m. pre†	0.05	0.04	0.03	0.04	0.1035	0.0307
Prescription 0-3 months pre†	0.79	0.77	0.83	0.76	-0.1018	0.0134
Prescription 4-12 months pre†	0.91	0.90	0.93	0.88	-0.0669	0.0776
Prescription 13-24 months pre†	0.93	0.92	0.94	0.91	-0.0451	0.0427
Prescription BC 0-3 months pre†	0.31	0.23	0.24	0.23	0.1626	0.0152
Prescription BC 4-12 months pre†	0.39	0.30	0.30	0.30	0.1887	-0.0016
Prescription BC 13-24 months pre†	0.31	0.25	0.30	0.24	0.0262	0.0310
Prescriptions 0-3 months pre	3.89	3.80	4.97	3.72	-0.2321	0.0157
Prescriptions 4-12 months pre	10.94	9.95	12.89	9.66	-0.1632	0.0240
Prescriptions 13-24 months pre	13.10	12.26	15.94	12.22	-0.1958	0.0025
Prescriptions BC 0-3 months pre	0.69	0.50	0.52	0.50	0.1616	0.0044
Prescriptions BC 4-12 months pre	1.46	1.09	1.12	1.09	0.1664	0.0008
Prescriptions BC 13-24 months pre	1.33	1.05	1.31	0.99	0.0113	0.0285
County-level information						
Average income	1,847	1,842	1,834	1,840	0.0606	0.0096
Academic degree†	0.16	0.16	0.16	0.17	-0.0338	-0.1231
Hospital beds/1,000 inhabitants	6.09	6.04	6.04	6.18	0.0181	-0.0464
Gynecologists/1,000K inhabitants	13.87	13.86	14.07	14.25	-0.0448	-0.0883
Average age of gynecologists	53.27	53.27	53.36	53.24	-0.0580	0.0210
Salaried gynecologists†	0.17	0.17	0.17	0.17	-0.0664	0.0015
Self-employed gynecologists†	0.74	0.74	0.73	0.74	0.0702	0.0065
Rural county sparsely populated†	0.10	0.10	0.09	0.09	0.0314	0.0580
Rural county moderately populated†	0.12	0.13	0.13	0.11	-0.0293	0.0478
Urban county†	0.43	0.43	0.42	0.42	0.0340	0.0137
Large city†	0.35	0.34	0.36	0.38	-0.0345	-0.0831
Regional-level information						
Radiologists/1,000K inhabitants	3.99	3.98	4.07	4.02	-0.1012	-0.0521
Average age of radiologists	52.18	52.19	52.17	52.23	0.0105	-0.0396
Salaried radiologists†	0.37	0.37	0.37	0.36	-0.0287	0.0616
Self-employed radiologists†	0.49	0.48	0.48	0.49	0.0794	-0.0877
Observations	983	667	89,962	1,089	-	-

Note: This table provides summary statistics for the intervention and control group (raw and reweighted after PSM). The first two columns present the means of the conditioning variables for the intervention group. The third and fourth column show the means for the control group. The last two columns show the D-statistic, whereby the D-statistic represents the standardized mean difference between intervention and control group. For continuous covariates x it is defined as $d_x = (\overline{x_1} - \overline{x_0})/\sqrt{0.5 * (\sigma_{x1}^2 + \sigma_{x0}^2)}$, where $\overline{x_1}$ and $\overline{x_0}$ denote the sample mean of the covariates in the intervention and control group, respectively, and σ_{s1}^2 and σ_{s0}^2 the corresponding variances. For dichotomous covariates p, it is defined as $d_p = (\widehat{p_1} - \widehat{p_0})/\sqrt{0.5 * (\widehat{p_1}(1 - \widehat{p_1}) + \widehat{p_0}(1 - \widehat{p_0}))}$, where $\widehat{p_1}$ and $\widehat{p_0}$ denote the mean in the intervention and control group, respectively. For a well-balanced sample, the absolute standardized mean difference should not be larger than 0.1 (Austin 2009). † The mean represents a share. Data: Techniker Krankenkasse.

Table 4-9: Descriptive Statistics for the Main Results

	Outcome means				(DiD) Estimator	
	Intervention		Control		Absolute	%
	Pre	Post	Pre	Post		
(i) Direct medical cost						
Total cost	€7,683	€8,149	€7,357	€6,130	€1,717	26.7 %
Inpatient cost	€1,501	€1,750	€1,486	€1,241	€743	73.8 %
Outpatient cost	€2,246	€2,257	€2,266	€1,890	€460	25.6 %
Pharmaceutical cost	€1,741	€2,251	€1,356	€1,325	€430	23.6 %
(ii) Medical resource consumption						
Hospitalization†	0.445	0.469	0.490	0.404	10.5 ppts.	28.8 %
Thereof BC specific†	0.190	0.156	0.200	0.078	8.8 ppts.	128.3 %
Hospital days	3.10	3.22	3.23	2.77	0.45	16.2 %
Thereof BC specific	0.85	1.15	1.36	0.73	0.15	15.4 %
Outpatient consultation†	1.000	1.000	1.000	0.998	0.2 ppts.	0.2 %
Thereof BC specific†	0.655	0.636	0.666	0.611	3.6 ppts.	6.0 %
Outpatient consultations	32.68	32.89	32.72	30.68	2.60	8.6 %
Thereof BC specific	3.35	3.72	3.17	2.81	0.40	12.1 %
Pharmaceutical prescription†	0.973	0.981	0.960	0.954	1.3 ppts.	1.3 %
Thereof BC specific†	0.337	0.366	0.325	0.320	3.4 ppts.	10.1 %
Pharmaceutical prescriptions	12.19	13.70	12.01	12.47	0.94	7.4 %
Thereof BC specific	1.10	1.46	1.07	1.20	0.09	6.7 %
(iii) Well-being						
Depressive episode†	0.340	0.378	0.353	0.387	0.4 ppts.	1.0 %
Sickness absence†	0.459	0.454	0.441	0.433	0.3 ppts.	0.7 %
(iv) BC specific procedures						
Biopsy†	-	0.047	-	0.002	4.5 ppts.	3000 %
Further imaging†	-	0.020	-	0.026	-0.7 ppts.	-25.6 %
Breast conserving surgery †	-	0.100	-	0.017	8.4 ppts.	508.5 %
Mastectomy†	-	0.036	-	0.005	3.1 ppts.	592.3 %
Anti HER2 therapy†	-	0.002	-	0.008	-0.6 ppts.	-80 %
Chemotherapy†	-	0.041	-	0.035	0.5 ppts.	15.1 %
Hormone therapy†	-	0.231	-	0.234	-0.3 ppts.	-1.3 %
Radiation therapy†	-	0.048	-	0.023	2.6 ppts.	113.3 %
Observations	667		1,089		-	

Note: Columns one to four display the means of the outcome variables for the intervention and control group after matching in the pre- and post-intervention period, respectively. For the BC specific outcomes, we consider a follow-up period of 90 days, while for all other outcomes the post-period lasts two years. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. For cost and countable medical resource consumption outcomes, the means represent yearly averages over the pre- and post-period, respectively. The fifths and sixths column provide the estimated changes in outcomes. For cost and countable outcomes, we estimated percentage changes. Thus, for instance, for total direct medical cost the absolute change is computed as follows: 0.267*€8,149/(1+0.267)=€1,717. For binary outcomes we estimated changes in percentage points. Hence, for instance, for the share of patients with at least one hospitalization the percentage change is calculated as follows: 0.105/(0.469-0.105)*100 = 28.8%. † The mean represents a share. Data: Techniker Krankenkasse.

Table 4-10: Robustness – Differences in Tumor Characteristics

	(1) Spread and size of the primary tumor (TNM)	(2) Regional lymph node involve-ment (TNM)	(3) Distant metastases spread (TNM)	(4) Patho-logical tumor grading	(5) Hormone receptor status	(6) Mani-festation distant metastases	(7) Mani-festation local recurrence
Pr<=P	0.667	0.287	0.007	0.602	1.000	0.438	0.504
Observations	237	237	229	236	236	379	379

Note: This table presents the results from Fisher's exact test for significant differences in average tumor characteristics between the non-conform MRI and the control group. Information on tumor characteristics is available for the subsample of breast cancer DMP participants only. When women are initially signed-up for the DMP, physicians are obligated to document certain tumor characteristics. While the tumor characteristics in Columns 1 to 5 only need to be documented for women with a primary tumor, the characteristics in Column 6 and Column 7 need to be documented for all women. DMP: Disease management program. TNM: International tumor classification systems, accounting for tumor size, node status and metastases. Data: Techniker Krankenkasse.

Table 4-11: Robustness – Results without Auditing ASHIPs

	(3) Main results		(16) Without auditing ASHIPs	
(i) Direct medical cost				
Total cost	0.267***	(0.057)	0.327***	(0.072)
Inpatient cost	0.738***	(0.217)	1.010***	(0.274)
Outpatient cost	0.256***	(0.044)	0.318***	(0.055)
Pharmaceutical cost	0.236**	(0.097)	0.328***	(0.123)
(ii) Medical resource consumption				
Hospitalization†	0.105***	(0.028)	0.187***	(0.036)
Thereof BC specific†	0.088***	(0.020)	0.158***	(0.026)
Hospital days	0.162***	(0.053)	0.258***	(0.066)
Thereof BC specific	0.154***	(0.037)	0.240***	(0.045)
Outpatient consultation†	0.002*	(0.001)	0.001	(0.001)
Thereof BC specific†	0.036*	(0.021)	0.017	(0.027)
Outpatient consultations	0.086***	(0.026)	0.132***	(0.032)
Thereof BC specific	0.121***	(0.035)	0.121***	(0.044)
Pharmaceutical prescription†	0.013	(0.011)	0.012	(0.013)
Thereof BC specific†	0.034**	(0.017)	0.066***	(0.022)
Pharmaceutical prescriptions	0.074**	(0.034)	0.123***	(0.042)
Thereof BC specific	0.067***	(0.025)	0.095***	(0.032)
(iii) Well-being				
Depressive episode†	0.004	(0.019)	0.028	(0.025)
Sickness absence†	0.003	(0.012)	-0.013	(0.025)
Observations	3,512		2,010	
(iv) BC specific procedures				
Biopsy†	0.045***	(0.008)	0.047***	(0.010)
Further imaging†	-0.009	(0.007)	-0.001	(0.008)
Breast conserving surgery †	0.087***	(0.011)	0.099**	(0.013)
Mastectomy†	0.030***	(0.007)	0.037***	(0.009)
Anti HER2 therapy†	-0.006***	(0.002)	-0.004*	(0.002)
Chemotherapy†	0.013*	(0.008)	0.010	(0.010)
Hormone therapy†	-0.009	(0.015)	0.020	(0.018)
Radiation therapy†	0.024***	(0.008)	0.027***	(0.010)
(v) Mortality				
Non-conform MRI	0.990	(0.004)	0.995	(0.004)
Control	0.977	(0.005)	0.973	(0.007)
Difference	0.013*		0.022**	
Log-Rank test	p=0.062		p=0.015	
Observations	1,756		1,005	

Note: This table shows the results for the subsample of MRI scans, which were provided in non-auditing ASHIP-regions only. Estimates in Column 3 coincide with the estimates in Column 3 of Table 4-1 and report our main results. Column 16 depicts the results when excluding MRI scans, which were provided by physicians who are a member of an auditing ASHIP (compare section 4.2). We conducted the analyses with newly matched control subjects, whereby the control group consists only of women, who live in non-auditing ASHIP-regions. We matched on the same set of covariates as in our main analyses. After matching the new sample is well-balanced over all covariates. Specifications for the estimation of outcomes in categories *(i)-(iii)* and for outcomes in category *(iv)* are based on our preferred DiD model from equation (1) and the LPM model from equation (2), respectively. For outcomes in categories *(i)-(iii)* we report standard errors clustered at the individual-level, and for outcomes in category *(iv)* we report heteroscedasticity robust standard errors. Cost and countable medical resource consumption outcomes are transformed using the natural logarithm of the outcome plus one. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. Regression models are weighted by the weights from the matching step. † The outcome represents a share. *** p<0.01, ** p<0.05, * p<0.1. Data: Techniker Krankenkasse.

Table 4-12: Descriptive Statistics by the Purpose of the MRI in the Course of BC Treatment

	MRI for primary diagnosis						MRI for surveillance					
	Outcome means				(DiD) Estimator		Outcome means				(DiD) Estimator	
	Intervention		Control		Absolute	%	Intervention		Control		Absolute	%
	Pre	Post	Pre	Post			Pre	Post	Pre	Post		
(i) Direct medical cost												
Total cost	€5,025	€15,473	€4,569	€10,317	€4,577	42.0 %	€7,961	€7,385	€7,647	€5,693	€1.482	25.1 %
Inpatient cost	€877	€3,431	€809	€2,157	€2,225	184.6 %	€1,566	€1,575	€1,557	€1,145	€605	62.3 %
Outpatient cost	€1,299	€3,676	€1,417	€2,595	€895	32.2 %	€2,344	€2,109	€2,354	€1,816	€420	24.9 %
Pharmaceutical cost	€1,555	€5,080	€889	€3,038	€2,137	72.6 %	€1,761	€1,955	€1,404	€1,146	€305	18.5 %
(ii) Medical resource consumption												
Hospitalization†	0.333	0.746	0.341	0.508	24.6 ppts.	49.2 %	0.462	0.440	0.505	0.393	9.0 ppts.	25.7 %
Thereof BC specific†	0.095	0.524	0.079	0.198	31.0 ppts.	144.6 %	0.200	0.118	0.213	0.065	6.5 ppts.	123.7 %
Hospital days	1.69	5.56	1.81	4.08	1.65	42.3 %	3.25	2.98	3.38	2.63	0.35	13.5 %
Thereof BC specific	0.30	4.12	0.27	1.23	1.48	56.0 %	0.91	0.84	1.47	0.67	0.08	11.2 %
Outpatient consultation†	1.000	1.000	1.000	0.992	0.8 ppts.	0.8 %	1.000	1.000	1.000	0.998	0.2 ppts.	0.2 %
Thereof BC specific†	0.254	0.524	0.310	0.397	18.3 ppts.	53.7 %	0.697	0.647	0.704	0.633	2.1 ppts.	3.4 %
Outpatient consultations	23.82	36.74	24.20	33.14	5.42	17.3 %	33.61	32.49	33.61	30.43	2.32	7.7 %
Thereof BC specific	0.42	3.70	0.72	2.33	1.15	45.1 %	3.66	3.72	3.42	2.86	0.29	8.6 %
Pharmaceutical prescription†	0.937	0.968	0.913	0.889	5.6 ppts.	6.1 %	0.977	0.982	0.964	0.961	0.8 ppts.	0.8 %
Thereof BC specific†	0.000	0.413	0.024	0.151	28.6 ppts.	225.2 %	0.373	0.361	0.357	0.338	0.7 ppts.	2.0 %
Pharmaceutical prescriptions	10.53	17.70	8.34	12.26	2.56	16.9 %	12.37	13.28	12.39	12.49	0.80	6.4 %
Thereof BC specific	0.00	1.66	0.05	0.58	0.51	44.4 %	1.21	1.43	1.02	1.27	0.04	2.8 %
(iii) Well-being												
Depressive episode†	0.222	0.302	0.278	0.333	2.4 ppts.	8.6 %	0.353	0.386	0.361	0.392	0.2 ppts.	0.5 %
Sickness absence†	0.556	0.603	0.373	0.429	-0.8 ppts.	-1.3 %	0.449	0.439	0.448	0.434	0.4 ppts.	0.9 %
(iv) BC specific procedures												
Biopsy†	-	0.079	-	0.008	7.1 ppts.	900 %	-	0.043	-	0.001	4.2 ppts.	5,148.8 %
Further imaging†	-	0.143	-	0.016	12.7 ppts.	793.75 %	-	0.007	-	0.027	-2.1 ppts.	-75.8 %
Breast conserving surgery †	-	0.429	-	0.071	35.7 ppts.	500.1 %	-	0.066	-	0.011	5.5 ppts.	515.4 %
Mastectomy†	-	0.079	-	0.008	7.1 ppts.	887.5 %	-	0.031	-	0.005	2.6 ppts.	533.4 %
Anti HER2 therapy†	-	0.016	-	0.008	0.8 ppts.	100 %	-	0.000	-	0.007	-0.7 ppts.	-100 %
Chemotherapy†	-	0.127	-	0.127	0.0 ppts.	0 %	-	0.031	-	0.026	0.6 ppts.	22.6 %
Hormone therapy†	-	0.175	-	0.071	10.3 ppts.	144.5 %	-	0.237	-	0.251	-1.4 ppts.	-5.6 %
Radiation therapy†	-	0.238	-	0.040	19.8 ppts.	495 %	-	0.028	-	0.021	0.7 ppts.	36 %
Observations	63		114		-	-	604		975		-	-

Note: This table shows the means of the outcome variables and the corresponding (DiD) estimators for the subgroup receiving the MRI for the purpose of primary diagnosis and for subgroup receiving the MRI for the purpose of surveillance. Samples are altered as described in Section 4.5.3. For the BC specific outcomes, we consider a follow-up period of 90 days, while for all other outcomes the post-period lasts two years. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. For cost and countable medical resource consumption outcomes, the means represent yearly averages over the pre- and post-period, respectively. † The mean represents a share. Data: Techniker Krankenkasse.

Table 4-13: Descriptive Statistics by the Reason of Non-Conformity of the MRI

	MRI premature						MRI as replacement					
	Outcome means				(DiD) Estimator		Outcome means				(DiD) Estimator	
	Intervention		Control		Absolute	%	Intervention		Control		Absolute	%
	Pre	Post	Pre	Post			Pre	Post	Pre	Post		
(i) Direct medical cost												
Total cost	€20,041	€7,414	€21,407	€4,789	€2,431	48.8 %	€5,448	€7,850	€5,320	€6,127	€1,625	26.1 %
Inpatient cost	€3,148	€2,349	€3,681	€718	€1,537	189.1 %	€1,151	€1,656	€1,115	€1,265	€643	63.5 %
Outpatient cost	€6,850	€2,102	€6,947	€1,919	€356	20.4 %	€1,562	€2,199	€1,663	€1,824	€477	27.7 %
Pharmaceutical cost	€4,319	€1,019	€4,772	€864	€349	52.1 %	€1,138	€2,118	€855	€1,332	€347	19.6 %
(ii) Medical resource consumption												
Hospitalization†	0.976	0.500	0.988	0.286	22.6 ppts.	82.5 %	0.364	0.461	0.407	0.410	9.4 ppts.	25.6 %
Thereof BC specific†	0.952	0.167	0.940	0.060	9.5 ppts.	132.6 %	0.074	0.152	0.088	0.074	9.3 ppts.	156.7 %
Hospital days	4.42	3.49	5.88	1.53	1.23	54.3 %	2.67	3.05	2.44	2.81	0.34	12.7 %
Thereof BC specific	3.29	1.51	4.86	0.73	0.36	31.6 %	0.33	1.08	0.55	0.62	0.14	15.4 %
Outpatient consultation†	1.000	1.000	1.000	1.000	0.0 ppts.	0.0 %	1.000	1.000	1.000	0.997	0.3 ppts.	0.3 %
Thereof BC specific†	0.857	0.905	0.976	0.905	11.9 ppts.	15.1 %	0.623	0.587	0.625	0.562	2.6 ppts.	4.6 %
Outpatient consultations	57.90	36.51	56.02	31.21	3.41	10.3 %	29.52	32.10	29.71	30.13	2.54	8.6 %
Thereof BC specific	7.36	7.29	7.05	5.77	1.11	17.9 %	2.79	3.09	2.64	2.36	0.30	10.6 %
Pharmaceutical prescription†	1.000	1.000	0.988	0.988	0.0 ppts.	0.0 %	0.972	0.979	0.958	0.951	1.4 ppts.	1.5 %
Thereof BC specific†	0.810	0.762	0.857	0.786	2.4 ppts.	3.3 %	0.280	0.303	0.271	0.257	3.6 ppts.	13.5 %
Pharmaceutical prescriptions	16.95	13.10	16.81	10.64	1.88	16.8 %	11.34	13.51	11.31	12.39	0.86	6.8 %
Thereof BC specific	2.06	3.15	2.35	3.21	0.20	6.8 %	1.00	1.21	0.98	0.94	0.09	7.8 %
(iii) Well-being												
Depressive episode†	0.333	0.357	0.321	0.405	-6.0 ppts.	-14.4 %	0.335	0.376	0.356	0.381	1.4 ppts.	3.9 %
Sickness absence†	0.643	0.524	0.667	0.536	1.2 ppts.	2.3 %	0.422	0.434	0.402	0.411	0.5 ppts.	1.2 %
(iv) BC specific procedures												
Biopsy†	-	0.071	-	0.000	7.1 ppts.	-	-	0.048	-	0.002	4.6 ppts.	2,281.0 %
Further imaging†	-	0.024	-	0.000	2.4 ppts.	-	-	0.019	-	0.031	-1.2 ppts.	-37.4 %
Breast conserving surgery †	-	0.119	-	0.000	11.9 ppts.	-	-	0.093	-	0.017	7.6 ppts.	453.9 %
Mastectomy†	-	0.048	-	0.000	4.8 ppts.	-	-	0.032	-	0.005	2.7 ppts.	504.8 %
Anti HER2 therapy†	-	0.000	-	0.024	-2.4 ppts.	-100 %	-	0.002	-	0.007	-0.5 ppts.	-75.1 %
Chemotherapy†	-	0.048	-	0.000	4.8 ppts.	-	-	0.035	-	0.032	0.4 ppts.	11.1 %
Hormone therapy†	-	0.642	-	0.690	-4.8 ppts.	-6.9 %	-	0.177	-	0.179	-0.2 ppts.	-1.2 %
Radiation therapy†	-	0.024	-	0.000	2.4 ppts.	-	-	0.042	-	0.013	2.9 ppts.	223.1 %
Observations	42		67		-	-	571		931		-	-

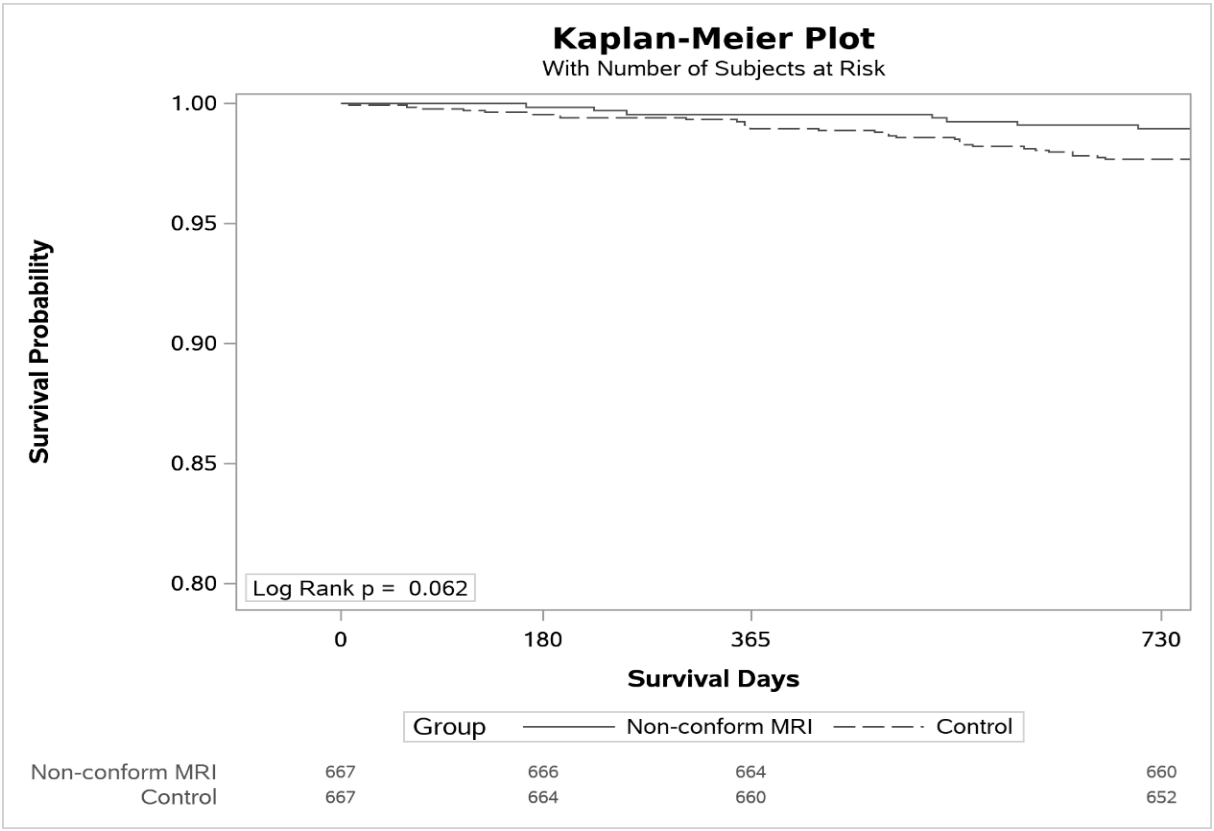
Note: This table shows the means of the outcome variables and the corresponding (DiD) estimators for the subgroup receiving a non-conform MRI that was performed prematurely and for the subgroup receiving a non-conform MRI that was performed as a replacement for conventional imaging. Samples are altered as described in Section 4.5.3. For the BC specific outcomes, we consider a follow-up period of 90 days, while for all other outcomes the post-period lasts two years. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. For cost and countable medical resource consumption outcomes, the means represent yearly averages over the pre- and post-period, respectively. † The mean represents a share. Data: Techniker Krankenkasse.

Table 4-14: Descriptive Statistics by the Settlement Structure in Women's Residential Area

	Rural area residents						Urban area residents					
	Outcome means				(DiD) Estimator		Outcome means				(DiD) Estimator	
	Intervention		Control		Absolute	%	Intervention		Control		Absolute	%
	Pre	Post	Pre	Post			Pre	Post	Pre	Post		
(i) Direct medical cost												
Total cost	€6,798	€7,930	€7,299	€5,520	€2,457	44.9 %	€7,949	€8,125	€7,371	€6,281	€1,459	21.6 %
Inpatient cost	€1,519	€1,911	€1,566	€1,123	€1,193	166.0 %	€1,496	€1,702	€1,467	€1,270	€564	49.5 %
Outpatient cost	€2,013	€2,128	€2,299	€1,830	€603	39.5 %	€2,315	€2,296	€2,258	€1,904	€406	21.5 %
Pharmaceutical cost	€1,130	€2,017	€990	€788	€535	36.1 %	€1,925	€2,321	€1,446	€1,457	€385	19.9 %
(ii) Medical resource consumption												
Hospitalization†	0.461	0.519	0.553	0.398	21.4 ppts.	70.1 %	0.446	0.454	0.474	0.406	7.6 ppts.	20.1 %
Thereof BC specific†	0.182	0.143	0.239	0.095	10.5 ppts.	277.4 %	0.193	0.160	0.191	0.074	8.4 ppts.	110.8 %
Hospital days	3.34	4.20	3.54	2.70	1.13	37.0 %	3.03	2.92	3.16	2.79	0.28	10.7 %
Thereof BC specific	0.84	1.98	1.26	0.73	0.32	19.2 %	0.86	0.90	1.38	0.72	0.11	14.4 %
Outpatient consultation†	1.000	1.000	1.000	1.000	0.0 ppts.	0.0 %	1.000	1.000	1.000	0.997	0.3 ppts.	0.3 %
Thereof BC specific†	0.695	0.623	0.667	0.636	-4.1 ppts.	-6.2 %	0.643	0.639	0.666	0.605	5.8 ppts.	10.0 %
Outpatient consultations	31.33	31.85	33.10	29.55	4.20	15.2 %	33.09	33.21	32.63	30.96	2.14	6.9 %
Thereof BC specific	3.37	3.23	3.25	3.14	-0.16	-4.7 %	3.35	3.87	3.15	2.73	0.56	16.8 %
Pharmaceutical prescription†	0.974	0.981	0.958	0.973	-0.9 ppts.	-0.9 %	0.973	0.981	0.960	0.950	1.8 ppts.	1.9 %
Thereof BC specific†	0.325	0.344	0.341	0.326	3.5 ppts.	11.3 %	0.341	0.372	0.321	0.319	3.4 ppts.	10.0 %
Pharmaceutical prescriptions	11.71	13.28	12.11	11.73	1.51	12.8 %	12.34	13.83	11.98	12.65	0.76	5.8 %
Thereof BC specific	1.06	1.41	1.04	1.16	0.08	6.2 %	1.11	1.47	1.08	1.21	0.09	6.9 %
(iii) Well-being												
Depressive episode†	0.338	0.344	0.333	0.398	-5.8 ppts.	-14.4 %	0.341	0.388	0.358	0.384	2.0 ppts.	5.4 %
Sickness absence†	0.409	0.409	0.424	0.413	1.1 ppts.	2.8 %	0.474	0.468	0.445	0.438	0.1 ppts.	0.2 %
(iv) BC specific procedures												
Biopsy†	-	0.026	-	0.000	2.6 ppts.	-	-	0.053	-	0.002	5.1 ppts.	2,715.9 %
Further imaging†	-	0.013	-	0.030	-1.7 ppts.	-57.1 %	-	0.021	-	0.025	-0.4 ppts.	-15 %
Breast conserving surgery †	-	0.065	-	0.011	5.4 ppts.	471.4 %	-	0.111	-	0.018	9.3 ppts.	525.7 %
Mastectomy†	-	0.052	-	0.004	4.8 ppts.	1,271.7 %	-	0.031	-	0.006	2.6 ppts.	456.9 %
Anti HER2 therapy†	-	0.000	-	0.004	-0.4 ppts.	-100 %	-	0.002	-	0.008	-0.6 ppts.	-76.8 %
Chemotherapy†	-	0.039	-	0.049	-1 ppts.	-20.9 %	-	0.041	-	0.032	0.9 ppts.	28.9 %
Hormone therapy†	-	0.234	-	0.216	1.8 ppts.	8.3 %	-	0.230	-	0.238	-0.8 ppts.	-3.5 %
Radiation therapy†	-	0.026	-	0.027	-0.1 ppts.	-2 %	-	0.055	-	0.021	3.3 ppts.	154 %
Observations	154		218		-	-	513		871		-	-

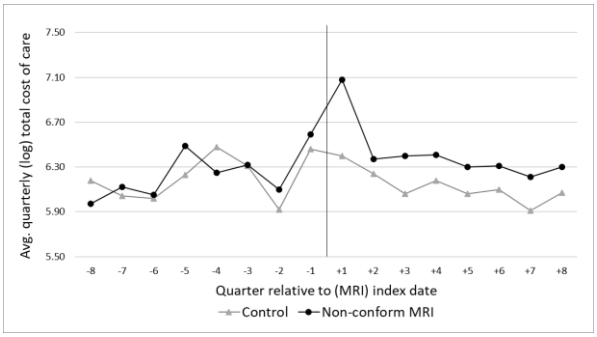
Note: This table shows the means of the outcome variables and the corresponding (DiD) estimators for the subgroup residing in rural areas and for the subgroup living in urban areas. Samples are altered as described in Section 4.5.3. For the BC specific outcomes, we consider a follow-up period of 90 days, while for all other outcomes the post-period lasts two years. Cost are corrected for inflation, expressed in 2018 Euros and winsorized at the 99% percentile. For cost and countable medical resource consumption outcomes, the means represent yearly averages over the pre- and post-period, respectively. † The mean represents a share. Data: Techniker Krankenkasse.

Figure 4-5: Kaplan-Meier Survival Curves

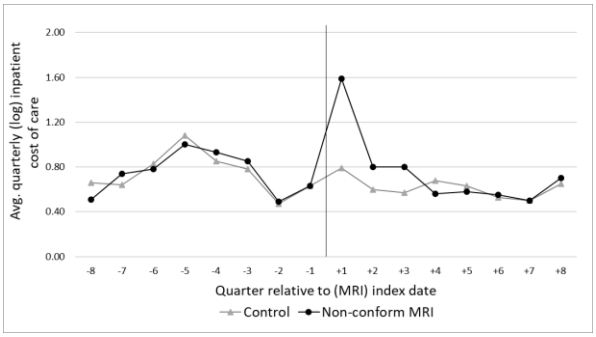


Note: This figure depicts the Kaplan-Meier survival curves for the non-conform MRI and control group. Survival functions are estimated as $S(t) = \Pr(T > t)$, where $S(t)$ describes the probability that an individual's lifetime T exceeds the threshold $t = 730$ days since index date. Data: Techniker Krankenkasse.

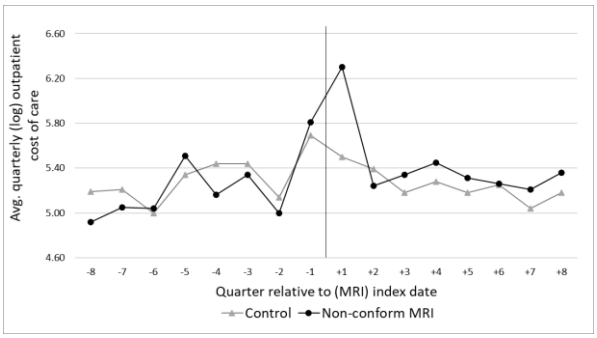
Figure 4-6: Trends



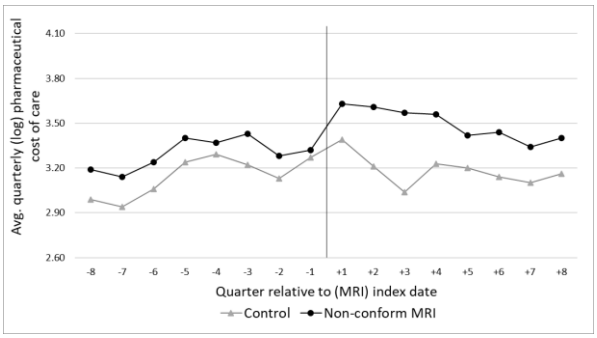
(a) (Log) Total Cost of Care ($\rho = 0.66$)



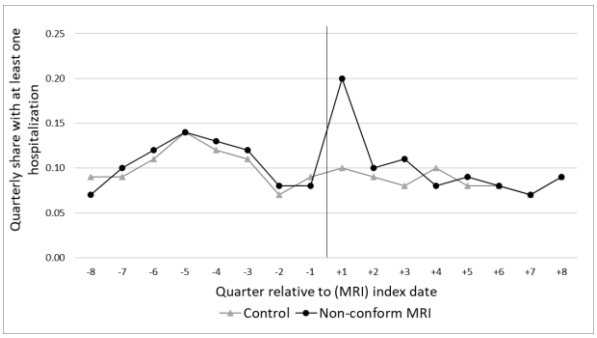
(b) (Log) Inpatient Cost of Care ($\rho = 0.89$)



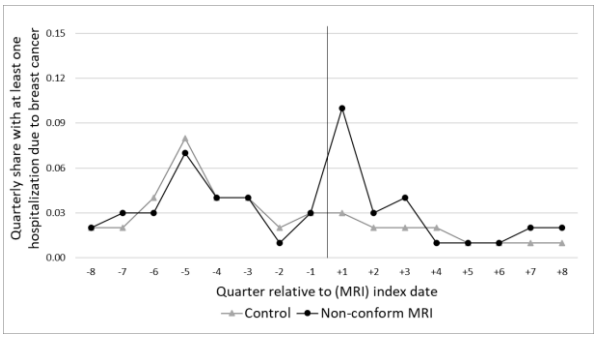
(c) (Log) Outpatient Cost of Care ($\rho = 0.83$)



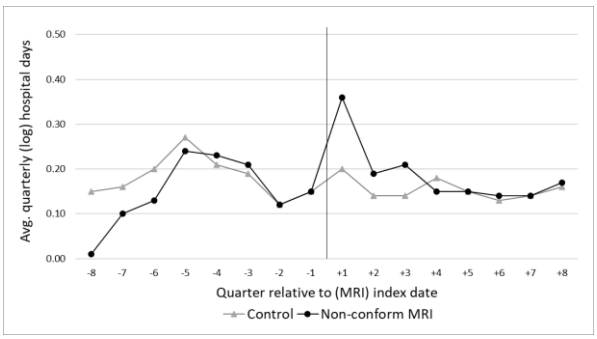
(d) (Log) Pharmaceutical Cost of Care ($\rho = 0.91$)



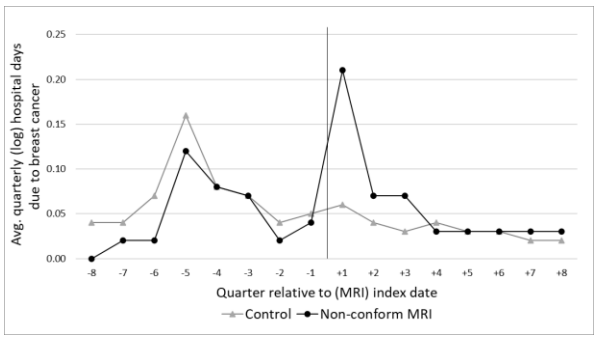
(e) Share with Hospitalization ($\rho = 0.90$)



(f) Share with Hospitalization due to BC ($\rho = 0.94$)

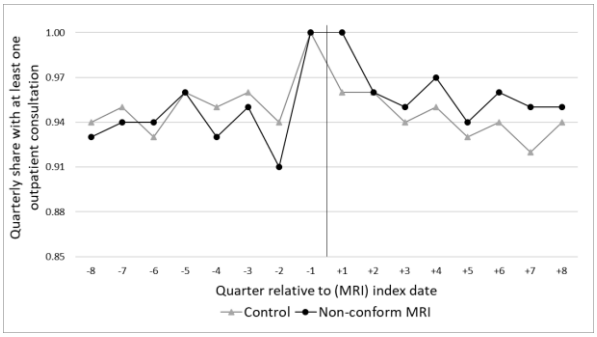


(g) (Log) Hospital Days ($\rho = 0.70$)

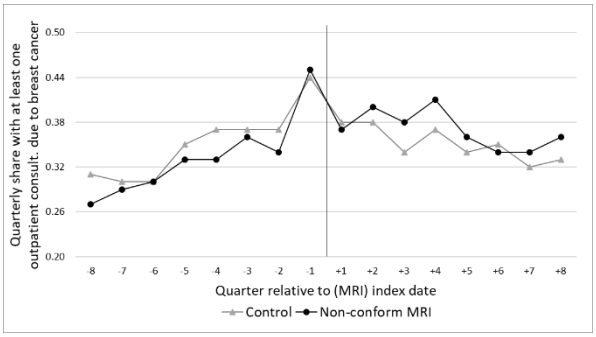


(h) (Log) Hospital Days due to BC ($\rho = 0.89$)

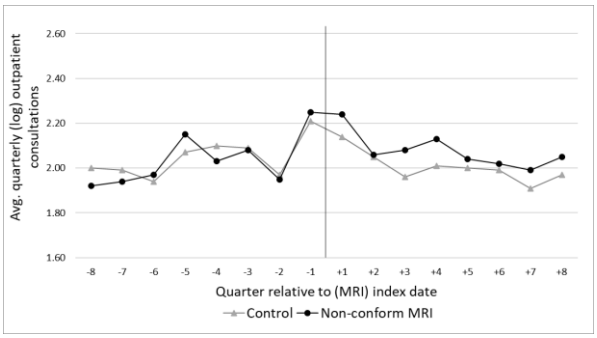
Figure 4-6 continued



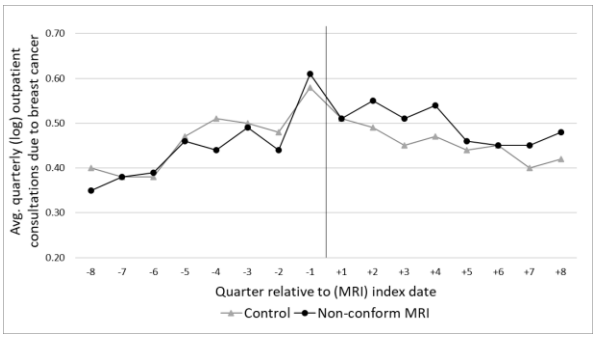
(i) Share with Outpat. Consultation ($\rho = 0.89$)



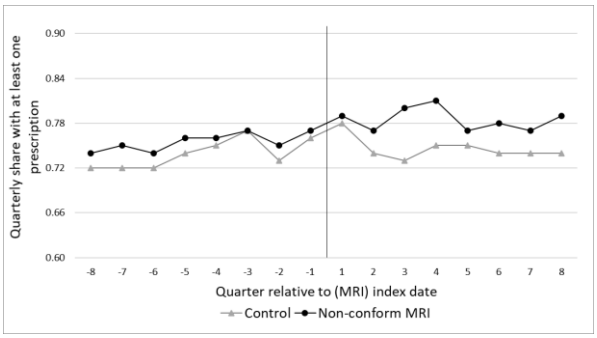
(j) Share with Outpat. Consult. due to BC ($\rho = 0.95$)



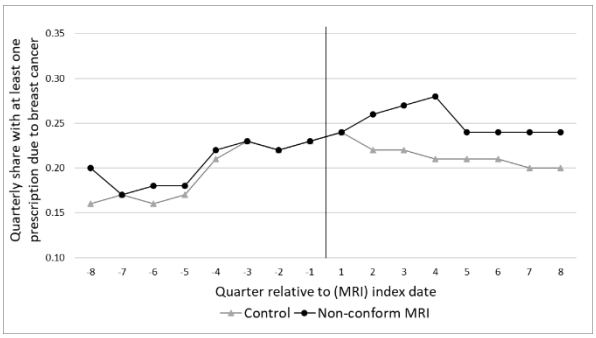
(k) (Log) Outpatient Consultations ($\rho = 0.88$)



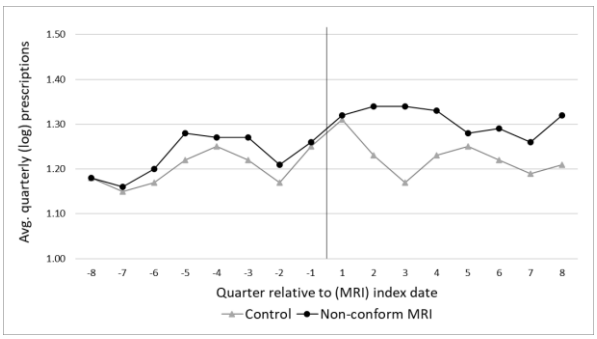
(l) (Log) Outpatient Consultations due to BC ($\rho = 0.91$)



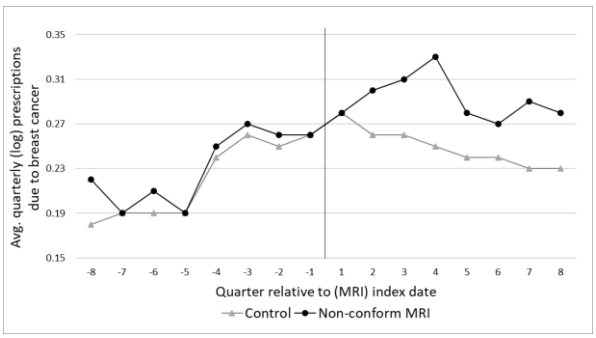
(m) Share with Prescription ($\rho = 0.95$)



(n) Share with Prescription due to BC ($\rho = 0.90$)

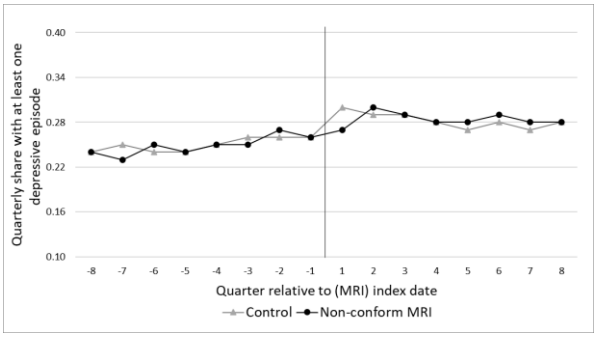


(o) (Log) Prescriptions ($\rho = 0.90$)

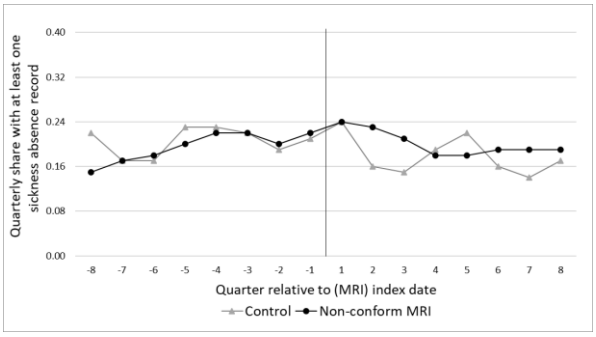


(p) (Log) Prescriptions due to BC ($\rho = 0.92$)

Figure 4-6 continued



(q) Share with a Depressive Episode ($\rho = 0.62$)



(r) Share with Sickness Absence Record ($\rho = 0.41$)

Note: This figure presents the trends in the outcomes per quarter relative to the (MRI) index date. Cost are reported in 2018 Euros and adjusted for inflation. Panel (a) shows the trends for average (log) total cost of care, Panel (b) for average (log) inpatient cost of care, Panel (c) for average (log) outpatient cost of care, Panel (d) for average (log) pharmaceutical cost of care, Panel (e) for the share of patients with at least one hospitalization, Panel (f) for the share of patients with at least one hospitalization due to BC, Panel (g) for average (log) number of hospital days, Panel (h) for average (log) hospital days due to BC, Panel (i) for the share of patients with at least one outpatient consultation, Panel (j) for the share of patients with at least one outpatient consultation due to BC, Panel (k) for the average (log) number of outpatient consultations, Panel (l) for the average (log) number of outpatient consultations due to BC, Panel (m) for the share of patients with at least one pharmaceutical prescription, Panel (n) for the share of patients with at least one pharmaceutical prescription due to BC, Panel (o) for the average (log) number of pharmaceutical prescriptions, Panel (p) for the average (log) number of pharmaceutical prescriptions due to BC, Panel (q) for the share of patients with at least one depressive episode and Panel (r) for the share patients with at least one sickness absence record. ρ is the pre-period correlation between the non-conform MRI and the control group. The vertical solid line implies the (MRI) index-date. Data: Techniker Krankenkasse.

Chapter 5

Contracts to Promote Optimal Diagnostic and Treatment Choice among Physicians with Heterogeneous Diagnostic Ability

Katharina Beenk

Abstract

Physicians' diagnostic performance is a crucial component of effective health care. However, heterogeneities in diagnostic ability and agency problems in the delivery of diagnostic and subsequent treatment services pose a challenge to regulators to provide optimal incentives. Considering a treatment choice model with endogenous diagnostic effort, I examine how a public insurer should optimally contract with providers with heterogeneous ability, whereby ability is measured by means of diagnostic accuracy or efficiency. Hence, a situation with simultaneous double moral hazard and adverse selection is assumed. A first-best analysis reveals that it may turn out suboptimal to incentivize all ability-types to exert costly diagnostic effort. While single policy mixed payment contracts, consisting of a capitation payment and conditional fee-for-service payments for each treatment option, are found to promote optimal incentives, they also entail costly information rent payments to the physicians. Therefore, in the second-best situation, particularly depending on the share of the high-ability types in the market, the parameter constellations for which all ability-types should be incentivized are further restricted. In contrast to standard principal-agent theory, a menu of policies that aims at inducing self-selection among the physicians does not improve on this outcome. Although medical expert associations frequently campaign for the special importance of a correct diagnosis to effective health care, considering the social cost of care provision may paint a more differentiated picture.

5.1 Introduction

The National Academies of Sciences, Engineering, and Medicine (2015) projected that on average 1 in 20 adults in the United States who seek outpatient care within one year experience a diagnostic error. This may not only lead to missed, delayed or unnecessary (harmful) treatment, it might also have psychological and/or financial repercussions. Against this background, physicians' performance of the correct diagnosis is seen as a crucial component of effective health care. Diagnostic errors can be defined as "the failure […] to establish an accurate and timely explanation of the patient's health problem(s)…" (National Academies of Sciences, Engineering, and Medicine 2015, p. 85) and physicians' abilities such as knowledge of basic science, disease patterns as well as experiential knowledge are identified as being particularly relevant to optimal diagnostic performance (Norman 2005). More specifically, these abilities affect the accuracy and the efficiency of a diagnosis. While accuracy stands for diagnostic precision, efficiency may relate to the cost of care caused by the amount of information or time needed to reach a certain level of diagnostic accuracy (Chatterjee et al. 2019).

As in other expert markets, diagnostic ability is heterogeneous across physicians and non-contractable, which poses a challenge to regulators to create incentives for effective care. Doyle et al. (2010) and, more recently, Chan et al. (2019), for instance, provide empirical evidence for heterogeneity in physicians' diagnostic ability. They studied physicians from various specialties in the inpatient sector and radiologists for diagnosis of pneumonia and found significant differences in ability levels measured by diagnostic efficiency and accuracy, respectively. Theoretical papers, which focus on the derivation of optimal payment contracts for physicians, address physicians' diagnostic ability (endogenous diagnostic effort) accordingly (Garcia Mariñoso and Jelovac 2003; Jelovac 2001). While in these papers diagnostic ability is assumed to be non-contractable, heterogeneities in ability levels are not accounted for so far. Still, the authors show that the provision of optimal incentives calls for sophisticated payment contracts and may entail costly rent payments to the physicians.

Moreover, physicians have further informational advantages over third parties and often take a dual role - besides performing a diagnosis, they also provide treatment. It is generally assumed that physicians need to exert costly effort to diagnose a patient. Yet, neither the patient, nor the payer can observe the physicians' effort or the result (Allard et al. 2014; Jelovac 2001). In addition, physician services represent a credence good, indicating that the appropriate treatment may not be verifiable ex-post by the patient. Theoretical and experimental evidence suggests

that these information asymmetries may result in inefficient renunciation of diagnostic effort and misreporting of the diagnostic outcome to the patient, possibly leading to over- or undertreatment (Dulleck and Kerschbamer 2006; Gottschalk et al. 2020; Liu et al. 2020). Hence, an optimal payment contract needs to account for both heterogeneity in physician ability and various agency problems.

In this paper, I will present a treatment choice model with endogenous diagnostic effort and examine how a public insurer (the payer) should optimally contract with providers with heterogeneous diagnostic ability (the physicians). The model builds on previous work by Jelovac (2001) and attributes a dual role to an outpatient physician. The physician either chooses to exert diagnostic effort and, therefore, to make an informed treatment choice or to forgo effort and to make a blind treatment choice. While exerting effort imposes a disutility on the physician, in return, a private imperfect signal on the patient's state of illness is received. The potential of the physician's effort depends on her diagnostic ability, whereby heterogeneities in ability are initially operationalized by means of heterogeneities in diagnostic accuracy. There are high and low-ability types in the market, with the type being private information to the physician. The patient is assumed to either suffer from a mild or a severe illness. A basic treatment only cures the patient in case he is mildly ill, whereas an advanced treatment always cures the patient. As the effort decision, the resulting diagnostic signal, and the ability-type are non-contractible, the model captures a situation with double moral hazard from hidden action and hidden information as well as adverse selection accruing from hidden knowledge.

Regarding the first-best situation, results show that it may turn out suboptimal to incentivize all physicians to exert costly diagnostic effort. Indeed, there are circumstances, in which it is optimal to only incentivize physicians with high diagnostic ability, or even to forgo incentivizing any physician at all. Subsequently relaxing the contractibility assumptions of the First-Best and focusing on the circumstances, in which it is optimal that at least the high-types exert effort, I find that mixed payment contracts provide optimal incentives. The contracts comprise a non-negative capitation and two strictly positive fee-for-service (FFS) payments, which are paid conditional on adequate treatment choice (compare Garcia Mariñoso and Jelovac 2003). All payment instruments depend on diagnostic ability. Given the payer offers a single policy contract, he may either offer a contract that is designed for the low-types or a contract that is designed for the high-types to all physicians. A distinguished contract offer is infeasible as the high-types would always have an incentive to mimic the low-types. The low-type contract is incentive-compatible for both types, yet, entails a strictly positive information rent accruing

from hidden knowledge that must be paid to the high-types. The high-type contract, by contrast, is not incentive-compatible for the low-types and, thus, implies an effort distortion. Further, depending on their reservation utility, physicians may earn an information rent resulting from the non-negativity limitation of the capitation payment under both contracts. This leads to a rent-efficiency trade-off and the dominating contract is particularly determined by the share of the high-types in the physician collective. Offering the low-type contract and, therefore, incentivizing both types to exert diagnostic effort and allowing the high-types to mimic the low-types, is only second-best optimal, if the share of the high-types is not too large. Distinguishing between the effects from moral hazard and adverse selection further reveals that, again, depending on the share of the high-types in the market, accounting for adverse selection in addition to moral hazard may have controversial effects on the extent of low-types' effort distortion. While all first-best solutions may also be second-best optimal, the parameter constellations, for which the provision of incentives for an informed treatment choice is second-best optimal, are restricted compared to the First-Best. Moreover, against standard principal-agent theory, I find that a contract consisting of a menu of policies that aims at inducing self-selection among the physicians does not improve on this outcome. Finally, the main findings are robust to an alternative operationalization of heterogeneities in physician ability by means of heterogeneities in diagnostic efficiency.

The single contracting problems outlined above have already been addressed by theoretical papers from the health economics field. Regarding the central problem of unobserved heterogeneous ability, the literature mainly focuses on referrals by gatekeeping general practitioners (GPs) rather than on physicians' diagnostic and treatment choice. Moreover, optimal incentives are not explicitly derived (Allard et al. 2011; Allard et al. 2014). The contributions by Jelovac (2001), Garcia Mariñoso and Jelovac (2003), Alger and Ma (2003) as well as Gottlieb and Moreira (2014) are particularly relevant to this paper, whereby Jelovac (2001) is the most closely related. The author considers a treatment choice model with endogenous diagnostic effort. Comparable to this paper, physicians may exert costly effort to get a private signal on the patients' illness. Yet, the model is limited to double moral hazard because it does not capture heterogeneities in ability. A payment scheme consisting of reward payments for successful treatment at first try is found to provide optimal incentives. Most notably, the scheme represents cost-sharing and has a pay-for-performance character. It is assumed that patients, who are not cured at first try, return to the physician and have to receive additional treatment, while the physician receives zero payments for these patients. Garcia Mariñoso and Jelovac (2003) applied the general findings by Jelovac (2001) to a gatekeeping

GP's referral problem and complemented the payment scheme with a capitation payment, which allows to incorporate physicians' reservation utility. Allard et al. (2011) and the companion paper by Allard et al. (2014) also model a gatekeeping GP's referral problem, whereby GPs are heterogeneous in both diagnostic ability and altruism. Allard et al. (2011) assume that the diagnostic accuracy depends on the exogenous ability of the GP. Based on a first-best analysis, they investigate, which pure payment system (FFS, fundholding or capitation) is most efficient given a GP's ability-altruism-profile. They find that it essentially depends on the distribution of profiles among the GPs. Allard et al. (2014) extend the analysis by considering endogenous diagnostic effort (ability) and adverse selection. By modelling GPs' behavior when they are paid by pure FFS, pure capitation or can self-select into either pure payment system, they show that GPs' self-selection is never optimal. The high-ability GPs will always mimic the low-ability GPs and, given their type, opt for the inefficient capitation system. This may call for a more sophisticated payment mechanism. Analyses by Alger and Ma (2003) reinforce the importance of the distribution of types in the market as shown by Allard et al. (2011). They examine optimal contracts for physicians that differ in their decision to collude with the patient when reporting to the payer. In case the payer offers a single policy contract, the distribution of the types essentially determines, whether a collusion- or a non-collusion-proof contract is superior. Only if the share of the dishonest physicians is sufficiently low, it is best to tolerate collusion. Moreover, they find that implementing a menu of policies would increase the overall welfare. Generally, classical principal-agent theory suggests that in situations with more than one type, contracts consisting of a menu of policies, which aim at inducing self-selection among the agents, provide optimal incentives (Laffont and Tirole 1993; Laffont and Martimort 2002). Yet, Gottlieb and Moreira (2014) show that for incentive problems in insurance and procurement with simultaneous adverse selection and moral hazard, offering a menu of policies may be inefficient. Moral hazard entails additional incentive-compatibility constraints (ICs) and agents, who exert low effort, may earn positive rents from mimicking high effort types. Because the ICs are binding, implementing a menu would imply a further effort distortion. This contrasts the *no distortion at the top* property of screening contracts.

The remainder of the paper is organized as follows: In the next section, I will introduce the model. Section 5.3 presents the First-Best. The implementation of the First-Best, considering single policy contracts as well as contracts that consist of a menu of policies, is established in Section 5.4. Section 5.5 thereupon presents the Second-Best. In Section 5.6, I account for the alternative operationalization of physicians' heterogeneous diagnostic ability by means of diagnostic efficiency. Finally, Section 5.7 concludes this paper.

5.2 The Model

5.2.1 The Patient

A patient who experiences symptoms of illness visits a physician. He suffers from a severe illness with probability $p \in (0;1)$ and from a mild illness with probability $(1-p)$. Probability p may be interpreted as the prevalence of the severe illness in the patient population that sees a physician and is common knowledge. While the mild illness may be cured with either a basic treatment T^B or an advanced treatment T^A, the severe illness may only be cured with an advanced treatment T^A. Eventually, the patient must be cured. Thus, in case a severely ill patient initially receives T^B, the treatment is insufficient to cure the illness and the patient returns to the physician to receive T^A on top. Following Jelovac (2001), I assume that the patient will not fall ill during the game, but only before. This implies that in case the severely ill patient returns to the physician to receive T^A on top of T^B, it is traced to the prior insufficient treatment choice. Further, in case a severely ill patient initially receives the insufficient T^B, he suffers from a health loss $\ell > 0$ due to a delayed cure. The patient is assumed to be fully insured and passive, in the sense that he always consents to the physician's actions.

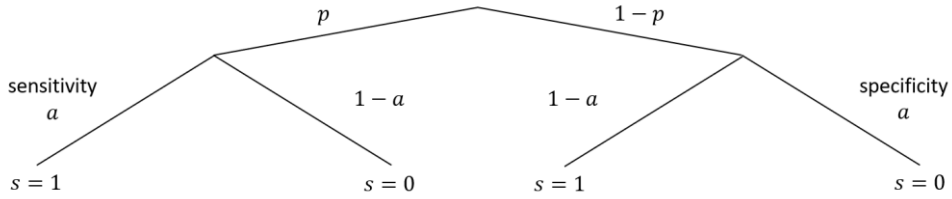
5.2.2 The Physician

When initially seeing a patient, the physician cannot tell whether the patient suffers from a mild or a severe illness. She only knows the prevalence of the severe illness p in the patient population that comes to see her.[1] To cure the patient, the physician makes two choices: Firstly, she decides on exerting diagnostic effort and, secondly, she decides on the treatment intensity. The choices are detailed in the following two paragraphs.

The diagnostic effort choice: At the beginning of a diagnosis and treatment course, the physician makes a discrete diagnostic effort choice $e \in \{0,1\}$.[2] Exerting effort ($e = 1$) yields a noisy diagnostic signal $s \in \{0,1\}$ on the patient's type of illness. The effort decision is hidden action, and the generated signal is hidden information. Hence, the payer can neither observe whether the physician exerted effort, nor which signal it resulted in. Figure 5-1 depicts the characteristics of the diagnostic signal. Following Allard et al. (2014), I suppose that the probability that signal s is correct is determined by the physician's individual diagnostic accuracy a.

[1] I adopt the linguistic convention that the patient and the payer are male, whereas the physician is female.
[2] While Jelovac (2001) allows for a continuous effort variable, for simplicity, I model effort as a binary choice.

Figure 5-1: The Diagnostic Signal



Note: This figure displays the characteristics of the private, noisy signal s on the patient's type of illness, which the physician receives from exerting diagnostic effort. p denotes the prevalence of the severe illness and $(1-p)$ denotes the prevalence of the mild illness in the population of patients that see a physician. a stands for the physician's diagnostic accuracy and may be interpreted as the sensitivity and specificity of her effort.

Thus, heterogeneities in diagnostic accuracy may serve as a measure for heterogeneities in the physician's exogenous diagnostic ability (Chatterjee et al. 2019; National Academies of Sciences, Engineering, and Medicine 2015). Diagnostic accuracy (ability) may be interpreted as the sensitivity (se) and specificity (sp) of the physician's effort and, to simplify the mathematical exposition, I suppose that $se = sp = a$ (compare Garcia Mariñoso and Jelovac 2003). While Allard et al. (2014) model ability continuous, for simplicity, I model ability binary with $a \in \{\underline{a}, \overline{a}\}$. Consequently, there are high ($\overline{a}$-types) and low ($\underline{a}$-types) ability physicians, whereby λ is the share of the $\overline{a}$-types in the physician collective. The ability-type is private information to the physician. The payer only has prior beliefs about the physician's type and is aware of the share of $\overline{a}$-types. Further, $0.5 < \underline{a} < \overline{a} < 1$, implicating that, for both ability-types, exerting diagnostic effort is informative yet imperfect and the signal is more precise for the $\overline{a}$-types. Moreover, exerting effort comes at uniform utility cost of $v > 0$ for both ability-types. Summing up, heterogeneities in physician ability are initially measured by differing production functions of diagnostic insights (accuracy), whereas the effort cost functions (efficiency) are similar.[3] Bayesian updating yields the following probabilities to suffer from the severe illness:

$$\begin{aligned} &\text{For signal } s = 0: \qquad p_0 = \frac{p(1-\mathrm{a})}{p(1-\mathrm{a})+(1-p)a} < p \\ &\text{For signal } s = 1: \qquad p_1 = \frac{pa}{pa+(1-p)(1-\mathrm{a})} > p \end{aligned} \tag{1}$$

[3] In Section 5.6, heterogeneities in ability will alternatively be operationalized by means of heterogeneities in diagnostic efficiency. Certainly, physicians may also be heterogeneous in both diagnostic accuracy and efficiency. Nevertheless, for simplicity and to identify possible differing effects, I consider both measures separately.

This implies $p_0^{\overline{a}} < p_0^{a} < p$ and $p < p_1^{a} < p_1^{\overline{a}}$.

In case the physician exerts no effort ($e = 0$), she receives no further information on the patient's state of health and, therefore, her ability-type is irrelevant.

The treatment choice: The physician decides on treatment intensity $\tau \in \{T^B, T^A\}$. She may choose basic treatment T^B or advanced treatment T^A. The cost for basic treatment are normalized to zero, whereas the cost for advanced treatment amount to $h > 0$. The physician's objective is to maximize her expected utility, which is defined as her expected income from the payments she receives from the payer minus the expected utility cost she incurs from diagnosing the patient (see equation (3) in the next subsection). I further assume that the physician is concerned only with her own income.

5.2.3 The Payer

The payer aims at maximizing the expected social welfare. Therefore, he accounts for the expected overall cost of care provision and the monetary equivalent of the patient's expected utility from a diagnosis and treatment course. While the former cost encompass the physician's effort cost, the additional cost for T^A compared to T^B and the payments to the physician, the latter arise from the patient's possible health loss from delayed T^A.[4] Thus, the payer's objective may also be stated as minimizing the sum of the expected cost from providing optimal diagnosis and treatment to the patient and the monetary equivalent of the patient's expected health loss. In the following, I refer to this as the expected social cost (ESC) as outlined in equation (2) below.

To achieve his objective, the payer designs a payment scheme (D, B, A) for the physician, which consists of three non-negative instruments: When seeing a patient, the physician always receives a capitation payment D for performing a diagnosis. Additionally, she receives FFS payments, each consisting of a cost-based component for the respective treatment and a mark-up, which is only paid conditional on adequate treatment choice.[5] For treatments T^B and T^A mark-ups B and A are paid, respectively, and treatment is presumed to be adequate if it cures the patient right away. Hence, in case the payer sets $B, A > 0$, the physician earns a positive mark-up on the respective treatment cost if she initially provides T^B for a mildly ill patient (mark-up B) or if she initially provides T^A for either patient (mark-up A). This implies that the

[4] It should be noted that, apart from the possible health loss from delayed advanced treatment, the patient's utility from a diagnosis and treatment course is normalized to zero.

[5] Garcia Mariñoso and Jelovac (2003) propose a similar payment scheme to incentivize general practitioners in a gatekeeping system to make an optimal referral decisions.

physician also receives mark-up A in case she provides T^A for a mildly ill patient. If, by contrast, the physician initially applies T^B for a severely ill patient, she receives no mark-up because she has to apply T^A on top of T^B to cure the patient. In case $B = A = 0$, the physician solely receives the cost-based FFS component for the respective treatment, independent of an adequate treatment choice. Therefore, the payer's objective function is given by:

$$\min_{(A,B,D)} ESC = EC(h, v) + EL(\ell) \tag{2}$$

Given payment scheme (D, B, A), the physician picks the diagnosis and treatment course that maximizes her expected utility:

$$\max_{e,\tau} EU = EI(D, B, A) - EC(v) \tag{3}$$

Motivated by the circumstance that, due to alternative income options, the physician has some negotiation power when signing a contract, I suppose that the payer must respect a participation constraint (PC). The physician's expected utility from a diagnosis and treatment course must be at least equal to her reservation utility ω. Furthermore, I require the capitation payment D to be non-negative. This assumption is based on the observation that payments by physicians to treat a patient do not exist. The most likely reason is that physicians will refuse contracts with $D < 0$ even if their PC is met.[6]

5.2.4 The Timing

The game comprises 8 stages. In stage 1, the payer sets the physician's payment contract (D, B, A). When designing the payment instruments, the payer is aware of the treatment related cost (h, ℓ), the physician's potential effort cost v, the prevalence p of the severe illness in the patient population that sees a physician, the share of the $\overline{a}$-types λ in the physician collective and the physician's reservation utility ω. In stage 2, the physician either accepts or rejects this contract. In the latter case, the game ends. In the former case, the game continues in stage 3 and nature determines the patient's state of health. With probability p the patient suffers from the severe illness and with probability $(1 - p)$ he suffers from the mild illness. Either way, the patient seeks care from the physician. In stage 4, the physician decides on exerting diagnostic effort $e \in \{0,1\}$. In case the physician exerts effort, in stage 5, nature sends an imperfect diagnostic signal. The precision of the signal increases in the physician's diagnostic ability. In case the physician forgoes exerting effort, the game directly continues in stage 6. In stage 6, the

[6] Mougeot and Naegelen (2005) provide a similar argument in the hospital context.

physician decides on treatment intensity $\tau \in \{T^B, T^A\}$. In case she exerted diagnostic effort before, she bases the decision on her private diagnostic signal. In case, she did not exert effort, she only knows prevalence p and makes a blind choice. If the chosen treatment cures the patient right away, the game ends and payments are made. If the chosen treatment turns out to be insufficient to cure the patient, in stage 8, the patient returns to receive sufficient treatment, and payments are made.

5.3 The First-Best

In this section, I derive the diagnosis and treatment courses, which minimize the expected social cost and assuming that there are no information asymmetries – this defines the First-Best. The payer can observe the physicians' diagnostic effort decision, the resulting signal as well as their ability-type. The ESC arise from the share of patients receiving T^A at cost h, the share of patients incurring a health loss due to a delayed cure with the monetary equivalent of ℓ, the share of patients for whom the physician exerts diagnostic effort at utility cost of v and the physicians' reservation utility ω. Given the fact that exerting diagnostic effort is costly and the resulting signal is indeed informative, I identify three first-best candidates:

- *Course 1 "Blind treatment choice, always T^B"*: The physician exerts no effort and initially treats the patient with T^B, for which cost are normalized to zero. Yet, with probability p, the patient suffers from the severe illness and T^B is insufficient to cure him. Thus, the patient is additionally treated with T^A at cost of h and incurs the health loss ℓ. This adds up to $ESC_1 = p(h + \ell) + \omega$.
- *Course 2 "Blind treatment choice, always T^A"*: The physician exerts no effort and initially treats the patient with T^A. As T^A always cures the patient and comes at cost of h, this amounts to $ESC_2 = h + \omega$.
- *Course 3 "Informed treatment choice*": The physician exerts diagnostic effort and incurs effort cost v. She follows the signal such that, in case $s = 0$, the patient receives T^B and, in case $s = 1$, the patient receives T^A. With probabilities a and $(1 - a)$ the physician receives signal $s = 1$ for a severely ill and a mildly ill patient, respectively. Cost h arise from T^A. With probability $(1 - a)$ signal $s = 0$ is received for a severely ill patient. Because T^B does not cure the patient, he additionally receives T^A at cost of h and incurs health loss ℓ. This yields $ESC_3 = v + [pa + (1 - p)(1 - a)]h + p(1 - a)(h + \ell) + \omega$.

The ESC from the blind courses are independent of the physicians' ability as no effort is exerted.

By contrast, the ESC from the informed treatment choice course decrease in ability $\left(\frac{\partial ESC_3}{\partial a} < 0\right)$ because the precision of the diagnostic signal increases in ability. The pairwise comparison of the ESC arising from the three courses by ability-type yields the first-solutions. Figure 5-2 depicts the first-best solutions depending on the treatment related cost, which are the additional cost h for T^A and the monetary equivalent of the patient's expected utility loss ℓ from delayed T^A. Formally, the first-best solutions are stated in Lemma 1:

Lemma 1:

Course 1 "Blind treatment choice, always T^B" is optimal if $\ell \leq \min\left\{\frac{1-p}{p}h,\ \frac{(1-p)(1-a)h+v}{pa}\right\}$.

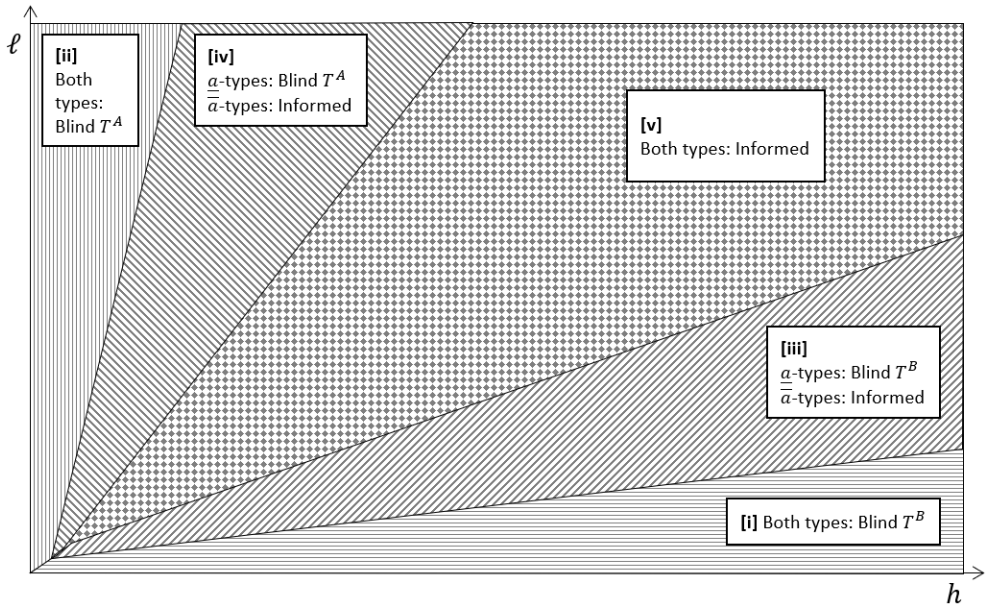
Course 2 "Blind treatment choice, always T^A" is optimal if $\ell \geq \max\left\{\frac{1-p}{p}h,\ \frac{(p-1)ah+v}{p(a-1)}\right\}$.

Course 3 "Informed treatment choice" is optimal if $\frac{(p-1)ah+v}{p(a-1)} \geq \ell \geq \frac{(1-p)(1-a)h+v}{pa}$.

Whereby, $a \in \{\underline{a}, \overline{a}\}$ *with* $\underline{a} < \overline{a}$.

For relatively low values of ℓ (h) and across values of h (ℓ), both ability-types should forego exerting diagnostic effort and pick a blind treatment strategy. In area [i], the health loss ℓ is small and T^A is relatively expensive. Therefore, independent of the type, it is optimal to first give T^B a try. In area [ii], the additional cost for T^A are small, while the negative effects of delayed treatment are relatively high. Hence, for both types, it is optimal to apply T^A right away. For moderate values of ℓ (h) and across values of h (ℓ), the ability-types should pick different courses: The $\overline{a}$-types should always exert diagnostic effort and base the treatment choice on the resulting signal. Because the stakes are higher than in areas [i] and [ii], for the $\overline{a}$-types it is worthwhile to exert costly diagnostic effort in order to subsequently save on the treatment related cost (h, ℓ). The $\underline{a}$-types should still forgo exerting diagnostic effort and continue to pick a blind treatment strategy. The signal, which the $\underline{a}$-types would receive from exerting diagnostic effort, would not be precise enough to generate savings on the treatment related cost that compensate for the effort cost. For moderate values of ℓ, the $\underline{a}$-types should try T^B first (area [iii]) and for moderate values of h, T^A should be chosen right away (area [iv]). In area [v], the treatment related cost (h, ℓ) are relatively high, as a result of which, also for the $\underline{a}$-types, the savings on the treatment related cost outweigh the effort cost. Hence, at the optimum, both types should exert diagnostic effort.

Figure 5-2: The First-Best Allocation given $a \in \{\underline{a}, \overline{a}\}$



Note: This figure depicts the first-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given physicians differ in their diagnostic ability, which is operationalized by means of heterogeneities in diagnostic accuracy (the production function) $a \in \{\underline{a}, \overline{a}\}$. There are high-ability physicians ($\overline{a}$-types) and low-ability physicians ($\underline{a}$-types). h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 v=2, $\underline{a}$=0.65 and $\overline{a}$=0.85.

5.4 Implementation of the First-Best Allocation and Rent Payments

In this section, I analyze the implementation of the first-best allocation in the context of asymmetric information. I suppose that the physicians' diagnostic effort decision is hidden action, the generated signal is hidden information and physicians' ability-type is hidden knowledge.

5.4.1 Both Ability-Types Make a Blind Treatment Choice

In case both ability-types should follow the same blind diagnosis and treatment course (see area [i] and area [ii] in Figure 5-2), the implementation is straight forward.[7] By setting $B = A = 0$, both types earn zero mark-ups, are indifferent between treatments and will, therefore, opt for the first-best. Because zero mark-ups are paid, there are no incentives to exert costly diagnostic effort. Subsequently, the capitation payment D may be set such that the physician just earns her reservation utility ω. Despite the information asymmetry regarding the physicians' ability-type, the payer can provide optimal incentives without paying an information rent.

[7] Hereinafter, I assume that with indifference, the physician will opt for the diagnostic and treatment course that causes the lower expected social cost – in this paper defined as the First-Best.

5.4.2 At Least One Ability-Types Makes an Informed Treatment Choice

The implementation of areas [iii] to [v] in Figure 5-2 is more demanding because they require at least the $\overline{a}$-types to exert costly diagnostic effort and to follow the resulting private signal. For now, I take a step back and suppose that the payer can observe the physicians' ability-types. This assumption will be relaxed again at a later stage in this subsection. To determine, if and at what cost areas [iii] to [v] may be implemented, I focus on the optimal incentives for the informed treatment choice course. The proposed payment scheme (D, B, A) needs to align the physicians' privately optimal choices with the socially optimal choices. Solving the problem by backwards induction, I start with the physicians' treatment decision (compare Beenk and Kifmann 2021; Garcia Mariñoso and Jelovac 2003).

Inciting the optimal treatment intensity: Firstly, the physician must be incentivized to comply with the signal, which she receives from exerting diagnostic effort. Patients, for whom she receives signal $s = 0$ should be treated with T^B, whereas for patients, for whom she receives signal $s = 1$, T^A should be applied. The physicians' expected payoffs from choosing either treatment amount to:

Signal	Payoff from T^B	Payoff from T^A
$s = 0$	$\boldsymbol{D + (1 - p_0)B}$	$D + A$
$s = 1$	$D + (1 - p_1)B$	$\boldsymbol{D + A}$

Consequently, the physician will only choose T^B if $(1 - p_0)B \geq A$ and only choose T^A if $(1 - p_1)B \leq A$. This gives the following incentive constraints (ICs):

$$A \leq (1 - p_0)B \tag{4}$$

$$A \geq (1 - p_1)B \tag{5}$$

Summing up, an informed treatment choice calls for $(1 - p_0)B \geq A \geq (1 - p_1)B$.

Inciting diagnostic effort: Secondly, the physician must be provided with incentives to exert diagnostic effort ($e = 1$) in the first place. She takes the effort decision independent of any preceding events. Thus, in case she exerts effort, and the above ICs are met, an informed treatment choice amounts to an expected utility of:

$$EU_3 = D - v + (1 - p)aB + [pa + (1 - p)(1 - a)]A \tag{6}$$

The physician is paid capitation D for each patient she sees. Exerting diagnostic effort comes at utility cost v. In case she receives signal $s = 0$, she initially chooses T^B and receives mark-up B for patient share $(1-p)a$. In case she receives signal $s = 1$, she applies T^A right away and receives mark-up A for patient share $pa + (1-p)(1-a)$.

In case she forgoes exerting diagnostic effort, she has the choice between the two blind courses, which result in the expected utilities:

$$EU_1 = D + (1-p)B \tag{7}$$

$$EU_2 = D + A \tag{8}$$

Thus, exerting diagnostic effort is also in the physicians' private best interest if mark-up payments B and A fulfill the condition $EU_3 \geq \max\{EU_1, EU_2\}$. Inserting from (6), (7) and (8) as well as substituting from (1) and rewriting, yields the additional incentive constraints:

$$A \geq (1-p_1)B + \frac{v}{pa+(1-p)(1-a)} \tag{9}$$

$$A \leq (1-p_0)B - \frac{v}{p(1-a)+(1-p)a} \tag{10}$$

Effort cost v are strictly positive. Thus, it is obvious that IC (9) implies IC (5) and IC (10) implies IC (4). Further, ICs (9) and (10) clearly show that $B > A > 0$ is a necessary condition to implement an informed treatment choice. If A was equal to or larger than B, the physician would always choose T^A right away, leading to overtreatment of the mildly ill patients.

Because capitation payment D has no incentive effect and low values of D should be avoided, focusing on values (B, A), which minimize the expected payments to the physician subject to the incentive-compatibility constraints, solves the payer's optimization problem:

$$\min_{(B,A)}\{D + (1-p)aB + [pa + (1-p)(1-a)]A\},$$

s.t. ICs (9) and (10).

As increasing B furthers (10), but counters (9) and increasing A furthers (9), but counters (10), both effort constraints bind.[8] Adding up ICs (9) and (10) yields:

$$A = (1-p)B \tag{11}$$

[8] It should be noted that higher values of A and B may also solve the incentive problem. However, due to the limitation of capitation payment D to non-negative values, the physician's earnings from higher reward payments may not be extracted ex-post and, therefore, higher reward payments may be inefficient.

Inserting for A into either effort constraint and, afterwards, using (11) yields

$$B^* = \frac{v}{p(1-p)(2a-1)} \qquad \text{and} \qquad A^* = \frac{v}{p(2a-1)}.$$

The optimal mark-up payments both decrease in ability $\left(\frac{\partial B^*}{\partial a} < 0,\ \frac{\partial A^*}{\partial a} < 0\right)$. The higher the physicians' ability, the more accurate is their signal from exerting diagnostic effort. Hence, given both types exert effort, $\overline{a}$-types will earn mark-up B^* for successfully treating mildly ill patients with T^B and mark-up A^* for treating severely ill patients with T^A with a higher probability than the $\underline{a}$-types. By contrast, mark-up A^* for treating mildly ill patients with T^A will be earned with a lower probability. The former effects always dominate because $B^* > A^*$. Thus, at the optimum, the mark-ups, which serve as a compensation for exerting costly effort, must be set lower for the $\overline{a}$-types than for the $\underline{a}$-types.

Derivation of the optimal capitation payment $D^*$and possible information rents: To determine the least costly combination of payments (D, B, A) that incites the physicians to perform an informed treatment choice, the payer sets D just only as high as necessary to satisfy the physicians' PC $EU_3 \geq \omega$. From inserting B^* and A^* into the physicians' utility function follows $D = \omega - \frac{v}{p(2a-1)}$. Because effort cost v are strictly positive, this possibly calls for a negative D. As already argued, it is unlikely that physicians accept a payment contract that entails paying upfront to treat a patient. Therefore, capitation payment D is essentially constraint by:

$$D^* = \max\left\{0, \omega - \frac{v}{p(2a-1)}\right\}$$

The findings are summarized in Lemma 2:

Lemma 2: *A mixed payment contract provides incentives for the implementation of an informed treatment choice course. The contract consists of FFS payments with mark-ups $B > A > 0$ on the respective treatment cost of T^B and T^A, conditional on adequate treatment choice, and a capitation $D \geq 0$, which is set as low as possible to fulfill the physicians' PC.*

In principal-agent situations, in which the agent cannot be punished because the principal's instruments are limited to reward payments only, the principal may have to grant the agent some information rent (Laffont and Martimort 2002, p. 155-57; Beenk and Kifmann 2021). This incentive problem also arises in my model: While mark-up payments (A, B) serve as rewards, capitation payment D cannot be negative. Thus, the physicians' PC $EU_3 \geq \omega$ may not bind, and

they may earn an information rent. Particularly, if $D^* > D$, providing optimal incentives to the physicians would require paying them an information rent:

$$r = \frac{v}{p(2a-1)} - \omega$$

Note that the rent decreases in the physicians' ability $\left(\frac{\partial r}{\partial a} < 0\right)$. The higher their ability, at the optimum, the lower is their income from mark-up payments (A^*, B^*) and the higher the capitation payment D^* can be set by the payer without having to pay a rent. Thus, the scope for D^* is larger. Moreover, the smaller the physicians' reservation utility, the larger the potential rent payments.

From now on, I account for the circumstance that the physicians' ability-type is hidden knowledge, which implies that the payment instruments cannot depend on ability. This does not mean that it will be impossible to induce different types to follow the respective first-best diagnosis and treatment course, but it may be costly to the payer to implement such allocation. In particular, the $\overline{a}$-types always have an incentive to mimic the $\underline{a}$-types and, therefore, extract a costly information rent. To implement the First-Best in classical principal-agent models with hidden knowledge and two types of agents, the payer can either offer a contract consisting of a single policy, or he can offer a more complex contract consisting of a menu of two policies. Usually, the menu is found to be superior (Laffont and Martimort 2002).

5.4.2.1 Implementation with a Contract Consisting of a Single Policy

Supposing that the payer may offer a contract consisting of a single policy, it is straightforward that he initially can design a contract that serves the $\overline{a}$-types or a contract that serves the $\underline{a}$-types. The contract candidates are outlined in Lemma 3:

Lemma 3: *When only considering single policy contracts, the payer may offer contract* $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ *or contract* $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$*, which are designed for the* $\overline{a}$*-types and the* $\underline{a}$*-types, respectively:*

- *Contract* $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ *consists of payments*
  $\overline{B}^* = \frac{v}{p(1-p)(2\overline{a}-1)}$, $\quad \overline{A}^* = \frac{v}{p(2\overline{a}-1)}$., $\quad \overline{D}^* = max\left\{0, \omega - \frac{v}{p(2\overline{a}-1)}\right\}$
- *Contract* $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ *consists of payments*
  $\underline{B}^* = \frac{v}{p(1-p)(2\underline{a}-1)}$, $\quad \underline{A}^* = \frac{v}{p(2\underline{a}-1)}$, $\quad \underline{D}^* = \max\left\{0, \omega - \frac{v}{p(2\underline{a}-1)}\right\}$

Whereby, $\underline{B}^* > \overline{B}^*$, $\underline{A}^* > \overline{A}^*$ *and* $\underline{D}^* \leq \overline{D}^*$.

Given an unchanged contract structure, the payer could also offer contracts that serve neither type directly. Yet, because effort is modelled binary, these would be inefficient. For instance, if the payer chose an average contract $(\tilde{B}, \tilde{A}, \tilde{D})$ with payments $\underline{B}^* > \tilde{B} > \overline{B}^*$, $\underline{A}^* > \tilde{A} > \overline{A}^*$ and $\underline{D}^* \leq \tilde{D} \leq \overline{D}^*$, then only the $\overline{a}$-types would be incentivized to perform an informed treatment choice at equal or higher cost (depending on the value of ω) than under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Moreover, there might also exist contract candidates with a different structure that provide adequate incentives. However, these analyses would go beyond the scope of this paper. Hence, I will only focus on the single policy contracts, which are defined by Lemma 3 and, hereinafter, also referred to as "incentive contracts".

Presented a payment contract, I assume that the physicians always decide on the diagnosis and treatment course that yields them the highest expected utility. The expected utilities under contracts $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ and $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ are derived in Appendix A.

In case contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ is offered

- $\overline{a}$-types' expected utilities amount to:
  $EU_1^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = EU_2^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = EU_3^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$
- $\underline{a}$-types' expected utilities amount to:
  $EU_1^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = EU_2^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$
  $EU_3^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\} - \frac{2v(\overline{a}-\underline{a})}{(2\overline{a}-1)}$

Comparing the expected utilities by type reveals that the $\overline{a}$-types are indifferent between courses and, thus, pick the informed treatment choice course.[9] Moreover, in case their PC does not bind, they earn a strictly positive information rent that accrues from the limitation on $\overline{D}^*$. The $\underline{a}$-types, by contrast, always generate a lower expected utility from performing an informed treatment choice than from following a blind course. They are indifferent between the blind courses and, hence, choose the one that yields the lower ESC. Thus, contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ is not incentive-compatible for the $\underline{a}$-types and offering the contract to both types of physicians results in separating diagnosis and treatment decisions. Interestingly, even though their effort is distorted, the $\underline{a}$-types still always earn the exact same information rent as the $\overline{a}$-types, which

[9] When deriving the optimal mark-up payments (A, B), the effort constraints were binding, which implies that, under their designated contract, physicians are always indifferent between courses. Therefore, they will pick the course that yields the lowest ESC given their type – that is the first-best course.

results from the limitation on $\overline{D}^*$. This may be explained by the fact that, when designing the optimal contract, the effort constraints are both binding and, by contrast to EU_3, the expected utilities for the blind courses, EU_1 and EU_2, are independent of diagnostic ability.

In case the payer offers contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$:

- $\overline{a}$-types' expected utilities amount to:

  $$EU_1^{\overline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = EU_2^{\overline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

  $$EU_3^{\overline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\} + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}$$

- $\underline{a}$-types' expected utilities amount to:

  $$EU_1^{\underline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = EU_2^{\underline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = EU_3^{\underline{a}}(\underline{B}^*, \underline{A}^*, \underline{D}^*) = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

Again, comparing the expected utilities separately for each type shows that the $\underline{a}$-types are indifferent between courses and, hence, perform an informed treatment choice. In case their PC does not bind, they earn a strictly positive information rent that results from the limitation on $\underline{D}^*$. The $\overline{a}$-types will also perform an informed treatment choice because this course always generates a higher expected utility than any blind course. Moreover, $\overline{a}$-types always earn a strictly positive information rent. The rent consists of two components: Firstly, a strictly positive component $\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}$, which results from the hidden knowledge about the type. The $\overline{a}$-types' earnings from $(\underline{B}^*, \underline{A}^*)$ are larger than the earnings of the $\underline{a}$-types. So, even if the $\underline{a}$-types' PC binds, $\underline{D}^*$ is too high and $\overline{a}$-types extract a rent. Secondly, in case the $\underline{a}$-types' PC does not bind, they earn the exact same rent as the $\underline{a}$-types, which is due to the limitation on $\underline{D}^*$. Reconciling the incentive effects from the two single policy contracts with the First-Best, I conclude with Lemma 4:

Lemma 4: *Contract* $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ *induces separating equilibria and implements the First-Best in area [iii] and area [iv] in Figure 5-2. The* $\overline{a}$*-types perform an informed treatment choice, while the* $\underline{a}$*-types forgo diagnostic effort and, in area [iii] (area [iv]) treat all patients with* T^B *(*T^A*) first (right away). The implementation may come at a cost: In case the* $\overline{a}$*-types' PC does not bind, both types must be paid the exact same information rent due to the limitation on* $\overline{D}^*$*. Rent payments add up to* $r_{(\overline{B}^*, \overline{A}^*, \overline{D}^*)} = \frac{v}{p(2\overline{a}-1)} - \omega$.

Contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ *induces a unifying equilibrium and implements the First-Best in area [v] in Figure 5-2. Both types perform an informed treatment choice. The implementation is always costly:* $\overline{a}$*-types always earn a strictly positive information rent, which accrues from the hidden knowledge about the type. In case the* $\underline{a}$*-types' PC does not bind, both types further earn an information rent, which results from the limitation on* $\underline{D}^*$.

Rent payments add up to $r_{(\underline{B}^*, \underline{A}^*, \underline{D}^*)} = \lambda \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)} + \frac{v}{p(2\underline{a}-1)} - \omega$.

5.4.2.2 Implementation with a Contract Consisting of a Menu of Policies

For comparison and to examine whether the payer can achieve a superior solution by offering a screening contract, I now suppose that the payer offers a contract consisting of a menu of two policies: $\left\{(\overline{A}^\circ, \overline{B}^\circ, \overline{D}^\circ), (\underline{A}^\circ, \underline{B}^\circ, \underline{D}^\circ)\right\}$. The screening contract aims at inducing self-selection among the physicians. Thus, both ability-types need to be incentivized to pick the policy that was designed for them, which is $(\overline{A}^\circ, \overline{B}^\circ, \overline{D}^\circ)$ for the $\overline{a}$-types and $(\underline{A}^\circ, \underline{B}^\circ, \underline{D}^\circ)$ for the $\underline{a}$-types.

The standard approach to this sort of incentive problem is to design a screening contract that trades off the information rent earned by the $\overline{a}$-types with the allocative efficiency of the $\underline{a}$-types. Ultimately, this leads to granting the $\overline{a}$-types less information rent and, in exchange, accepting an effort distortion by the $\underline{a}$-types, which is penalized by paying zero rents to the latter (Bolton and Dewatripont 2005, p. 94). However, in this model, the $\underline{a}$-types' outside option always pays them the exact same rent as the $\overline{a}$-types receive. So even if $\underline{a}$-types' effort is distorted, they may earn a strictly positive information rent. This possible equilibrium is already implemented by single policy contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ because effort is binary. Thus, I conclude with Lemma 5.

Lemma 5: *The classical finding from binary principal-agent models with hidden knowledge, which is that, compared to a single policy contract, a contract consisting of a menu of policies improves on the outcome, does not apply to this model. In fact, the optimal menu consists of a single non-discriminatory policy* $(\overline{A}^\circ, \overline{B}^\circ, \overline{D}^\circ) = (\underline{A}^\circ, \underline{B}^\circ, \underline{D}^\circ)$*, which coincides with single policy contract* $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$.

Proof. See Appendix B.

5.5 The Second-Best

In the following, I determine the diagnosis and treatment courses as well as the corresponding contracts, which minimize the expected social cost and assuming that the payer can neither observe the physicians' diagnostic effort decision, nor the resulting signal or their diagnostic ability – defined as the Second-Best. I will only focus on the single policy contracts because contracts consisting of a menu of policies were found to not further improve on the outcome.

The provision of optimal incentives to the physicians may entail paying them a costly information rent. Consequently, in the second-best situation, the payer faces a rent-efficiency trade-off. In particular, focusing on areas [iii] to [v] in Figure 5-2, which are implemented by incentive contracts $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ and $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$, he has to trade-off the rent payments resulting from a possible limitation on $\underline{D}^*$ to both types and from the hidden knowledge about the ability-type to the $\overline{a}$-types against the effort distortion cost due to the $\underline{a}$-types' inefficient diagnosis and treatment choice and possible savings in rent payments from the $\underline{a}$-types' lower earnings from $(\overline{B}^*, \overline{A}^*)$. Analogously to the first-best analyses, the ESC arise from the shares of patients causing treatment related cost (ℓ, h), the share of patients for whom the physicians bear effort cost v and the physicians' reservation utility ω. Further, the rent payments, expressed in terms of the previous parameters, must be considered.

5.5.1 Identification of the Dominant Incentive Contract

In a first step, the dominant single policy incentive contract needs to be identified by comparing the ESC that amount when either contract is offered to the physicians.

Under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$, the $\overline{a}$-types perform an informed treatment choice, which causes effort cost v, while the $\underline{a}$-types pick the blind course that yields the lower ESC. For each type, the treatment related cost (ℓ, h) are determined analogously to Section 5.3. Both types earn at least their reservation utility ω and, in case the $\overline{a}$-types' PC does not bind, further earn a strictly positive information rent. This adds up to

$$ESC(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \lambda[v + [p\overline{a} + (1-p)(1-\overline{a})]h + p(1-\overline{a})(h+\ell)] + (1-\lambda)\min\{p(h+\ell), h\} + \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}.$$

Under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$, both types perform an informed treatment choice and exert diagnostic effort, which causes effort cost v. Both types earn at least their reservation utility ω and, in case the $\underline{a}$-types' PC does not bind, further earn a strictly positive information rent. The

$\overline{a}$-types (with a share of λ) additionally generate an excess rent from the hidden knowledge about the type. This sums up to

$$ESC\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = v + \lambda\left[[p\overline{a} + (1-p)(1-\overline{a})]h + p(1-\overline{a})(h+\ell) + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] + (1-\lambda)\left[\left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell)\right] + \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}.$$

Contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ dominates contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ if and only if

$$ESC(\overline{B}^*, \overline{A}^*, \overline{D}^*) \geq ESC\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) \tag{16}$$

Plugging in for $ESC(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ and $ESC\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ into equation (16) yields Lemma 6:

Lemma 6:

Contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ dominates contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ if and only if:

- *(I) In case ω takes relatively large values, $\omega \geq \frac{v}{p(2\underline{a}-1)}$, which is in case both ability-types' PCs bind: $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} > \ell > \frac{(1-p)(1-\underline{a})h+v}{p\underline{a}}$ and, for $\ell \geq \frac{(1-p)h}{p}$ $\left(\ell \leq \frac{(1-p)h}{p}\right)$,*

  $\lambda \leq \underbrace{\frac{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell}{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_1}$ $\left(\lambda \leq \underbrace{\frac{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell}{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_2}\right)$

- *(II) In case ω takes intermediate values, $\frac{v}{p(2\underline{a}-1)} > \omega \geq \frac{v}{p(2\overline{a}-1)}$, which is in case only the $\overline{a}$-types' PC binds: $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)} - \frac{\omega}{p(\underline{a}-1)} > \ell > \frac{v+(1-p)\left(1-\underline{a}\right)h-\omega}{p\underline{a}} + \frac{v}{p^2\underline{a}(2\underline{a}-1)} - \frac{\omega}{p\underline{a}}$ and, for $\ell \geq \frac{(1-p)h}{p}$ $\left(\ell \leq \frac{(1-p)h}{p}\right)$,*

  $\lambda \leq \underbrace{\frac{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_3}$ $\left(\lambda \leq \underbrace{\frac{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_4}\right)$

- *(III) In case ω takes relatively small values, $\frac{v}{p(2\overline{a}-1)} > \omega$, which is in case neither ability-types' PC binds: $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} > \ell > \frac{v+(1-p)\left(1-\underline{a}\right)h}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)}\Big)$, and, for $\ell \geq \frac{(1-p)h}{p}$ $\left(\ell \leq \frac{(1-p)h}{p}\right)$*

  $\lambda \leq \underbrace{\frac{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(p-1)\underline{a}h+p\left(1-\underline{a}\right)\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_5}$ $\left(\lambda \leq \underbrace{\frac{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}}_{X_6}\right)$

Proof. See Appendix C.

If and only if the treatment related cost (ℓ, h) are relatively high and the cost arising from both cost sources are relatively balanced, it may generally be worth incentivizing both types to perform an informed treatment choice by implementing contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$, compared to only incentivizing the $\overline{a}$-types by implementing contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. The closer the parameter constellations (ℓ, h) are located to the relation $p\ell = (1-p)h$, the more balanced they are. In case the expected cost from delayed adequate treatment for severely ill patients, $p\ell$, equal the expected cost of unnecessarily applying T^A for mildly ill patients, $(1-p)h$, cost sources are perfectly balanced. Generally, the higher and the more uncertain the stakes, the greater the benefit from an informed treatment choice. Further, the share of the $\overline{a}$-types, λ, must not be too large, otherwise the excess rent payments to the $\overline{a}$-types under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ always exceed the effort distortion cost, which arise under the alternative contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$.

5.5.2 The Second-Best Allocation

In a second step, it needs to be verified whether offering the dominant incentive contract is also second-best optimal. Comparing the ESC that arise from the implementation of the dominant incentive contract to the ESC of implementing the alternative blind courses, leads to the second-best solutions, which are defined in Proposition 1.

Proposition 1: *Given heterogeneities in physicians' diagnostic ability are operationalized by means of heterogeneities in diagnostic accuracy, in the second-best situation, offering a non-incentivizing contract with FFS mark-up payments $B = A = 0$ and capitation payment $D = \omega$ is optimal if and only if*

$$\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega, 0\right\}+\lambda[(p-1)\overline{a}h+v]}{\lambda p(\overline{a}-1)}\right\} \quad or$$

$$\ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega, 0\right\}+\lambda[(1-p)(1-\overline{a})h+v]}{\lambda p\overline{a}}\right\}.$$

The conditions, under which offering an incentive contract that either serves the high-ability types or the low-ability types is second-best optimal, particularly depend on the share of the high-ability types, λ, in the physician collective. The contracts are outlined in Lemma 3 and the optimality conditions are presented in Table 5-1.

*Proof. S*ee Appendix D.

In the second-best situation, the implementation of area [i] and area [ii] in Figure 5-2 is always costless, whereas the implementation of area [iii] and area [iv] comes at a cost in case the $\overline{a}$-types' PC does not bind and implementing area [v] is always costly. Therefore, compared to the first-best allocation, I expect the following changes for the second-best allocation:

- In case ω takes relatively large or intermediate values, area [i] and area [ii] Figure 5-2 remain unaffected because incentive contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ may be implemented without paying any information rents to the physicians. Area [iii] and area [iv] expand at the expense of area [v], which results from the trade-off of the $\underline{a}$-types' effort distortion cost under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ against the excess rent payments to the $\overline{a}$-types (and the rents accruing from the limitation $\underline{D}^*$) under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$. The expansion of area [iii] and area [iv] grows in λ, the share of the $\overline{a}$-types.
- In case ω takes relatively small values, area [i] and area [ii] Figure 5-2 expand at the expense of areas [iii] - [v] as the implementation of both incentivizing contracts is costly. The smaller λ, the further expand area [i] and area [ii]. Moreover, area [iii] and area [iv] expand at the expense of area [v], which results from the trade-off of the $\underline{a}$-types' effort distortion cost and the rents accruing from the limitation $\overline{D}^*$ under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ against the excess rent payments to the $\overline{a}$-types and the rents accruing from the limitation $\underline{D}^*$ under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$.

The second-best allocation and the corresponding optimal contracts are depicted in Figure 5-3 to Figure 5-5, separately for the three value ranges of the physicians' reservation utility ω. The figures are based on the same set of parameters that were chosen for the presentation of the first-best allocation in Figure 5-2.

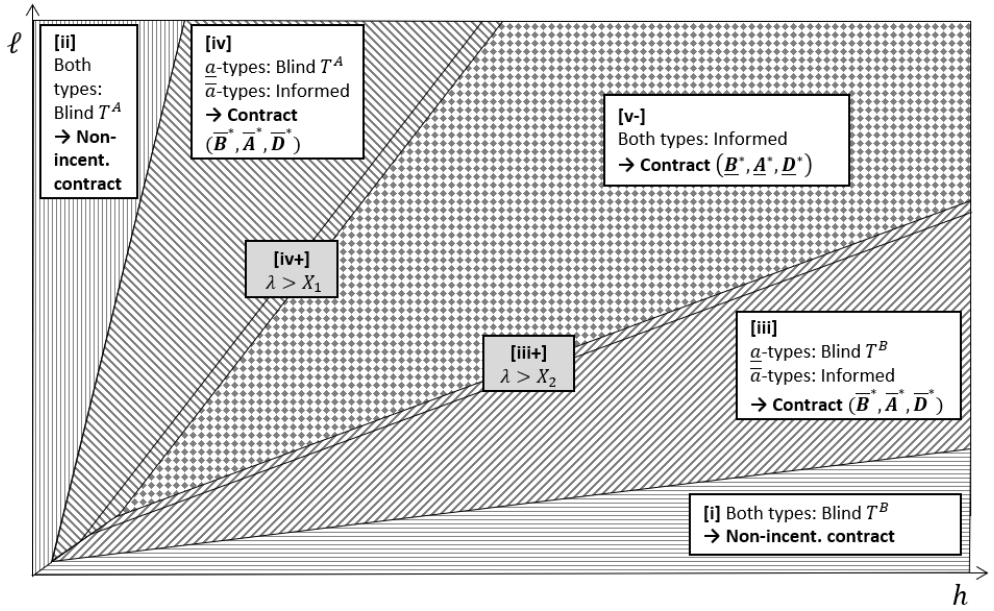
Table 5-1: The Second-Best Optimality Conditions for the Implementation of the Incentivizing Contracts given $a \in \{\underline{a}, \overline{a}\}$

Value range for ω	***Optimality conditions for the implementation of contract $(\overline{\boldsymbol{B}}^*, \overline{\boldsymbol{A}}^*, \overline{\boldsymbol{D}}^*)$, which is designed for the $\overline{\boldsymbol{a}}$-types***	***Optimality conditions for the implementation of contract $(\underline{\boldsymbol{B}}^*, \underline{\boldsymbol{A}}^*, \underline{\boldsymbol{D}}^*)$, which is designed for the $\underline{\boldsymbol{a}}$-types***
(I) The physicians' reservation utility ω takes relatively large values: $\omega \geq \frac{v}{p(2\underline{a}-1)}$	$\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > X_1$ *or* $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > X_2$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(p-1)\overline{a}h+v]}{\lambda p(\overline{a}-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}\right\}$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(1-p)(1-\overline{a})h+v]}{\lambda p\overline{a}} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}\right\}$	$\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_1$ *or* $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_2$
(II) The physicians' reservation utility ω takes intermediate, values: $\frac{v}{p(2\underline{a}-1)} > \omega \geq \frac{v}{p(2\overline{a}-1)}$	$\frac{v+(p-1)\underline{a}h-\omega}{p(\underline{a}-1)} + \frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > X_3$ *or* $\frac{v+(1-p)(1-\underline{a})h-\omega}{p\underline{a}} + \frac{v}{p^2\underline{a}(2\underline{a}-1)} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > X_4$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(p-1)\overline{a}h+v]}{\lambda p(\overline{a}-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{v+(p-1)\underline{a}h-\omega}{p(\underline{a}-1)} + \frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)}\right\}$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(1-p)(1-\overline{a})h+v]}{\lambda p\overline{a}} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{v+(1-p)(1-\underline{a})h-\omega}{p\underline{a}} + \frac{v}{p^2\underline{a}(2\underline{a}-1)}\right\}$	$\frac{v+(p-1)\underline{a}h-\omega}{p(\underline{a}-1)} + \frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_3$ *or* $\frac{v+(1-p)(1-\underline{a})h-\omega}{p\underline{a}} + \frac{v}{p^2\underline{a}(2\underline{a}-1)} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_4$
(*III) The physicians' reservation utility ω takes relatively small values:* $\frac{v}{p(2\overline{a}-1)} > \omega$	For $\lambda \geq \frac{(2\underline{a}-1)\left[[(2\overline{a}-1)-2(\overline{a}-\underline{a})]v-p(2\overline{a}-1)(2\underline{a}-1)\omega\right]}{4(\overline{a}-\underline{a})^2 v}$	
	$\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > X_5$ *or* $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > X_6$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(p-1)\overline{a}h+v]}{\lambda p(\overline{a}-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)}\right\}$ *or* $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(1-p)(1-\overline{a})h+v]}{\lambda p\overline{a}} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)}\right\}$	$\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_5$ *or* $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_6$

Value range for ω	***Optimality conditions for the implementation of contract $(\overline{\boldsymbol{B}}^*, \overline{\boldsymbol{A}}^*, \overline{\boldsymbol{D}}^*)$, which is designed for the $\overline{\boldsymbol{a}}$-types***	***Optimality conditions for the implementation of contract $(\underline{\boldsymbol{B}}^*, \underline{\boldsymbol{A}}^*, \underline{\boldsymbol{D}}^*)$, which is designed for the $\underline{\boldsymbol{a}}$-types***
(III) The physicians' reservation utility ω takes relatively small values: $\frac{v}{p(2\overline{a}-1)} > \omega$	For $\lambda < \frac{(2\underline{a}-1)\left[[(2\overline{a}-1)-2(\overline{a}-\underline{a})]v-p(2\overline{a}-1)(2\underline{a}-1)\omega\right]}{4(\overline{a}-\underline{a})^2 v}$	
	$\min\left\{\frac{\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\right\}$ $> \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > X_5$ *or*	$\min\left\{\frac{\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\right\}$ $> \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_5$ *or*
	$\max\left\{\frac{\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}\right\}$ $< \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > X_6$ *or*	$\max\left\{\frac{\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}\right\}$ $< \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq X_6$
	$\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(p-1)\overline{a}h+v]}{\lambda p(\overline{a}-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \min\left\{\frac{\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\right\}\right\}$ *or*	
	$\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[(1-p)(1-\overline{a})h+v]}{\lambda p\overline{a}} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \max\left\{\frac{\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)},\ \frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}\right\}\right\}$	

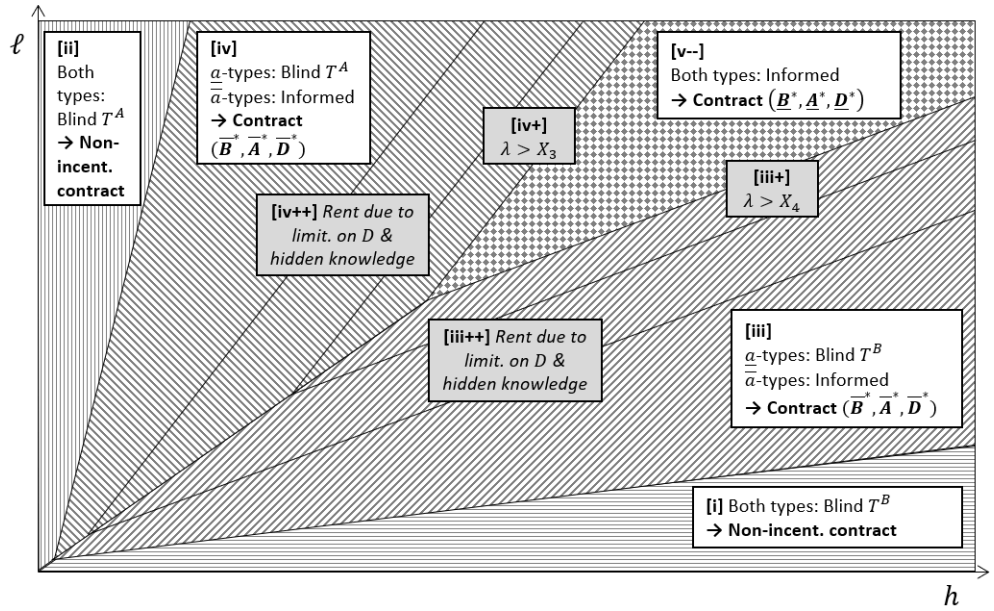
Note: This table presents the second-best optimality conditions for the implementation of the incentivizing contracts given physicians differ in their diagnostic ability, which is operationalized by means of heterogeneities in diagnostic accuracy $a \in \{\underline{a}, \overline{a}\}$. There are high-ability physicians ($\overline{a}$-types) and low-ability physicians ($\underline{a}$-types) in the market, whereby λ stands for the share of the $\overline{a}$-types. v denotes the physicians' utility cost for exerting diagnostic effort and ω denotes their reservation utility. p is the probability that patients suffer from the severe illness. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. The values for X_i with $i \in \{1, 2, 3, 4, 5, 6\}$ are presented in Lemma 6.

Figure 5-3: The Second-Best Allocation given (I) ω is Relatively Large and $a \in \{\underline{a}, \overline{a}\}$



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes relatively large values $\left(\omega \geq v/p\left(2\underline{a} - 1\right)\right)$, which is in case both ability-types' PCs bind. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{a}$-types) and low-ability physicians ($\underline{a}$-types). λ refers to the share of the $\overline{a}$-types in the physician collective. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 v=2, $\underline{a}$=0.65, $\overline{a}$=0.85, λ=1/3 and ω=25.

Figure 5-4: The Second-Best Allocation given (II) ω is Intermediate and $a \in \{\underline{a}, \overline{a}\}$



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes intermediate values $\left(v/p\left(2\underline{a} - 1\right) > \omega \geq v/p(2\overline{a} - 1)\right)$, which is in case only the $\overline{a}$-types' PC binds. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{a}$-types) and low-ability physicians ($\underline{a}$-types). λ refers to the share of the $\overline{a}$-types in the physician collective. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 v=2, $\underline{a}$=0.65, $\overline{a}$=0.85, λ=1/3 and ω=12.5.

Figure 5-3 shows the second-best allocation for relatively large values of ω. Because both types' PCs bind, both incentive contracts may be implemented without paying any information rents accruing from the limitation on capitation payment D. Consequently, there is no trade-off between the incentives provided by the non-incentivizing contract and by incentive contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Area [i] and area [ii] remain unaffected (compare Figure 5-2). When incentive contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is implemented, however, the $\overline{a}$-types always earn an excess rent accruing from to the hidden knowledge about the type. This may make alternative contracts superior. Particularly, contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ dominates contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ if and only if the treatment related cost (ℓ, h) reach a certain level and λ does not exceed a certain threshold. For relatively large values of ω, the necessary conditions for ℓ (and h) coincide with the first-best condition for the $\underline{a}$-types' informed treatment choice. This simplifies the graphical analysis.[10] If λ exceeded the threshold, the excess rent payments to the increasing share of $\overline{a}$-types under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ would be higher than the cost resulting from the $\underline{a}$-types' effort distortion under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. In this exemplary figure, λ exceeds the threshold for parameter constellations (ℓ, h) in area [iii+] and area [iv+] as $\lambda > X_2$ and $\lambda > X_1$, respectively. Consequently, contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ becomes second-best optimal for parameter constellations for which contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ used to be first-best and implies an effort distortion for the $\underline{a}$-types. Generally, the threshold for λ increases in the treatment related cost (ℓ, h) and in the balance of parameter pairs (ℓ, h), which is the closer they are to the relation $\ell = (1-p)h/p$. If the stakes are rather high and uncertain, the effort distortion cost, which increase in h and ℓ, clearly outweigh the excess rent payments to the $\overline{a}$-types. Summing up, in comparison to the first-best allocation, the areas, in which only the $\overline{a}$-types perform an informed choice (area [iii] and area [iv], respectively) expand at the expense of area [v-], in which it is optimal that both types perform an informed treatment choice.

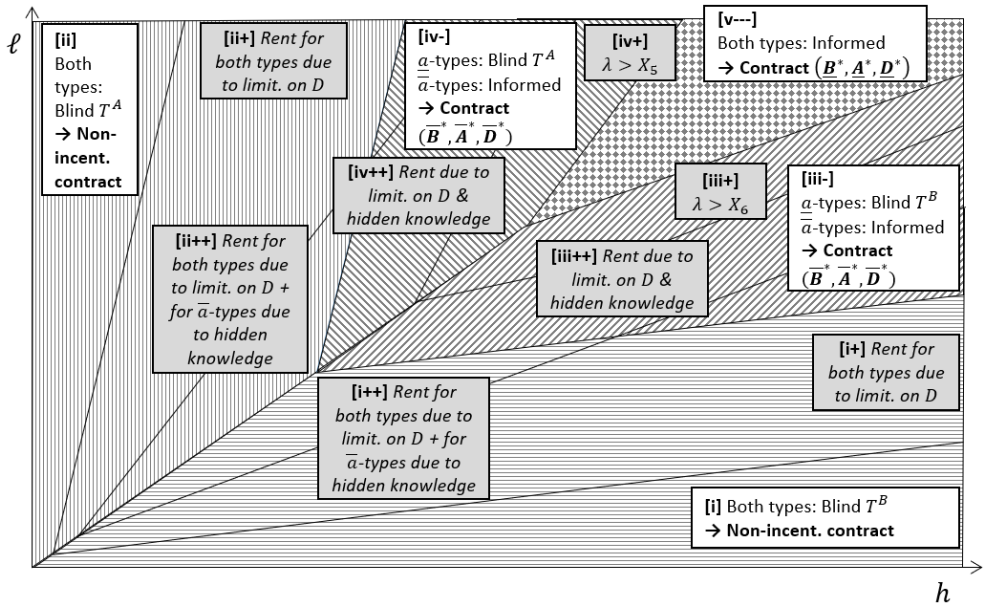
Figure 5-4 depicts the second-best allocation for intermediate values of ω. As the $\overline{a}$-types' PC still binds, neither type earns a rent from the limitation on $\overline{D}^*$ when contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ is implemented. Area [i] and area [ii] remain unaffected. When contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is implemented though, both types earn the information rent from the limitation on $\underline{D}^*$ and the $\overline{a}$-types additionally gain the excess rent. For moderate values of ℓ and h (see area [iii++] and area [iv++]), the sum of the rent payments under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ always exceeds the effort

[10] The values for ℓ (and h) as well as the thresholds for λ are defined by Proposition 1.

distortion cost under the alternative contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$). With a lower ω, providing incentives with contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ becomes increasingly costly to the payer. Hence, contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$) becomes second-best optimal for parameter constellations, for which contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ was found to be first-best. Only for relatively high values of h and ℓ, the necessary conditions for ℓ (and h), under which contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ may dominate contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$), are met (see area [iii+], area [iv+] and area [v--]). Further, λ needs to be below a certain threshold. Only if both conditions are met, incentivizing both types is worthwhile. In area [iii+] and area [iv+], λ exceeds the threshold ($\lambda > X_4$ and $\lambda > X_3$, respectively). Therefore, the $\underline{a}$-types' effort is distorted in the second-best situation. Summing up, compared to the previous case for relatively large values of ω, the areas, in which only the $\overline{a}$-types perform an informed choice (area [iii] and area [iv]), even further expand at the expense of area [v--], in which it is optimal that both types perform an informed treatment choice.

Finally, Figure 5-5 presents the second-best allocation for relatively small values of ω. Because neither types' PC binds, the implementation of either incentive contract requires paying an information rent resulting from the limitation on capitation payment D. When contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is implemented, the $\overline{a}$-types further earn an excess rent accruing from to the hidden knowledge about their type. Thus, compared to the first-best allocation, area [i] expands by area [i+] and area [i++] at the expense of area [iii] and area [v], respectively. Area [ii] expands by area [ii+] and area [ii++] at the expense of area [iv] and area [v], respectively. This implies that, in the second-best situation, both types' effort is distorted. For moderate values of h and ℓ (see area [iii++] and area [iv++]), again, the sum of the rent payments under contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ always exceeds the rent payments and the effort distortion cost under the alternative contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$). Therefore, contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$) becomes second-best optimal for parameter constellations, for which contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ was found to be first-best. Finally, only for relatively high values of ℓ and h, contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ may potentially be second-best optimal. While in area [iii+] and area [iv+] λ exceeds the threshold ($\lambda > X_6$ and $\lambda > X_5$, respectively), such that contract ($\overline{B}^*, \overline{A}^*, \overline{D}^*$) becomes second-best optimal and the $\underline{a}$-types' effort is further distorted, only in area [v---] incentivizing both types is also second-best optimal. Summing up, due to the relatively small value of ω and the associated increased information rent payments to both types of physicians, the provision of a non-incentivizing contract becomes relatively more attractive. While the areas, in which neither type performs an informed treatment choice expand, the areas in which at least the $\overline{a}$-types perform an informed treatment choice shrink.

Figure 5-5: The Second-Best Allocation given (III) ω is Relatively Small and $a \in \{\underline{a}, \overline{a}\}$



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes relatively small values ($v/p(2\overline{a} - 1) > \omega$), which is in case neither ability-types' PC binds. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{a}$-types) and low-ability physicians ($\underline{a}$-types). λ refers to the share of the $\overline{a}$-types in the physician collective. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 v=2, $\underline{a}$=0.65, $\overline{a}$=0.85, λ=1/3 and ω=0.

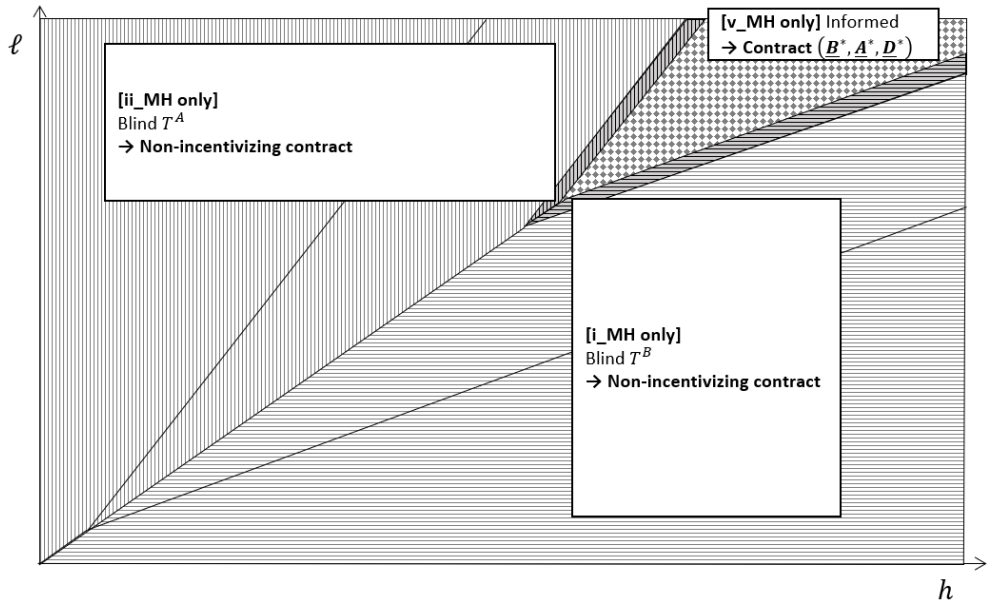
5.5.3 Distinction of the Effects from Moral Hazard and Adverse Selection

Figure 5-3 to Figure 5-5 depict the second-best allocation when accounting for both, double moral hazard and adverse selection. Yet, the illustrations do not allow for a distinction between the effects from either kind of information asymmetry on the $\underline{a}$-types' effort distortion. In this subsection, I therefore consider a benchmark situation with moral hazard only to differentiate between the two effects. Assuming the physician collective consists of $\underline{a}$-types only and the payer can neither observe the physicians' diagnostic effort decision, nor the resulting signal, contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ provides optimal incentives for an informed treatment choice (follows from Lemma 2 and Lemma 3). Lemma 7 defines the conditions, under which the implementation of contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is second-best optimal. The proof is in Appendix E.

Lemma 7:*Given there are only low-ability types in the physician collective, offering incentive contract* $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ *is second-best optimal if and only if*

$$\frac{v+\underline{a}(p-1)h+\max\left\{\frac{v}{p(2\underline{a}-1)}-w,0\right\}}{-p(1-\underline{a})} \geq \ell \geq \frac{v+(1-p)(1-\underline{a})h+\max\left\{\frac{v}{p(2\underline{a}-1)}-w,0\right\}}{p\underline{a}}$$

Figure 5-6: The Second-Best Allocation given Moral Hazard Only and ω is Relatively Small



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes relatively small values ($v/p(2\overline{a} - 1) > \omega$), which is in case neither ability-types' PC binds. There are only low-ability physicians ($\underline{a}$-types) in the physician collective, referring to a situation with moral hazard only. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 v=2, $\underline{a}$=0.65 and ω=0.

Figure 5-6 depicts the second-best allocation given there are only $\underline{a}$-types in the physician collective and ω is relatively small. As small values for ω represent the most interesting case, in this sub-analysis, I will solely focus on said parameter range. For reasons of comparability, Figure 5-6 is based on the same set of parameters as Figure 5-5. Generally, the $\underline{a}$-types' effort is distorted in the second-best case (compare the first-best allocation in Figure 5-2). For parameter constellations in area [i_MH only] (area [ii_MH only]), comprising the horizontally (vertically) striped subordinate areas, the physician should forgo diagnostic effort and treat all patients with T^B first (T^A right away). Only in area [v_MH only] it is second-best optimal for the $\underline{a}$-types to perform an informed treatment choice. Comparing Figure 5-6 to Figure 5-5 reveals that, given moral hazard only, the $\underline{a}$-types' effort is even further distorted than in the situation with moral hazard and adverse selection: For parameter constellations in the striped, grey shaded areas, the $\underline{a}$-types' effort is distorted in a situation with moral hazard only but not in a situation with additional adverse selection.

In this model, adding adverse selection to the moral hazard problem may have controversial effects on the $\underline{a}$-types' effort distortion: When additionally accounting for adverse selection,

the next best alternative to incentivizing the $\underline{a}$-types (and the $\overline{a}$-types) by implementing contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is to only incentivize the $\overline{a}$-types with contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. (By contrast, given moral hazard only, the alternative to incentivizing the $\underline{a}$-types with contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is to provide a non-incentivizing contract.) In case λ is below a threshold, the effort distortion cost caused by the $\underline{a}$-types and the rent payments (to both types) under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ already outweigh the sum of the rent payments from contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ for parameter constellations, for which, given moral hazard only, the effort distortion cost caused by the $\underline{a}$-types under the non-incentivizing contract do not yet outweigh the rent payments to the $\underline{a}$-types from contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$. Hence, additionally accounting for adverse selection may alleviate the $\underline{a}$-types' effort distortion resulting from pure moral hazard. In case λ exceeds the threshold though, the increasing share of the $\overline{a}$-types drives up the sum of the rent payments from contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$, while the effort distortion cost caused by the $\underline{a}$-types and the rent payments (to the $\underline{a}$-types) under contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ decrease. Hence, for parameter constellations, for which, given moral hazard only, the effort distortion cost caused by the $\underline{a}$-types under the non-incentivizing contract already outweigh the rent payments to the $\underline{a}$-types from contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$, the $\underline{a}$-types' effort remains still distorted. Adverse selection therefore reinforces the $\underline{a}$-types' effort distortion. The threshold for λ is derived in Appendix E.

5.6 Operationalizing Heterogeneous Diagnostic Ability by Means of Heterogeneous Diagnostic Efficiency

The altered model assumptions: In this subsection, heterogeneities in physicians' diagnostic ability are operationalized by means of heterogeneities in diagnostic efficiency rather than by heterogeneities in diagnostic accuracy. The rationale is that physicians differ in the amount time or information needed to reach a certain level of diagnostic precision and not in the level of diagnostic precision itself (National Academies of Sciences, Engineering, and Medicine 2015; Chatterjee et al. 2019). Analogously to the previous operationalization, exerting diagnostic effort yields a private, imperfect signal s on the patient's state of illness for both ability-types. Yet, I suppose that the informational value of the signal depends on a uniform measure of the physicians' diagnostic accuracy a with $se = sp = a$, while physicians incur heterogeneous effort cost v to reach the uniform level of diagnostic accuracy. Again, diagnostic ability is modelled binary with $v \in \{\overline{v}, \underline{v}\}$. Low-ability physicians ($\underline{v}$-types) incur cost $\underline{v}$, whereas high-ability physicians ($\overline{v}$-types) incur cost $\overline{v}$. Effort cost are strictly positive and $\overline{v}$-types incur

smaller cost and, therefore, work more efficiently than $\underline{v}$-types, implicating $0 < \overline{v} < \underline{v}$. Because the analyses follow the same pattern as before, the reader is referred to Appendix E for the details. In this section, I will solely present the main results.

The First-Best: Figure 5-7 depicts the first-best solutions. Analogously to the previous operationalization by means of diagnostic accuracy, for relatively low values of ℓ (h) and across values of h (ℓ), both types should pick a blind treatment strategy. In area [v.i], it is optimal to first give T^B a try and in area [v.ii] applying T^A right away is optimal. For moderate values of ℓ (h) and across values of h (ℓ), the $\overline{v}$-types should perform an informed treatment choice, while the $\underline{v}$-types should try T^B first (choose T^A right away). Since, in area [v.iii] and area [v.iv], the stakes are higher than in area [v.i] and area [v.ii], respectively, for the $\overline{v}$-types it is worthwhile to exert costly diagnostic effort to subsequently save on the treatment related cost (ℓ, h). The $\underline{v}$-types should continue to pick a blind treatment strategy because their higher effort cost would still exceed the savings on the treatment related cost. For relatively high values of (ℓ, h), exerting effort is worthwhile for both types (see area [v.v]).

Implementation of the First-Best and potential rent payments: For parameter constellations, for which it is optimal that both types follow a blind course (area [v.i] and area [v.ii] in Figure 5-7), again, setting $B = A = 0$ implements the first-best. Capitation payment D will be set such that the physician just earns her reservation utility ω. This allows for a costless implementation.

In case it is optimal that at least one type makes an informed treatment choice (areas [v.iii] - [v.v]), the optimal FFS mark-up payments (B, A) are derived in the same manner as for the operationalization by means of diagnostic accuracy. They are given by

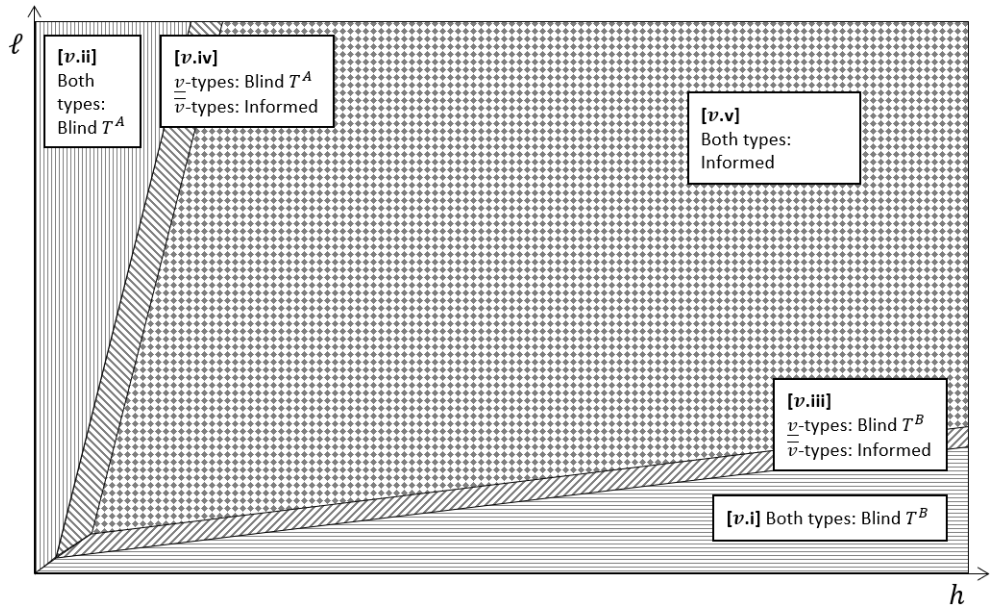
$$B^{*,v} = \frac{v}{p(1-p)(2a-1)} \quad \text{and} \quad A^{*,v} = \frac{v}{p(2a-1)}.$$

Both mark-ups decrease in ability because more able physicians incur smaller effort cost than less able physicians to generate a diagnostic signal of the exact same precision $\left(\frac{\partial B^{*,v}}{\partial v} > 0, \ \frac{\partial A^{*,v}}{\partial v} > 0\right)$. Thus, at the optimum, the mark-ups, which serve as a compensation for exerting costly effort, may be set lower for the $\overline{v}$-types than for the $\underline{v}$-types. The optimal capitation payment D is determined by inserting $B^{*,v}$ and $A^{*,v}$ into the physicians' utility function. Again, D must not be negative. Hence, the incentive-compatible D is given by

$$D^{*,v} = \max\left\{0, \omega - \frac{v}{p(2a-1)}\right\}.$$

Rent payments amount to $r^v = \frac{v}{p(2a-1)} - \omega$ and decrease in ability.

Figure 5-7: The First-Best Allocation given $v \in \{\overline{v}, \underline{v}\}$



Note: This figure depicts the first-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given that physicians differ in their diagnostic ability, which is operationalized by means of heterogeneities in diagnostic efficiency (the cost function) $v \in \{\overline{v}, \underline{v}\}$. There are high-ability physicians ($\overline{v}$-types) and low-ability physicians ($\underline{v}$-types). h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. T^B and T^A stand for basic and advanced treatment, respectively. The figure is exemplary for values of p=0.3 a=0.85, $\underline{v}$=6 and $\overline{v}$=2.

Implementation with a contract consisting of a single policy: The payment instruments cannot depend on effort cost v because the ability-type is hidden knowledge. Therefore, supposing the payer may only offer a single policy contract and, following the same reasoning as in section 5.4.2.1, I focus on two contracts. The contracts are defined by Lemma 8:

Lemma 8: *When only considering single policy contracts and operationalizing heterogeneous diagnostic ability by means of heterogeneous diagnostic efficiency, the payer may offer contract* $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ *or contract* $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$*, which serve the* $\overline{v}$*-types and the* $\underline{v}$*-types, respectively:*

- *Contract* $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ *consists of payments*
  $\overline{B}^{*,v} = \frac{\overline{v}}{p(1-p)(2a-1)}$, $\overline{A}^{*,v} = \frac{\overline{v}}{p(2a-1)}$., $\overline{D}^{*,v} = \max\left\{0, \omega - \frac{\overline{v}}{p(2a-1)}\right\}$
- *Contract* $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$ *consists of payments*
  $\underline{B}^{*,v} = \frac{\underline{v}}{p(1-p)(2a-1)}$, $\underline{A}^{*,v} = \frac{\underline{v}}{p(2a-1)}$, $\underline{D}^{*,v} = \max\left\{0, \omega - \frac{\underline{v}}{p(2a-1)}\right\}$

Whereby, $\underline{B}^{*,v} > \overline{B}^{*,v}$, $\underline{A}^{*,v} > \overline{A}^{*,v}$ *and* $\underline{D}^{*,v} \leq \overline{D}^{*,v}$.

Inspecting the incentive effects from the single policy contracts reveals that contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ is not incentive-compatible for the $\underline{v}$-types and, therefore, leads to a separating equilibrium. The contract implements the first-best solutions in area [v.iii] and area [v.iv] in Figure 5-7. Moreover, in case the $\overline{v}$-types' PC does not bind, both types earn a strictly positive information rent that accrues from the limitation on $\overline{D}^{*,v}$. Rent payments amount to $r^{v}_{(\overline{B}^{*,v},\overline{A}^{*,v},\overline{D}^{*,v})} = \frac{\overline{v}}{p(2a-1)} - \omega$.

Offering contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is incentive-compatible for both types, but always comes at a cost. The $\overline{v}$-types earn a strictly positive excess information rent $(\underline{v} - \overline{v})$, which results from the hidden knowledge about the type, and, in case the $\underline{v}$-types' PC does not bind, both types earn a rent due to the limitation on capitation $\underline{D}^{*,v}$. The contract implements the First-Best in area [v.v] and rent payments sum up to $r^{v}_{(\underline{B}^{*,v},\underline{A}^{*,v},\underline{D}^{*,v})} = \lambda\left(\underline{v} - \overline{v}\right) + \frac{\underline{v}}{p(2a-1)} - \omega$.

Implementation with a contract consisting of a menu of policies: The single policy contracts in Lemma 8 have the same structure and create similar incentive effects as the contracts for the operationalization by means of diagnostic accuracy (compare Lemma 3). Hence, the findings from Lemma 5 may also be applied to this operationalization: As the $\underline{v}$-types may still earn a strictly positive information rent when their effort is distorted and because effort is assumed to be binary, a menu of policies does not improve on the outcome compared to single policy contracts $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ and $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$.

The Second-Best: The second-best solutions are determined analogously to the previous case. Though, compared to the operationalization by means of diagnostic accuracy, the derivation is less complex. In the $h - \ell$ space, the optimality constraints differ in the intercept, which is determined by effort cost v (efficiency), rather than in the slope, which is determined by diagnostic precision a (accuracy). The second-best solutions are given by Proposition 2.

Proposition 2: *Given heterogeneities in physicians' diagnostic ability are operationalized by means of heterogeneities in diagnostic efficiency, in the second-best situation, offering a non-incentivizing contract with FFS mark-up payments $B = A = 0$ and capitation payment $D = \omega$ is optimal only if and only if*

$$\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)}\right\} \text{ or}$$

$$\ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa}\right\}.$$

The conditions, under which offering an incentive contract that either serves the high or the low-ability types is second-best optimal, particularly depend on the share of the high-ability types, λ, in the physician collective. The contracts are outlined in Lemma 8 and the optimality conditions are presented in Table 5-2.

*Proof. S*ee Appendix F.

The second-best allocations and contracts for the three value ranges of the physicians' reservation utility ω are presented in Figure 5-8 to Figure 5-10. To the extent possible, they are based on the same set of parameters that were chosen for the preceding figures.

Figure 5-8 shows the second-best allocation given relatively large values of ω and $v \in \{\underline{v}, \overline{v}\}$. Because the $\overline{v}$-types' (and the $\underline{v}$-types') PC binds, implementing incentive contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ is costless to the payer and area [v.i] and area [v.ii] remain unaffected. Implementing incentive contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$, however, is costly as the $\overline{v}$-types earn the excess rent accruing from the hidden knowledge about their type. Particularly, if λ exceeds a certain threshold, the excess rent payments to the increasing share of $\overline{v}$-types are higher than the effort distortion cost caused by the $\underline{v}$-types under contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$.[11] This makes contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ second-best optimal for parameter constellations for which contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$ used to be first-best and implies an effort distortion for the $\underline{v}$-types. Thus, in comparison to the first-best allocation (compare Figure 5-7), area [v.iii] and area [v.iv] expand by area [v.iii+] and area [v.iv+], respectively, at the expense of area [v.v-].

Figure 5-9 depicts the second-best allocation for intermediate values of ω and given $v \in \{\underline{v}, \overline{v}\}$. Only the $\overline{v}$-types' PC binds. Therefore, when contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ is implemented, no rents are earned from the limitation on $\overline{D}^{*,v}$ and area [v.i] and area [v.ii] remain unaffected. Yet, when contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$ is implemented, both types earn a rent from the limitation on $\underline{D}^{*,v}$ and the $\overline{v}$-types additionally earn the excess rent.

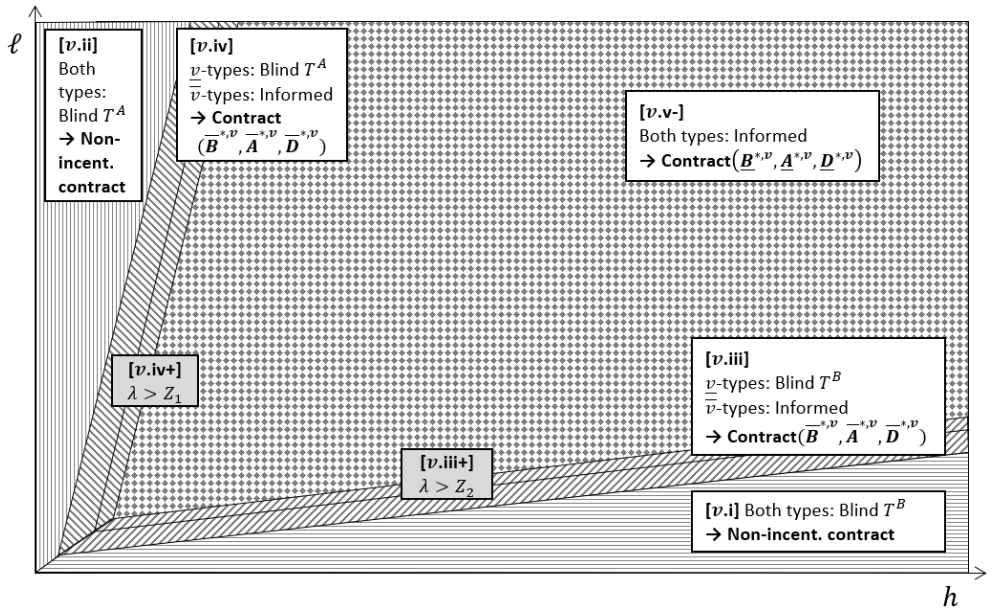
[11] If ω takes relatively large or intermediate values, the necessary conditions for ℓ (and h), under which contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$ dominates contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$, coincide with the second-best condition for the implementation of contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$. This simplifies the graphical analysis. The conditions for ℓ (and h) as well as the thresholds for λ are defined by Proposition 2.

Table 5-2: The Second-Best Optimality Conditions for the Implementation of the Incentivizing Contracts given $v \in \{\underline{v}, \overline{v}\}$

Value range for ω	***Optimality conditions for the implementation of contract $(\overline{\boldsymbol{B}}^{*,v}, \overline{\boldsymbol{A}}^{*,v}, \overline{\boldsymbol{D}}^{*,v})$, which is designed for the $\overline{v}$-types***	***Optimality conditions for the implementation of contract $(\underline{\boldsymbol{B}}^{*,v}, \underline{\boldsymbol{A}}^{*,v}, \underline{\boldsymbol{D}}^{*,v})$, which is designed for the $\underline{v}$-types***
(I) The physicians' reservation utility ω takes relatively large values: $\omega \geq \frac{\underline{v}}{p(2a-1)}$	$\frac{\underline{v}+(p-1)ah}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(p-1)ah+p(1-a)\ell}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(1-p)(1-a)h-pa\ell}{\overline{v}+(1-p)(1-a)h-pa\ell}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(p-1)ah]}{\lambda p(a-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(p-1)ah}{p(a-1)}\right\}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(1-p)(1-a)h]}{\lambda pa} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(1-p)(1-a)h}{pa}\right\}$	$\frac{\underline{v}+(p-1)ah}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell}{\overline{v}+(1-p)(1-a)h-pa\ell}$
(II) The physicians' reservation utility ω takes intermediate values: $\frac{\underline{v}}{p(2a-1)} > \omega \geq \frac{\overline{v}}{p(2a-1)}$	$\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(1-p)(1-a)h-pa\ell}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(p-1)ah]}{\lambda p(a-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)}\right\}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(1-p)(1-a)h]}{\lambda pa} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa}\right\}$	$\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(1-p)(1-a)h-pa\ell}$
(III) The physicians' reservation utility ω takes relatively small values: $\frac{\overline{v}}{p(2a-1)} > \omega$	$\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda > \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(1-p)(1-a)h-pa\ell}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(p-1)ah]}{\lambda p(a-1)} \geq \ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)}\right\}$ *or* $\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(1-p)(1-a)h]}{\lambda pa} \leq \ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa}\right\}$	$\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} > \ell \geq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(p-1)ah+p(1-a)\ell}$ *or* $\frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa} < \ell \leq \frac{(1-p)h}{p}$ *and* $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(1-p)(1-a)h-pa\ell}$

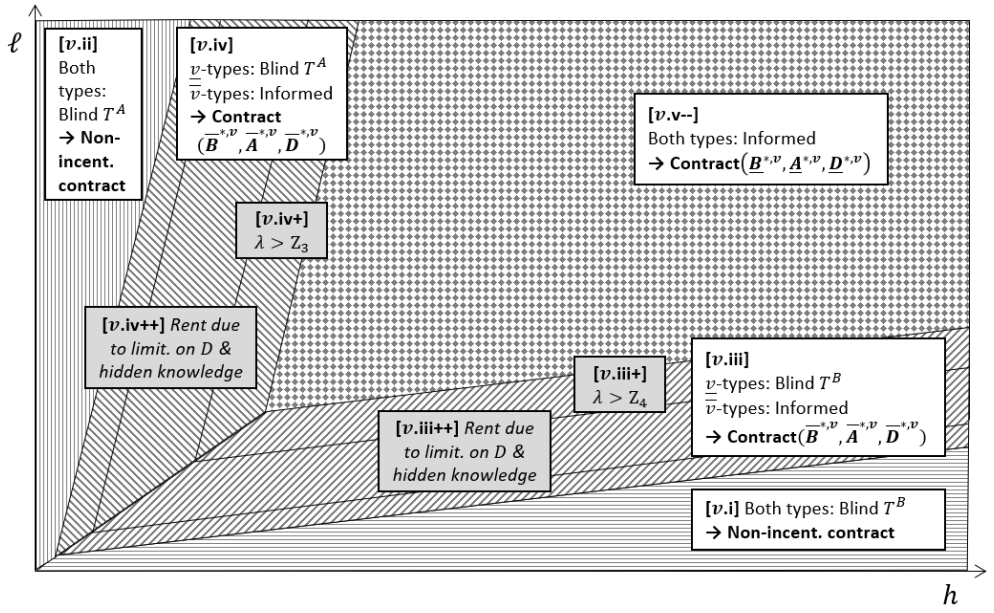
Note: This table presents the second-best optimality conditions for the implementation of the incentivizing contracts given physicians differ in their diagnostic ability, which is operationalized by means of heterogeneities in diagnostic efficiency $v \in \{\underline{v}, \overline{v}\}$. There are high-ability physicians ($\overline{v}$-types) and low-ability physicians ($\underline{v}$-types) in the market, whereby λ stands for the share of the $\overline{v}$-types. a denotes the physicians' diagnostic accuracy and ω denotes their reservation utility. p is the probability that patients suffer from the severe illness. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment.

Figure 5-8: The Second-Best Allocation given (I) ω is Relatively Large and $v \in \{\underline{v}, \overline{v}\}$



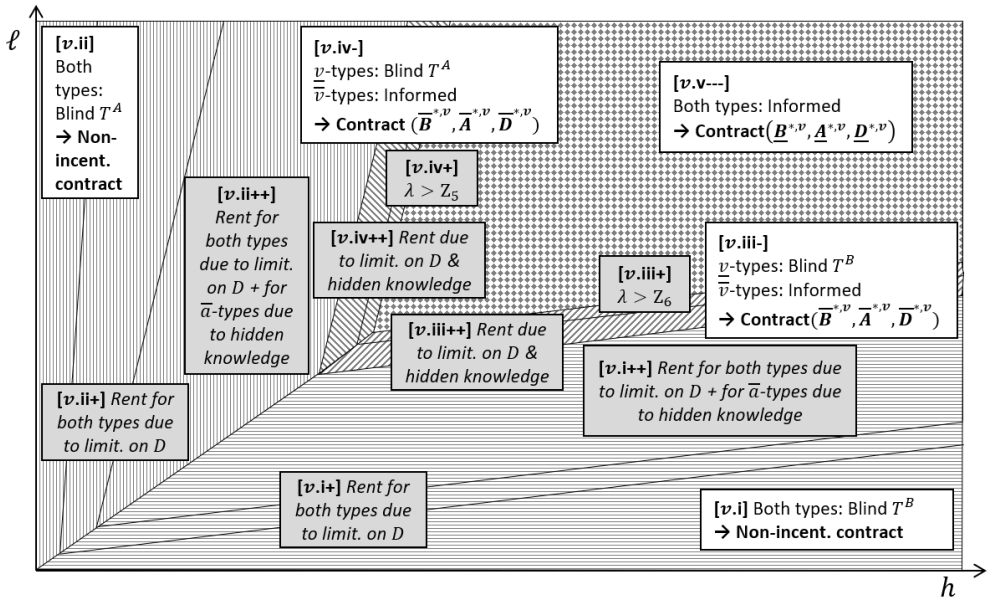
Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes relatively large values $\left(\omega \geq \underline{v}/p(2a-1)\right)$, which is in case both ability-types' PCs bind. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{v}$-types) and low-ability physicians ($\underline{v}$-types). λ refers to the share of the $\overline{v}$-types in the physician collective. T^B and T^A stand for basic treatment and advanced treatment, respectively. The figure is exemplary for values of p=0.3 a=0.85, $\underline{v}$=6 and $\overline{v}$=2, λ=1/3 and ω=35.

Figure 5-9: The Second-Best Allocation given (II) ω is Intermediate and $v \in \{\underline{v}, \overline{v}\}$



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes intermediate values $\left(\underline{v}/p(2a-1) > \omega \geq \overline{v}/p(2a-1)\right)$, which is in case only the $\overline{v}$-types' PC binds. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{v}$-types) and low-ability physicians ($\underline{v}$-types). λ refers to the share of the $\overline{v}$-types in the physician collective. T^B and T^A stand for basic treatment and advanced treatment, respectively. The figure is exemplary for values of p=0.3 a=0.85, $\underline{v}$=6 and $\overline{v}$=2, λ=1/3 and ω=17.5.

Figure 5-10: The Second-Best Allocation given (III) ω is Relatively Small and $v \in \{\underline{v}, \overline{v}\}$



Note: This figure depicts the second-best allocation of diagnosis and treatment courses in the $h - \ell$ space, given the physicians' reservation utility ω takes relatively small values ($\overline{v}/p(2a-1) > \omega$), which is in case neither types' PC binds. h stands for the additional cost for advanced compared to basic treatment. ℓ denotes the monetary equivalent of the patient's expected health loss from delayed advanced treatment. There are high-ability physicians ($\overline{v}$-types) and low-ability physicians ($\underline{v}$-types). λ refers to the share of the $\overline{v}$-types in the physician collective. T^B and T^A stand for basic treatment and advanced treatment, respectively. The figure is exemplary for values of p=0.3 a=0.85, $\underline{v}$=6 and $\overline{v}$=2, λ=1/3 and ω=0.

Hence, similar to the operationalization by means of diagnostic accuracy, for moderate values of ℓ and h, contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ always becomes second-best optimal (see area [v.iii++] and area [v.iv++]). For relatively high values of ℓ and h, it, again, depends on the size of λ. In this figure, λ exceeds the threshold in area [v.iii+] and area [v.iv+]) and the $\underline{v}$-types' effort is further distorted in the second-best situation. In conclusion, area [v.iii] and area [v.iv] expand by areas [v.iii+, v.iii++] and areas [v.iv+, v.iv++], respectively, at the expense of area [v.v--].

Lastly, Figure 5-10 presents the second-best allocation for relatively small values of ω and given $v \in \{\underline{v}, \overline{v}\}$. As neither types' PC binds, implementing either incentive contract becomes increasingly costly to the payer. Therefore, area [v.i] and area [v.ii] expand by areas [v.i+, v.ii+] and areas [v.i++, v.ii++], respectively, at the expense of area [v.iii-], area [v.iv-] and area [v.v---]. Both types' effort is distorted. As before, for moderate values of ℓ and h, the sum of the rent payments under contract $(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v})$ always exceeds the sum of the rent payments and the effort distortion cost under contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$. Hence, in area [$v$.iii++] and area [$v$.iv++], contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ becomes second-best optimal. For high values of ℓ and h,

again, it depends on the size of λ. In area [v.iii+] and area [v.iv+], λ exceeds the threshold and contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ becomes second-best optimal for parameter constellation, for which contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ used to be first-best. Therefore, only in area [v.v---], incentivizing both types is still worthwhile.

Finally, comparing both operationalization approaches yields that the results from the operationalization by means of diagnostic efficiency (the cost function) resemble the results from the operationalization by means of diagnostic accuracy (the production function). In both cases, the implementation of the first-best allocation is never second-best optimal. This finding may be explained by two effects: Firstly, when implementing the First-Best, the high-types always earn a strictly positive information rent by pretending to be low-types. This leaves the incentivization of both types suboptimal for certain parameter constellations. Particularly, offering a contract that incentivizes both types to perform an informed treatment choice and, therefore, allowing the high-types to mimic the low-types, is second-best optimal, only if the share of the high-types λ is not too big, treatment related cost (ℓ, h) are neither too small nor too unbalanced, and the heterogeneities in physician ability are not too great. Secondly, depending on the value of the physicians' reservation utility ω, the payer may further have to grant a rent, which accrues from the limitation on capitation payment D. The physicians' earnings from FFS mark-up payments (A, B) are higher under the contract that incentivizes both types than under the contract that incentivizes the high-types only. Consequently, with decreasing ω, providing incentives to both or to the high-types only becomes successively suboptimal.

5.7 Conclusion

When purchasing diagnosis and treatment services from multiple physicians, insurers need to account for heterogeneities in physicians' diagnostic ability and for potential agency problems. Both conditions may entail costly rent payments to the physicians and, therefore, have a special influence on the provision of optimal incentives. To the best of my knowledge, this is the first paper that examines how a public insurer should optimally create incentives for physicians' diagnostic and treatment choice given hidden action and hidden information (double moral hazard) with heterogeneous diagnostic ability (adverse selection). The paper shows that, in the second-best situation, mixed payment contracts provide optimal incentives. The contracts comprise a non-negative capitation and two strictly positive fee-for-service (FFS) payments, which are paid conditional on adequate treatment choice. All three payment instruments depend

on diagnostic ability. Given the payer offers a single policy contract, to all physicians, he may either offer a contract that is designed for the low-types or a contract that is designed for the high-types. A distinguished contract offer is infeasible as the high-types would always have an incentive to mimic the low-types. This leads to a rent-efficiency trade-off and the dominating contract is particularly determined by the share of the high-types in the physician collective. Moreover, distinguishing between the effects from moral hazard and adverse selection reveals that accounting for adverse selection in addition to moral hazard may have controversial effects on the extent of the low-types' effort distortion. While all first-best solutions may also be second-best optimal, the parameter constellations, for which the provision of incentives for an informed treatment choice is second-best optimal, are restricted compared to the First-Best. A contract consisting of a menu of policies that aims at inducing self-selection among the physicians does not improve on this outcome and the results are robust to different operationalizations of heterogeneities in ability.

The paper is based on a treatment choice model with endogenous diagnostic effort and relies on several assumptions. Hence, the findings of the paper should be seen in light of these assumptions and the structure of the derived payment contract. Firstly, diagnostic ability is modelled binary. While other authors (i.a. Allard et al. 2014) assume a continuum of types, I limit the model to two types as the main insights should be generalizable to a situation with more than two types. Secondly, I assume that the physicians' reservation utility is type-independent. Yet, it could well be argued that high-ability physicians have superior alternative income options to low-ability physicians (Jullien 2000). Depending on the original level of the low-types' reservation utility and the spread between the two types' differing reservation utilities, I conjecture that this model extension would intensify the incentive problem because it might require increased rent payments to the high-types only or even to both types. Thirdly, although the assumption of altruistic physicians and its importance for payment scheme design has been widely acknowledged in the literature (for instance, Jack 2005; Allard et al. 2014), I make the simplifying assumption that physicians are only concerned with their own income. Alternatively, supposing physicians also accounted for the patients' well-being, for instance, by internalizing the health loss from delayed advanced treatment, would not alter the fundamental results. The FFS reward payments, which are conditional on adequate treatment choice, could be lowered and thereupon, depending on the level of the physicians' reservation utility, I hypothesize that the incentive problem would be alleviated. A further limitation is that the model solely focuses on physicians as decision-makers in the care process. Even though physicians essentially determine cost and outcomes with their decisions, patient-centered care

is increasing in significance. That is, patients may ever more be regarded as active counterparts in the care process by not only demanding certain services but also by actively participating in it (see, for instance, Little et al. 2004). Hence, activating the fully-insured patient might reinforce the physicians' incentives to forgo diagnostic effort and to provide the advanced treatment right away. Advanced treatment always cures the patient without side-effects and is, therefore, preferred by the patients. This would make the provision of incentives even more costly.

Although the model has limitations, I surmise that the main findings of the paper remain robust: Given the insurer offers a single policy contract, it may turn out suboptimal to incentivize all physicians to exert costly diagnostic effort in the second-best situation. While the National Academies of Sciences, Engineering, and Medicine (2015) and other medical expert associations campaign for the special importance of a correct diagnosis to effective health care, considering the social cost of care provision may paint a more differentiated picture. In consequence of heterogeneities in diagnostic ability and agency problems, which are found to entail costly rent payments to the physicians, the circumstances under which all physicians should be incentivized to perform a diagnosis are restricted in the second-best situation. The results are found to particularly depend on the share of the high-ability physicians in the physician collective. Incentivizing both ability-types to exert diagnostic effort by offering a contract that is designed for the low-ability types and allowing the high-types to mimic the low-types, is only second-best optimal if the share of the high-types is not too large. By contrast, if the share of the high-types exceeds a certain threshold, offering a contract that serves the high-types and accepting the low-types' effort distortion is the dominating strategy. In this respect, the findings are congruent with the results from Alger and Ma (2003). Apart from that, there may also be circumstances, in which it is second-best optimal to forgo incentivizing any physician at all because the provision of incentives turns out too costly. Moreover, the paper reinforces the general findings from Gottlieb and Moreira (2014) and shows that a limitation to single policy contracts may be efficient. By contrast to classical principal-agent theory, I find that for diagnostic and treatment choice problems with simultaneous moral hazard and adverse selection, a menu of policies does not improve on the outcome implemented by single policy contracts.

This insight reasserts the observation that there are only few health care systems, which directly offer a menu of policies and allow outpatient physicians to self-select themselves into a payment policy. Across provinces in Canada, for instance, primary care physicians can choose between FFS and alternative forms of remuneration schemes (Institute of Health Economics 2009;

Rudoler et al. 2015). Besides, implementing these menus in practice is challenging and may impose extra administrative cost on the payer. Finally, the results are robust to different operationalizations of heterogeneities in ability and are therefore generalizable to a broad range of applications.

References

Alger, Ingela; Ma, Ching-to Albert (2003). Moral hazard, insurance, and some collusion. Journal of Economic Behavior & Organization 50 (2), pp. 225–247. DOI: 10.1016/S0167-2681(02)00049-5.

Allard, Marie; Jelovac, Izabela; Léger, Pierre Thomas (2011). Treatment and referral decisions under different physician payment mechanisms. Journal of Health Economics 30 (5), pp. 880–893. DOI: 10.1016/j.jhealeco.2011.05.016.

Allard, Marie; Jelovac, Izabela; Léger, Pierre-Thomas (2014). Payment mechanism and GP self-selection: capitation versus fee for service. International Journal of Health Care Finance and Economics 14 (2), pp. 143–160. DOI: 10.1007/s10754-014-9143-z.

Beenk, Katharina; Kifmann, Mathias (2021). Optimal Financial Incentives for Physician's Sequential Diagnostic Testing and Treatment Choice. Working Paper.

Bolton, Patrick; Dewatripont, Mathias (2005). Contract theory. Cambridge, Mass.: MIT Press.

Chan, David; Gentzkow, Matthew; Yu, Chuan (2019). Selection with Variation in Diagnostic Skill: Evidence from Radiologists. NBER Working Paper Series. Working Paper No. 26467. DOI: 10.3386/w26467.

Chatterjee, Souvik; Desai, Sanjay; Manesh, Reza; Sun, Junfeng; Nundy, Shantanu; Wright, Scott M. (2019). Assessment of a Simulated Case-Based Measurement of Physician Diagnostic Performance. JAMA Network Open 2 (1), e187006. DOI: 10.1001/jamanetworkopen.2018.7006.

Doyle, Joseph J.; Ewer, Steven M.; Wagner, Todd H. (2010). Returns to physician human capital: evidence from patients randomized to physician teams. Journal of Health Economics 29 (6), pp. 866–882. DOI: 10.1016/j.jhealeco.2010.08.004.

Dulleck, Uwe; Kerschbamer, Rudolf (2006). On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. Journal of Economic Literature 44 (1), pp. 5–42. DOI: 10.1257/002205106776162717.

Garcia Mariñoso, Begoña; Jelovac, Izabela (2003). GPs' payment contracts and their referral practice. Journal of Health Economics 22 (4), pp. 617–635. DOI: 10.1016/S0167-6296(03)00008-0.

Gottlieb, Daniel; Moreira, Humberto (2014). Simultaneous Adverse Selection and Moral Hazard. Working Paper. Available online at https://epge.fgv.br/users/humberto/wp-content/uploads/2015/07/ASMH_09072014.20.pdf, checked on 1/17/2021.

Gottschalk, Felix; Mimra, Wanda; Waibel, Christian (2020). Health Services as Credence Goods: a Field Experiment. The Economic Journal 130 (629), pp. 1346–1383. DOI: 10.1093/ej/ueaa024.

Institute of Health Economics (2009). Financial Incentives to Physician Practices. A literature review of evaluations of physician remuneration models. Available online at https://www.ihe.ca/download/provider_financing_forum_literature_review.pdf, checked on 1/17/2021.

Jack, William (2005). Purchasing health care services from providers with unknown altruism. Journal of Health Economics 24 (1), pp. 73–93. DOI: 10.1016/j.jhealeco.2004.06.001.

Jelovac, Izabela (2001). Physicians' payment contracts, treatment decisions and diagnosis accuracy. Health economics 10 (1), pp. 9–25. DOI: 10.1111/poms.12780.

Jullien, Bruno (2000). Participation Constraints in Adverse Selection Models. Journal of Economic Theory 93 (1), pp. 1–47. DOI: 10.1006/jeth.1999.2641.

Laffont, Jean-Jacques; Martimort, David (2002). The Theory of Incentives. The Principal-Agent Model. Princeton: Princeton University Press. DOI: 10.2307/j.ctv7h0rwr.

Laffont, Jean-Jacques; Tirole, Jean (1993). A Theory of Incentives in Procurement and Regulation. 5th ed. Cambridge, Massachusetts: The MIT Press.

Little, Paul; Dorward, Martina; Warner, Greg; Stephens, Katharine; Senior, Jane; Moore, Michael (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. BMJ (Clinical research ed.) 328 (7437), p. 444. DOI: 10.1136/bmj.38013.644086.7C.

Liu, Fang; Rasch, Alexander; Schwarz, Marco A.; Waibel, Christian (2020). The Role of Diagnostic Ability in Markets for Expert Services. CESifo Working Paper No. 8704. Available online at SSRN: https://ssrn.com/abstract=3736486, checked on 12/29/2020.

Mougeot, Michel; Naegelen, Florence (2005). Hospital price regulation and expenditure cap policy. Journal of Health Economics 24 (1), pp. 55–72. DOI: 10.1016/j.jhealeco.2004.04.007.

National Academies of Sciences, Engineering, and Medicine (2015). Improving Diagnosis in Health Care. Washington (DC): National Academies Press (US). DOI: 10.17226/21794.

Norman, Geoffrey (2005). Research in clinical reasoning: past history and current trends. Medical Education 39 (4), pp. 418–427. DOI: 10.1111/j.1365-2929.2005.02127.x.

Rudoler, David; Laporte, Audrey; Barnsley, Janet; Glazier, Richard H.; Deber, Raisa B. (2015). Paying for primary care: a cross-sectional analysis of cost and morbidity distributions across primary care payment models in Ontario Canada. Social Science & Medicine 124, pp. 18–28. DOI: 10.1016/j.socscimed.2014.11.001.

Appendix A

Contracts $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ and $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ are also incentive-compatible for the ability-type they were not designed for (which is the $\underline{a}$-type and the $\overline{a}$-type, respectively), if the treatment and the effort constraints hold.

In case contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ is offered to the $\underline{a}$-types, inserting into treatment constraints (4) and (5) yields:

$$(1-p_0)\overline{B}^* \geq \overline{A}^* \quad \leftrightarrow$$

$$\frac{(1-p)\underline{a}}{p(1-\underline{a})+(1-p)\underline{a}} * \frac{v}{p(1-p)(2\overline{a}-1)} \geq \frac{v}{p(2\overline{a}-1)} \quad \leftrightarrow$$

$$\underline{a} \geq \left[p(1-\underline{a}) + (1-p)\underline{a}\right] \quad \leftrightarrow$$

$$2\underline{a} \geq 1$$

$$\overline{A}^* \geq (1-p_1)\overline{B}^* \quad \leftrightarrow$$

$$\frac{v}{p(2\overline{a}-1)} \geq \frac{(1-p)(1-\underline{a})}{p\underline{a}+(1-p)(1-\underline{a})} * \frac{v}{p(1-p)(2\overline{a}-1)} \quad \leftrightarrow$$

$$\left[p\underline{a} + (1-p)(1-\underline{a})\right] \geq (1-\underline{a}) \quad \leftrightarrow$$

$$2\underline{a} \geq 1$$

In case contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is offered to the $\overline{a}$-types, inserting into treatment constraints (4) and (5) yields:

$$(1-p_0)\underline{B}^* \geq \underline{A}^* \quad \leftrightarrow$$

$$\frac{(1-p)\overline{a}}{p(1-\overline{a})+(1-p)\overline{a}} * \frac{v}{p(1-p)(2\underline{a}-1)} \geq \frac{v}{p(2\underline{a}-1)} \quad \leftrightarrow$$

$$\overline{a} \geq [p(1-\overline{a}) + (1-p)\overline{a}] \quad \leftrightarrow$$

$$2\overline{a} \geq 1$$

$$\underline{A}^* \geq (1-p_1)\underline{B}^* \quad \leftrightarrow$$

$$\frac{v}{p(2\underline{a}-1)} \geq \frac{(1-p)(1-\overline{a})}{p\overline{a}+(1-p)(1-\overline{a})} * \frac{v}{p(1-p)(2\underline{a}-1)} \quad \leftrightarrow$$

$$[p\overline{a} + (1-p)(1-\overline{a})] \geq (1-\overline{a}) \quad \leftrightarrow$$

$$2\overline{a} \geq 1$$

As shown above, the treatment constraints hold. Since the effort constraints result from a comparison of the physicians' expected utility from course 3 with their expected utilities from courses 1 and 2, respectively, I directly verify the expected utilities.

In case contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ is offered to the $\underline{a}$-types, expected utilities amount to:

$$EU_1^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + (1-p)\,\overline{B}^* = \overline{D}^* + \frac{v}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

$$EU_2^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + \overline{A}^* = \overline{D}^* + \frac{v}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

$$EU_3^{\underline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + (1-p)\underline{a}\,\overline{B}^* + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]\overline{A}^* - v$$

$$= \overline{D}^* + \frac{v[1-2p(\overline{a}-\underline{a})]}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\} - \frac{2v(\overline{a}-\underline{a})}{(2\overline{a}-1)}$$

In case contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ is offered to the $\overline{a}$-types, expected utilities amount to:

$$EU_1^{\overline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + (1-p)\underline{B}^* = \underline{D}^* + \frac{v}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$EU_2^{\overline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + \underline{A}^* = \underline{D}^* + \frac{v}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$EU_3^{\overline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + (1-p)\overline{a}\underline{B}^* + [p\overline{a} + (1-p)(1-\overline{a})]\underline{A}^* - v$$

$$= \underline{D}^* + \frac{v[1+2p(\overline{a}-\underline{a})]}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\} + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}$$

For completeness, the $\overline{a}$-types' and the $\underline{a}$-types' expected utilities from their designated contracts $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ and $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$, respectively, are given by:

$$EU_1^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + (1-p)\,\overline{B}^* = \overline{D}^* + \frac{v}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

$$EU_2^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + \overline{A}^* = \overline{D}^* + \frac{v}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

$$EU_3^{\overline{a}}(\overline{B}^*, \overline{A}^*, \overline{D}^*) = \overline{D}^* + (1-p)\overline{a}\,\overline{B}^* + [p\overline{a} + (1-p)(1-\overline{a})]\overline{A}^* - v$$

$$= \overline{D}^* + \frac{v}{p(2\overline{a}-1)} = \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

$$EU_1^{\underline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + (1-p)\underline{B}^* = \underline{D}^* + \frac{v}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$EU_2^{\underline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + \underline{A}^* = \underline{D}^* + \frac{v}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$EU_3^{\underline{a}}\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right) = \underline{D}^* + (1-p)\underline{a}\underline{B}^* + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]\underline{A}^* - v$$

$$= \underline{D}^* + \frac{v}{p(2\underline{a}-1)} = \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

Appendix B

Proof for Lemma 5:

Independent of their type, physicians pick the policy that maximizes their expected utility, which, in turn, depends on their choice of a diagnosis and treatment course. Thus, the following revelation constraints must hold:

$$\max\left\{EU_1^{\overline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ}), EU_2^{\overline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ}), EU_3^{\overline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ})\right\} \geq \max\left\{EU_1^{\overline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ}), EU_2^{\overline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ}), EU_3^{\overline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ})\right\} \tag{13}$$

$$\max\left\{EU_1^{\underline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ}), EU_2^{\underline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ}), EU_3^{\underline{a}}(\underline{A}^{\circ},\underline{B}^{\circ},\underline{D}^{\circ})\right\} \geq \max\left\{EU_1^{\underline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ}), EU_2^{\underline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ}), EU_3^{\underline{a}}(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ})\right\} \tag{14}$$

Moreover, the physicians' participation constraint, $EU \geq \omega$, must be respected and the $\overline{a}$-types' effort must not be distorted. The latter results from the monotonicity constraint. Accordingly, from ICs (9) and (10) follows:

$$\overline{A}^{\circ} \geq \frac{(1-p)(1-\overline{a})}{p\overline{a}+(1-p)(1-\overline{a})}\overline{B}^{\circ} + \frac{v}{p\overline{a}+(1-p)(1-\overline{a})} \tag{15}$$

$$\frac{(1-p)\overline{a}}{p(1-\overline{a})+(1-p)\overline{a}}\overline{B}^{\circ} - \frac{v}{p(1-\overline{a})+(1-p)\overline{a}} \geq \overline{A}^{\circ} \tag{16}$$

The optimal menu for separating equilibria (areas [iii] and [iv] in Figure 5-2): Single policy contract $(\overline{B}^{*},\overline{A}^{*},\overline{D}^{*})$ implements the First-Best. Thus, it is taken as a starting point for policy design and as a benchmark, to verify whether a menu of policies can improve on the outcome.

- In case $\omega \geq \frac{v}{p(2\overline{a}-1)}$, it follows from Lemma 4 that, under contract $(\overline{B}^{*},\overline{A}^{*},\overline{D}^{*})$, the implementation of the First-Best is costless. Thus, improvement is not possible.
- In case $\frac{v}{p(2\overline{a}-1)} > \omega$, both types earn the exact same strictly positive information rent. Focusing on policy $(\overline{A}^{\circ},\overline{B}^{\circ},\overline{D}^{\circ})$, FFS payments $(\overline{A}^{\circ},\overline{B}^{\circ})$ must not be lowered because otherwise IC (15) and/or IC (16) would be violated. Capitation $\overline{D}^{\circ}$ must not be lowered because otherwise the non-negativity assumption would be violated. Hence, $\overline{A}^{\circ}$ and $\overline{B}^{\circ}$ can only jointly be raised at a ratio of $\overline{A}^{\circ} = (1-p)\,\overline{B}^{\circ}$. A separate raise is not feasible, because otherwise IC (15) or IC (16) would be violated. If $\overline{A}^{\circ}$ and $\overline{B}^{\circ}$ were jointly raised, while $\overline{D}^{\circ}$ was not lowered, the rent payments to the $\overline{a}$-types would increase. Hence, to improve on

the overall outcome, policy $(\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$ would need to economize on the rent payments to the $\underline{a}$-types. Yet, it follows from Lemma 4 that the $\underline{a}$-types always earn the exact same rent as the $\overline{a}$-types when picking $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ})$ instead of $(\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$. Thus, there exists no policy $(\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$ that satisfies IC (14) and economizes on the rent payments. Hence, the optimal menu consists of a non-discriminatory policy $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ}) = (\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$, which coincides with contract $(\overline{B}^{*}, \overline{A}^{*}, \overline{D}^{*})$.

The optimal menu for a unifying equilibrium (area [v] in Figure *5-2):* Because single policy contract $(\underline{B}^{*}, \underline{A}^{*}, \underline{D}^{*})$ implements the First-Best, it serves as a starting point and as a benchmark.

- For all ω, $\overline{a}$-types earn an excess strictly positive information rent accruing from the hidden knowledge about the type.
- In case $\omega \geq \frac{v}{p(2\underline{a}-1)}$, I first examine policy $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ})$. Jointly raising FFS payments $(\overline{A}^{\circ}, \overline{B}^{\circ})$ up to the point, at which lowering $\overline{D}^{\circ}$ still more than offsets the higher FFS earnings or, setting $(\overline{A}^{\circ}, \overline{B}^{\circ}) = (\underline{B}^{*}, \underline{A}^{*})$ and lowering $\overline{D}^{\circ}$ or, jointly lowering $(\overline{A}^{\circ}, \overline{B}^{\circ})$ until ICs (15) and (16) bind and, if necessary, increasing $\overline{D}^{\circ}$, saves on the rent payments to the $\overline{a}$-types, while still providing effort incentives. Turning to policy $(\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$, it becomes obvious that the latter cannot be incentivizing. If the policy was incentivizing, IC (13) would be violated as the $\overline{a}$-types would also strictly prefer policy $(\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ})$ over $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ})$. This results in a costly effort distortion for the $\underline{a}$-types. Note, that single policy contract $(\overline{B}^{*}, \overline{A}^{*}, \overline{D}^{*})$ already implements said allocation by paying zero rents to either type. Hence, the optimal menu consists of a single non-discriminatory policy $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ}) = (\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ}) = (\overline{B}^{*}, \overline{A}^{*}, \overline{D}^{*})$.
- In case $\frac{v}{p(2\underline{a}-1)} > \omega$, both types additionally earn the exact same rent resulting from the limitation on $\underline{D}^{*}$. Initially focusing on policy $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ})$ shows that jointly lowering $(\overline{A}^{\circ}, \overline{B}^{\circ})$ until ICs (15) and (16) bind and, if necessary, increasing $\overline{D}^{\circ}$ to meet the PC, reduces the rent payments to the $\overline{a}$-types. Analogous to the previous case, the optimal policy for the $\underline{a}$-types can only be non-incentivizing as otherwise IC (13) would be violated. Again, contract $(\overline{B}^{*}, \overline{A}^{*}, \overline{D}^{*})$ already implements said allocation at the lowest possible cost, which,

interestingly, includes paying a strictly positive information rent to both types. Hence, the optimal menu consists of a single non-discriminatory policy $(\overline{A}^{\circ}, \overline{B}^{\circ}, \overline{D}^{\circ}) = (\underline{A}^{\circ}, \underline{B}^{\circ}, \underline{D}^{\circ}) = (\overline{B}^{*}, \overline{A}^{*}, \overline{D}^{*})$. □

Appendix C

Proof for Lemma 6:

Plugging in for $ESC(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ and $ESC(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ into equation (16) and rewriting yields

$$\lambda\left[v + \left[p\underline{a} + (1-p)(1-\underline{a})\right]h + p(1-\underline{a})(h+\ell) - \min\{p(h+\ell), h\} - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]$$
$$\geq v + \left[p\underline{a} + (1-p)(1-\underline{a})\right]h + p(1-\underline{a})(h+\ell) - \min\{p(h+\ell), h\} +$$
$$\max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\} - \max\left\{\frac{v}{p(2\overline{a}-1)}, \omega\right\}$$

The above constraint may be simplified by distinguishing between three value ranges for the physicians' reservation utility ω:

(I) In case ω takes relatively large values, $\omega \geq \frac{v}{p(2\underline{a}-1)}$, which is in case both ability-types' PCs bind:

$$\lambda\left[v + \left[p\underline{a} + (1-p)(1-\underline{a})\right]h + p(1-\underline{a})(h+\ell) - \min\{p(h+\ell), h\} - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]$$
$$\geq v + \left[p\underline{a} + (1-p)(1-\underline{a})\right]h + p(1-\underline{a})(h+\ell) - \min\{p(h+\ell), h\}$$

- For $\ell \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting into the constraint above gives

  $$\lambda\left[v + (p-1)\underline{a}h + p(1-\underline{a})\ell - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v + (p-1)\underline{a}h + p(1-\underline{a})\ell$$

  - For $\ell > \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ never dominates.
  - For $\ell = \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left side is zero. Thus, $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ never dominates.
  - For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} \leq \ell < \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ never dominates.
  - For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} > \ell$, both sides are negative. Thus, $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} > \ell$ is a necessary condition for the dominance of contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Further, $\lambda \leq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting into the constraint above gives

  $$\lambda\left[v + (1-p)\left(1-\underline{a}\right)h - p\underline{a}\ell - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v + (1-p)\left(1-\underline{a}\right)h - p\underline{a}\ell$$

  - For $\ell < \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$ both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\ell = \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left side is zero. Thus, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} \geq \ell > \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} < \ell$, both sides are negative. Thus, $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} < \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Further, $\lambda \leq \frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

(II) In case ω takes intermediate values, $\frac{v}{p(2\underline{a}-1)} > \omega \geq \frac{v}{p(2\overline{a}-1)}$, which is in case only the $\overline{a}$-types' PC binds:

$$\lambda\left[v + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) - \min\{p(h+\ell), h\} - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]$$
$$\geq v + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) - \min\{p(h+\ell), h\} + \frac{v}{p(2\underline{a}-1)} - \omega$$

- For $\ell \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting into the constraint above gives

  $$\lambda\left[v + (p-1)\underline{a}h + p\left(1-\underline{a}\right)\ell - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v + (p-1)\underline{a}h + p\left(1-\underline{a}\right)\ell + \frac{v}{p(2\underline{a}-1)} - \omega$$

  - For $\ell > \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\ell = \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left side is zero. Thus, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.

- For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)}-\frac{\omega}{p(\underline{a}-1)} \leq \ell < \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
- For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)}-\frac{\omega}{p(\underline{a}-1)} > \ell$, both sides are negative. Thus, $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{v}{p^2(\underline{a}-1)(2\underline{a}-1)}-\frac{\omega}{p(\underline{a}-1)} > \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Further, $\lambda \leq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting into the constraint above gives $\lambda\left[v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v+(1-p)\left(1-\underline{a}\right)h-p\underline{a}\ell+\frac{v}{p(2\underline{a}-1)}-\omega$
  - For $\ell < \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}-\frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\ell = \frac{v+(1-p)(1-\overline{a})h}{p\overline{a}}-\frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left side is zero. Thus, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{v}{p^2\underline{a}(2\underline{a}-1)}-\frac{\omega}{p\underline{a}} \geq \ell > \frac{v+(1-p)(1-\overline{a})h}{p\overline{a}}-\frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{v}{p^2\underline{a}(2\underline{a}-1)}-\frac{\omega}{p\underline{a}} < \ell$, both sides are negative. Thus, $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{v}{p^2\underline{a}(2\underline{a}-1)}-\frac{\omega}{p\underline{a}} < \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Further, $\lambda \leq \frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell+\frac{v}{p(2\underline{a}-1)}-\omega}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

(III) In case ω takes relatively small values, $\frac{v}{p(2\overline{a}-1)} > \omega$, which is in case neither ability-types' PC binds:

$$\lambda\left[v+\left[p\underline{a}+(1-p)\left(1-\underline{a}\right)\right]h+p\left(1-\underline{a}\right)(h+\ell)-\min\{p(h+\ell), h\}-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]$$

$$\geq v + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) - \min\{p(h+\ell), h\} + \frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}$$

- For $\ell \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting into the constraint above gives $\lambda\left[v + (p-1)\underline{a}h + p\left(1-\underline{a}\right)\ell - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v + (p-1)\underline{a}h + p\left(1-\underline{a}\right)\ell + \frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}$
  - For $\ell > \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\ell = \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left side is zero. Thus, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} \leq \ell < \frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} - \frac{2v(\overline{a}-\underline{a})}{p(\underline{a}-1)(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} > \ell$, both sides are negative. Thus, $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)} > \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$. Further, $\lambda \leq \frac{v+(p-1)\underline{a}h+p(1-\underline{a})\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(p-1)\underline{a}h+p(1-\underline{a})\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting into the constraint above gives $\lambda\left[v + (1-p)\left(1-\underline{a}\right)h - p\underline{a}\ell - \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \geq v + (1-p)\left(1-\underline{a}\right)h - p\underline{a}\ell + \frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}$
  - For $\ell < \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}} > 1$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\ell = \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left side is zero. Thus, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.
  - For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)} \geq \ell > \frac{v+(1-p)(1-\underline{a})h}{p\underline{a}} - \frac{2v(\overline{a}-\underline{a})}{p\underline{a}(2\underline{a}-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ never dominates.

- For $\frac{v+(1-p)(1-\underline{a})h}{p\underline{a}}+\frac{2v(\overline{a}-\underline{a})}{p^2\underline{a}(2\underline{a}-1)(2\overline{a}-1)}<\ell$, both sides are negative. Thus, $\frac{v+(p-1)\underline{a}h}{p(\underline{a}-1)}+\frac{2v(\overline{a}-\underline{a})}{p^2(\underline{a}-1)(2\underline{a}-1)(2\overline{a}-1)}>\ell$ is a necessary condition for the dominance of contract $(\underline{B}^*,\underline{A}^*,\underline{D}^*)$ over contract $(\overline{B}^*,\overline{A}^*,\overline{D}^*)$. Further, $\lambda\leq\frac{v+(1-p)(1-\underline{a})h-p\underline{a}\ell+\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{v+(1-p)(1-\underline{a})h-p\underline{a}\ell-\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$.

□

Appendix D

Proof for Proposition 1:

The dominance of contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ over contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ implies that contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is also second-best optimal, only if the SB conditions for contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ are not stronger than the dominance conditions. While the dominance conditions are defined by Lemma 6, the conditions for the second-best solutions, given the payer offers contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$, are defined below:

- Implementing course 1 for both types (area [i] in Figure 5-2) is SB optimal only if

$$\ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}\right\}.$$

- Implementing course 2 for both types (area [ii]) is SB optimal only if

$$\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\right\}.$$

- Implementing course 3 for both types (area [v]) is SB optimal only if

$$\max\left\{\frac{(1-p)h}{p}, \frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\right\} \geq \ell \geq$$
$$\min\left\{\frac{(1-p)h}{p}, \frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}\right\}.$$

Subsequently comparing the conditions, under which the implementation of contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ is second-best, to the necessary conditions for ℓ, under which contract $(\underline{B}^*, \underline{A}^*, \underline{D}^*)$ possibly dominates contract $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ for the three different value ranges of the physicians' reservation utility, ω, yields:

(I) In case ω takes relatively large values, $\omega \geq \frac{v}{p(2\underline{a}-1)}$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB and the dominance constraint on ℓ increase in h with slopes $> \frac{(1-p)}{p}$. For the value $h = \frac{v}{(1-p)(2\underline{a}-1)}$, they are just equal to $\frac{(1-p)h}{p}$.

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} = \frac{(1-p)h}{p} \quad \leftrightarrow$$

$$v+(p-1)\underline{a}h+\lambda\left[(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]=(1-p)h\left[\underline{a}-1+\lambda\left(\overline{a}-\underline{a}\right)\right] \quad \leftrightarrow$$

$$v+\lambda\left[\frac{2(\overline{a}-\underline{a})[(p-1)(2\underline{a}-1)h+v]}{(2\underline{a}-1)}\right]=(1-p)(2\underline{a}-1)h \quad \leftrightarrow$$

$$v\left[\left(2\underline{a}-1\right)+2\lambda\left(\overline{a}-\underline{a}\right)\right]=(1-p)\left(2\underline{a}-1\right)h\left[\left(2\underline{a}-1\right)+2\lambda\left(\overline{a}-\underline{a}\right)\right] \quad \leftrightarrow$$

$$h=\frac{v}{(1-p)(2\underline{a}-1)}$$

$$\frac{(p-1)\underline{a}h+v}{p(\underline{a}-1)}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$(p-1)\underline{a}h+v=(1-p)h\left(\underline{a}-1\right) \quad \leftrightarrow$$

$$h=\frac{v}{(2\underline{a}-1)(1-p)} \quad \leftrightarrow$$

The SB constraint is not stronger than the dominance constraint on ℓ only if:

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}\geq\frac{(p-1)\underline{a}h+v}{p(\underline{a}-1)} \quad \leftrightarrow$$

$$v+(p-1)\underline{a}h+\lambda\left[(p-1)\overline{a}h-(p-1)\underline{a}h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]\leq\frac{[(p-1)\underline{a}h+v][\underline{a}-1+\lambda(\overline{a}-\underline{a})]}{(\underline{a}-1)} \quad \leftrightarrow$$

$$\left(\underline{a}-1\right)\lambda\left[(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]\geq\lambda\left[(p-1)\underline{a}\left(\overline{a}-\underline{a}\right)h+\left(\overline{a}-\underline{a}\right)v\right] \quad \leftrightarrow$$

$$\left(\underline{a}-1\right)\left[(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]\geq(p-1)\underline{a}\left(\overline{a}-\underline{a}\right)h+\left(\overline{a}-\underline{a}\right)v \quad \leftrightarrow$$

$$-(p-1)h+\frac{2v(\underline{a}-1)}{(2\underline{a}-1)}\geq v \quad \leftrightarrow$$

$$h\geq\frac{v}{(2\underline{a}-1)(1-p)}$$

- For all $\ell<\frac{(1-p)h}{p}$, the SB and dominance constraint on ℓ increase in h with slopes $<\frac{(1-p)}{p}$. For the value $h=\frac{v}{(1-p)(2\underline{a}-1)}$, they are just equal to $\frac{(1-p)h}{p}$.

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$v+(1-p)\left(1-\underline{a}\right)h+\lambda\left[(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]=(1-p)h\left[\underline{a}+\lambda\left(\overline{a}-\underline{a}\right)\right] \quad \leftrightarrow$$

$$v\left[\left(2\underline{a}-1\right)+2\lambda\left(\overline{a}-\underline{a}\right)\right]=(1-p)\left(2\underline{a}-1\right)h\left[\left(2\underline{a}-1\right)+2\lambda\left(\overline{a}-\underline{a}\right)\right] \quad \leftrightarrow$$

$$h=\frac{v}{(1-p)(2\underline{a}-1)}$$

$$\frac{(1-p)(1-\underline{a})h+v}{p\underline{a}}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$v=(1-p)h\left(\underline{a}+\underline{a}-1\right) \quad \leftrightarrow$$

$$h = \frac{v}{(2\underline{a}-1)(1-p)}$$

The SB constraint is not stronger than the dominance constraint on ℓ only if:

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]} \leq \frac{(1-p)(1-\underline{a})h+v}{p\underline{a}} \quad \leftrightarrow$$

$$v(1-p)(1-\underline{a})h + \lambda\left[(p-1)(\overline{a}-\underline{a})h + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \leq \frac{[(1-p)(1-\underline{a})h+v][\underline{a}+\lambda(\overline{a}-\underline{a})]}{\underline{a}} \quad \leftrightarrow$$

$$v + (1-p)(1-\underline{a})h + \lambda\left[(p-1)(\overline{a}-\underline{a})h + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \leq$$

$$(1-p)(1-\underline{a})h + v(1-\lambda) + \lambda h\left[(p-1)(\overline{a}-\underline{a}+1)\right] + \frac{\overline{a}\lambda(v+h(1-p))}{\underline{a}} \quad \leftrightarrow$$

$$\lambda\left[(p-1)(\overline{a}-\underline{a})h + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \leq -\lambda v + \lambda h\left[(p-1)(\overline{a}-\underline{a}+1)\right] + \frac{\overline{a}\lambda(v+h(1-p))}{\underline{a}} \quad \leftrightarrow$$

$$(p-1)(\overline{a}-\underline{a})h + \frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)} \leq -v + h(p-1)(\overline{a}-\underline{a}) + (p-1)h + \frac{\overline{a}(v+h(1-p))}{\underline{a}} \quad \leftrightarrow$$

$$v(\overline{a}-\underline{a}) \leq h\left[(2\underline{a}-1)(\overline{a}-\underline{a}) - p(2\underline{a}-1)(\overline{a}-\underline{a})\right] \quad \leftrightarrow$$

$$h \geq \frac{v}{(2\underline{a}-1)(1-p)}$$

(II) In case ω takes intermediate values, $\frac{v}{p(2\underline{a}-1)} > \omega \geq \frac{v}{p(2\overline{a}-1)}$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB and the dominance constraint on ℓ increase in h with slopes $> \frac{(1-p)}{p}$. They are equal to $\frac{(1-p)h}{p}$ only if:

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} = \frac{(1-p)h}{p} \quad \leftrightarrow$$

$$2(p-1)\underline{a}h - (p-1)h + 2\lambda(p-1)(\overline{a}-\underline{a})h = -v + \omega - \frac{v}{p(2\underline{a}-1)} - \lambda\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)} \quad \leftrightarrow$$

$$h(p-1)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right) = -\frac{v[p(2\underline{a}-1)+1+\lambda 2p(\overline{a}-\underline{a})]}{p(2\underline{a}-1)} + \omega \quad \leftrightarrow$$

$$h = \frac{v[1+p(2\underline{a}-1)+\lambda 2p(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]} - \frac{\omega}{(1-p)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]}$$

$$\frac{(p-1)\underline{a}h+v-\omega}{p(\underline{a}-1)} + \frac{v}{p^2(2\underline{a}-1)(\underline{a}-1)} = \frac{(1-p)h}{p} \quad \leftrightarrow$$

$$(p-1)\underline{a}h + v - \omega + \frac{v}{p(2\underline{a}-1)} = (1-p)(\underline{a}-1)h \quad \leftrightarrow$$

$$h = \frac{v\left(1+p(2\underline{a}-1)\right)}{p(1-p)(2\underline{a}-1)^2} - \frac{\omega}{(1-p)(2\underline{a}-1)}$$

The SB constraint is not stronger than the dominance constraint on ℓ only if:

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} \geq \frac{(p-1)\underline{a}h+v-\omega}{p(\underline{a}-1)}+\frac{v}{p^2(2\underline{a}-1)(\underline{a}-1)} \leftrightarrow$$

$$v-\omega+(1-\lambda)\left[(p-1)\underline{a}h+\frac{v}{p(2\underline{a}-1)}\right]+\lambda\left[(p-1)\overline{a}h+\frac{v[1+2p(\overline{a}-\underline{a})]}{p(2\underline{a}-1)}\right] \leq$$

$$\frac{(p-1)\underline{a}h[\underline{a}-1+\lambda(\overline{a}-\underline{a})]+v[\underline{a}-1+\lambda(\overline{a}-\underline{a})]-\omega[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}{(\underline{a}-1)}+\frac{v[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}{p(2\underline{a}-1)(\underline{a}-1)} \leftrightarrow$$

$$\lambda\left[(p-1)(\overline{a}-\underline{a})h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right] \leq \frac{\lambda(p-1)\underline{a}h(\overline{a}-\underline{a})+v\lambda(\overline{a}-\underline{a})-\omega\lambda(\overline{a}-\underline{a})}{(\underline{a}-1)}+\frac{v\lambda(\overline{a}-\underline{a})}{p(2\underline{a}-1)(\underline{a}-1)} \leftrightarrow$$

$$(\underline{a}-1)\left[(p-1)h+\frac{2v}{(2\underline{a}-1)}\right] \geq \left[(p-1)\underline{a}h+v-\omega+\frac{v}{p(2\underline{a}-1)}\right] \leftrightarrow$$

$$h \geq \frac{v(1+p)}{p(1-p)(2\underline{a}-1)}-\frac{\omega}{(1-p)}$$

The above condition for h always holds because $\frac{v(1+p)}{p(1-p)(2\underline{a}-1)}-\frac{\omega}{(1-p)} <$

$$\frac{v[1+p(2\underline{a}-1)+p2\lambda(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)}-\frac{\omega}{(1-p)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)} < \frac{v\left(1+p(2\underline{a}-1)\right)}{p(1-p)(2\underline{a}-1)^2}-\frac{\omega}{(1-p)(2\underline{a}-1)}.$$

$$\frac{v(1+p)}{p(1-p)(2\underline{a}-1)}-\frac{\omega}{(1-p)} < \frac{v[1+p(2\underline{a}-1)+p2\lambda(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)}-\frac{\omega}{(1-p)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)} \leftrightarrow$$

$$-\omega p(2\underline{a}-1)^2-\omega p(2\underline{a}-1)2\lambda(\overline{a}-\underline{a})+\omega p(2\underline{a}-1) < v-v(2\underline{a}-1)-v2\lambda(\overline{a}-\underline{a}) \leftrightarrow$$

$$\omega p(2\underline{a}-1)\left(-(2\underline{a}-1)-2\lambda(\overline{a}-\underline{a})+1\right) < v\left(1-(2\underline{a}-1)-2\lambda(\overline{a}-\underline{a})\right) \leftrightarrow$$

$$2\omega p(2\underline{a}-1)\left(1-\underline{a}-\lambda(\overline{a}-\underline{a})\right) < 2v\left(1-\underline{a}-\lambda(\overline{a}-\underline{a})\right) \leftrightarrow$$

$$\omega < \frac{v}{p(2\underline{a}-1)}$$

$$\frac{v[1+p(2\underline{a}-1)+p2\lambda(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)}-\frac{\omega}{(1-p)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right)} < \frac{v\left(1+p(2\underline{a}-1)\right)}{p(1-p)(2\underline{a}-1)^2}-\frac{\omega}{(1-p)(2\underline{a}-1)} \leftrightarrow$$

$$0 < v2\lambda(\overline{a}-\underline{a})-\omega p(2\underline{a}-1)2\lambda(\overline{a}-\underline{a}) \leftrightarrow$$

$$\omega p(2\underline{a}-1) < v \leftrightarrow$$

$$\omega < \frac{v}{p(2\underline{a}-1)}$$

- For all $\ell < \frac{(1-p)h}{p}$, the SB and the dominance constraint on ℓ increase in h with slopes $< \frac{(1-p)}{p}$. They are equal to $\frac{(1-p)h}{p}$ only if:

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]} = \frac{(1-p)h}{p} \leftrightarrow$$

$$\frac{v-\omega+(1-p)h+(p-1)\underline{a}h+\frac{v}{p(2\underline{a}-1)}+\lambda\left[(p-1)(\overline{a}-\underline{a})h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{[\underline{a}+\lambda(\overline{a}-\underline{a})]} = (1-p)h \leftrightarrow$$

$$v-\omega+(1-p)\left(1-\underline{a}\right)h+\frac{v}{p\left(2\underline{a}-1\right)}+\lambda(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{\lambda 2v\left(\overline{a}-\underline{a}\right)}{\left(2\underline{a}-1\right)}=$$

$$(1-p)h\underline{a}+\lambda(1-p)\left(\overline{a}-\underline{a}\right)h \quad \leftrightarrow$$

$$h(p-1)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]=-\frac{v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right]}{p\left(2\underline{a}-1\right)}+\omega \quad \leftrightarrow$$

$$h=\frac{v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right]}{p(1-p)\left(2\underline{a}-1\right)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}-\frac{\omega}{(1-p)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}$$

$$\frac{(1-p)\left(1-\underline{a}\right)h+v-\omega}{p\underline{a}}+\frac{v}{p^2\left(2\underline{a}-1\right)\underline{a}}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$(1-p)\left(1-\underline{a}\right)hp\left(2\underline{a}-1\right)+vp\left(2\underline{a}-1\right)-\omega p\left(2\underline{a}-1\right)+v=(1-p)h\underline{a}p\left(2\underline{a}-1\right) \quad \leftrightarrow$$

$$(1-p)\left(1-\underline{a}\right)hp\left(2\underline{a}-1\right)-(1-p)h\underline{a}p\left(2\underline{a}-1\right)=-vp\left(2\underline{a}-1\right)-v+\omega p\left(2\underline{a}-1\right) \quad \leftrightarrow$$

$$hp(1-p)\left(2\underline{a}-1\right)\left(1-2\underline{a}\right)=-v\left[p\left(2\underline{a}-1\right)+1\right]+\omega p\left(2\underline{a}-1\right) \quad \leftrightarrow$$

$$h=\frac{v\left[1+p\left(2\underline{a}-1\right)\right]}{p(1-p)\left(2\underline{a}-1\right)^2}-\frac{\omega}{(1-p)\left(2\underline{a}-1\right)}$$

The SB constraint is not stronger than the dominance constraint on ℓ only if:

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p\left(2\underline{a}-1\right)},\omega\right\}+(1-\lambda)\left[(p-1)\underline{a}+1\right]h+\lambda\left[\left[(p-1)\overline{a}+1\right]h+\frac{2v\left(\overline{a}-\underline{a}\right)}{\left(2\underline{a}-1\right)}\right]}{p\left[\underline{a}+\lambda\left(\overline{a}-\underline{a}\right)\right]}\leq\frac{(1-p)\left(1-\underline{a}\right)h+v-\omega}{p\underline{a}}+\frac{v}{p^2\left(2\underline{a}-1\right)\underline{a}} \quad \leftrightarrow$$

$$\frac{v}{p\left(2\underline{a}-1\right)}+\lambda\left[(p-1)\left(\overline{a}-\underline{a}\right)h+\frac{2v\left(\overline{a}-\underline{a}\right)}{\left(2\underline{a}-1\right)}\right]\leq\frac{(1-p)\left(1-\underline{a}\right)\lambda\left(\overline{a}-\underline{a}\right)h+v\lambda\left(\overline{a}-\underline{a}\right)-\omega\lambda\left(\overline{a}-\underline{a}\right)}{\underline{a}}+\frac{v\left[\underline{a}+\lambda\left(\overline{a}-\underline{a}\right)\right]}{p\left(2\underline{a}-1\right)\underline{a}} \quad \leftrightarrow$$

$$\lambda h(p-1)\left(\overline{a}-\underline{a}\right)\leq\lambda\left(\overline{a}-\underline{a}\right)\left[\frac{v\left[p\left(2\underline{a}-1\right)+1-2p\underline{a}\right]}{p\left(2\underline{a}-1\right)}-\omega\right] \quad \leftrightarrow$$

$$h\geq\frac{v(p-1)}{p(1-p)\left(2\underline{a}-1\right)}+\frac{\omega}{(1-p)}$$

The above condition for h always holds because $\frac{v(p-1)}{p(1-p)\left(2\underline{a}-1\right)}+\frac{\omega}{(1-p)}<$

$$\frac{v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right]}{p(1-p)\left(2\underline{a}-1\right)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}-\frac{\omega}{(1-p)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}<\frac{v\left[1+p\left(2\underline{a}-1\right)\right]}{p(1-p)\left(2\underline{a}-1\right)^2}-\frac{\omega}{(1-p)\left(2\underline{a}-1\right)}.$$

$$\frac{v(p-1)}{p(1-p)\left(2\underline{a}-1\right)}+\frac{\omega}{(1-p)}<\frac{v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right]}{p(1-p)\left(2\underline{a}-1\right)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}-\frac{\omega}{(1-p)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]} \quad \leftrightarrow$$

$$\omega p\left(2\underline{a}-1\right)\left[2\underline{a}+2\lambda\left(\overline{a}-\underline{a}\right)\right]<v\left[2\underline{a}+2\lambda\left(\overline{a}-\underline{a}\right)\right] \quad \leftrightarrow$$

$$\omega<\frac{v}{p\left(2\underline{a}-1\right)}$$

$$\frac{v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right]}{p(1-p)\left(2\underline{a}-1\right)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}-\frac{\omega}{(1-p)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]}<\frac{v\left[1+p\left(2\underline{a}-1\right)\right]}{p(1-p)\left(2\underline{a}-1\right)^2}-\frac{\omega}{(1-p)\left(2\underline{a}-1\right)} \quad \leftrightarrow$$

$$p\omega\left(2\underline{a}-1\right)2\lambda\left(\overline{a}-\underline{a}\right)<v2\lambda\left(\overline{a}-\underline{a}\right) \quad \leftrightarrow$$

$$\omega<\frac{v}{p\left(2\underline{a}-1\right)}$$

(III) In case ω takes relatively small values, $\frac{v}{p(2\overline{a}-1)} > \omega$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB and the dominance constraint on ℓ increase in h with slopes

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}$$

$$\frac{\partial f}{\partial h} = \frac{(p-1)\underline{a}+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]}$$

$$\frac{(p-1)\underline{a}h+v}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(2\underline{a}-1)(2\overline{a}-1)(\underline{a}-1)}$$

$$\frac{\partial f}{\partial h} = \frac{(p-1)\underline{a}}{p(\underline{a}-1)}$$

Comparing the slopes yields $\frac{(p-1)\underline{a}+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} > \frac{(p-1)\underline{a}}{p(\underline{a}-1)} > \frac{(1-p)}{p}$.

$$\frac{(p-1)\underline{a}+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} > \frac{(p-1)\underline{a}}{p(\underline{a}-1)} \quad \leftrightarrow$$

$$(\underline{a}-1)[\underline{a}+\lambda(\overline{a}-\underline{a})] < \underline{a}[\underline{a}-1+\lambda(\overline{a}-\underline{a})] \quad \leftrightarrow$$

$$\underline{a}^2-\underline{a}+\lambda\overline{a}\underline{a}-\lambda\underline{a}^2-\lambda\overline{a}+\lambda\underline{a} < \underline{a}^2-\underline{a}+\lambda\overline{a}\underline{a}-\lambda\underline{a}^2 \quad \leftrightarrow$$

$$\underline{a} < \overline{a}$$

$$\frac{(p-1)\underline{a}}{p(\underline{a}-1)} > \frac{(1-p)}{p} \quad \leftrightarrow$$

$$(p-1)\underline{a} < (1-p)(\underline{a}-1) \quad \leftrightarrow$$

$$(p-1)\underline{a} < (p-1)(1-\underline{a}) \quad \leftrightarrow$$

$$2\underline{a} > 1$$

The constraints are equal to $\frac{(1-p)h}{p}$ only if:

$$\frac{v-\omega-h+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}-1+\lambda(\overline{a}-\underline{a})]} = \frac{(1-p)h}{p} \quad \leftrightarrow$$

$$2(p-1)\underline{a}h-(p-1)h+2\lambda(p-1)(\overline{a}-\underline{a})h = -v+\omega-\frac{v}{p(2\underline{a}-1)}-\lambda\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)} \quad \leftrightarrow$$

$$h(p-1)\left(2\underline{a}-1+2\lambda(\overline{a}-\underline{a})\right) = -\frac{v[p(2\underline{a}-1)+1+\lambda 2p(\overline{a}-\underline{a})]}{p(2\underline{a}-1)}+\omega \quad \leftrightarrow$$

$$h = \frac{v[1+p(2\underline{a}-1)+\lambda 2p(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]} - \frac{\omega}{(1-p)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]}$$

$$\frac{(p-1)\underline{a}h+v}{p(\underline{a}-1)} + \frac{2v(\overline{a}-\underline{a})}{p^2(2\underline{a}-1)(2\overline{a}-1)(\underline{a}-1)} = \frac{(1-p)h}{p} \quad \leftrightarrow$$

$$(p-1)\underline{a}h + v + \frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)} = (1-p)h(\underline{a}-1) \quad \leftrightarrow$$

$$(p-1)(2\underline{a}-1)h = -\frac{v[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})]}{p(2\underline{a}-1)(2\overline{a}-1)} \quad \leftrightarrow$$

$$h = \frac{v[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)^2(2\overline{a}-1)}$$

Thus, the SB constraint is not stronger than the dominance constraint on ℓ only if:

$$\frac{v[1+p(2\underline{a}-1)+\lambda 2p(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]} - \frac{\omega}{(1-p)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]} \leq \frac{v[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)^2(2\overline{a}-1)} \quad \leftrightarrow$$

$$(2\underline{a}-1)(2\overline{a}-1)[1+p(2\underline{a}-1)+\lambda 2p(\overline{a}-\underline{a})]v - p(2\underline{a}-1)^2(2\overline{a}-1)\omega \leq$$

$$[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})][2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]v \quad \leftrightarrow$$

$$(2\underline{a}-1)[(2\overline{a}-1)v - 2(\overline{a}-\underline{a})v - p(2\underline{a}-1)(2\overline{a}-1)\omega] \leq 4\lambda(\overline{a}-\underline{a})^2 v \quad \leftrightarrow$$

$$\frac{(2\underline{a}-1)[[(2\overline{a}-1)-2(\overline{a}-\underline{a})]v - p(2\underline{a}-1)(2\overline{a}-1)\omega]}{4(\overline{a}-\underline{a})^2 v} \leq \lambda$$

- For all $\ell < \frac{(1-p)h}{p}$, the SB and the dominance constraint on ℓ increase in h with slopes

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}$$

$$\frac{\partial f}{\partial h} = \frac{(1-p)(1-\underline{a})+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}$$

$$\frac{(1-p)(1-\underline{a})h+v}{p\underline{a}} + \frac{2v(\overline{a}-\underline{a})}{p^2(2\underline{a}-1)(2\overline{a}-1)\underline{a}}$$

$$\frac{\partial f}{\partial h} = \frac{(1-p)(1-\underline{a})}{p\underline{a}}$$

Comparing the slopes yields $\frac{(1-p)}{p} > \frac{(1-p)(1-\underline{a})}{p\underline{a}} > \frac{(1-p)(1-\underline{a})+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}$.

$$\frac{(1-p)}{p} > \frac{(1-p)(1-\underline{a})}{p\underline{a}} \quad \leftrightarrow$$

$$(1-p)\underline{a} > (1-p)(1-\underline{a}) \quad \leftrightarrow$$

$$2\underline{a} > 1$$

$$\frac{(1-p)(1-\underline{a})}{p\underline{a}} > \frac{(1-p)(1-\underline{a})+\lambda(p-1)(\overline{a}-\underline{a})}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]} \quad \leftrightarrow$$

$$(1-\underline{a})[\underline{a}+\lambda(\overline{a}-\underline{a})] > (1-\underline{a})\underline{a} - \lambda(\overline{a}-\underline{a})\underline{a}$$

$$\underline{a}+\lambda(\overline{a}-\underline{a}) - \underline{a}^2 - \lambda(\overline{a}\underline{a}-\underline{a}^2) > \underline{a} - \underline{a}^2 - \lambda(\overline{a}\underline{a}-\underline{a}^2) \quad \leftrightarrow$$

$$\overline{a} > \underline{a}$$

The constraints are equal to $\frac{(1-p)h}{p}$ only if:

$$\frac{v-\omega-ph+\max\left\{\frac{v}{p(2\underline{a}-1)},\omega\right\}+(1-\lambda)[(p-1)\underline{a}+1]h+\lambda\left[[(p-1)\overline{a}+1]h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{p[\underline{a}+\lambda(\overline{a}-\underline{a})]}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$\frac{v-\omega+(1-p)h+(p-1)\underline{a}h+\frac{v}{p(2\underline{a}-1)}+\lambda\left[(p-1)(\overline{a}-\underline{a})h+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right]}{[\underline{a}+\lambda(\overline{a}-\underline{a})]}=(1-p)h \quad \leftrightarrow$$

$$h(p-1)\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]=-\frac{v[p(2\underline{a}-1)+1+\lambda 2p(\overline{a}-\underline{a})]}{p(2\underline{a}-1)}+\omega \quad \leftrightarrow$$

$$h=\frac{v[p(2\underline{a}-1)+1+\lambda 2p(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]}-\frac{\omega}{(1-p)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]}$$

$$\frac{(1-p)(1-\underline{a})h+v}{p\underline{a}}+\frac{2v(\overline{a}-\underline{a})}{p^2(2\underline{a}-1)(2\overline{a}-1)\underline{a}}=\frac{(1-p)h}{p} \quad \leftrightarrow$$

$$p(1-p)\left(2\underline{a}-1\right)(2\overline{a}-1)\left(1-2\underline{a}\right)h=-\left[p\left(2\underline{a}-1\right)(2\overline{a}-1)+2\left(\overline{a}-\underline{a}\right)\right]v \quad \leftrightarrow$$

$$h=\frac{[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})]v}{p(1-p)\left(2\underline{a}-1\right)^2(2\overline{a}-1)}$$

Thus, the SB constraint is not stronger than the dominance constraint on ℓ if:

$$\frac{[p(2\underline{a}-1)(2\overline{a}-1)+2(\overline{a}-\underline{a})]v}{p(1-p)\left(2\underline{a}-1\right)^2(2\overline{a}-1)}\geq\frac{v[p(2\underline{a}-1)+1+\lambda 2p(\overline{a}-\underline{a})]}{p(1-p)(2\underline{a}-1)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]}-\frac{\omega}{(1-p)[2\underline{a}-1+2\lambda(\overline{a}-\underline{a})]} \quad \leftrightarrow$$

$$\left[p\left(2\underline{a}-1\right)(2\overline{a}-1)+2\left(\overline{a}-\underline{a}\right)\right]\left[2\underline{a}-1+2\lambda\left(\overline{a}-\underline{a}\right)\right]v\geq$$

$$v\left[p\left(2\underline{a}-1\right)+1+\lambda 2p\left(\overline{a}-\underline{a}\right)\right](2\overline{a}-1)\left(2\underline{a}-1\right)-\omega p(2\overline{a}-1)\left(2\underline{a}-1\right)^2 \quad \leftrightarrow$$

$$2\left(\overline{a}-\underline{a}\right)\left(2\underline{a}-1\right)v+4\lambda\left(\overline{a}-\underline{a}\right)^2v\geq(2\overline{a}-1)\left(2\underline{a}-1\right)v-\omega p(2\overline{a}-1)\left(2\underline{a}-1\right)^2 \quad \leftrightarrow$$

$$4\lambda\left(\overline{a}-\underline{a}\right)^2v\geq\left(2\underline{a}-1\right)\left[\left[(2\overline{a}-1)-2\left(\overline{a}-\underline{a}\right)\right]v-p(2\overline{a}-1)\left(2\underline{a}-1\right)\omega\right] \quad \leftrightarrow$$

$$\lambda\geq\frac{(2\underline{a}-1)\left[\left[(2\overline{a}-1)-2(\overline{a}-\underline{a})\right]v-p(2\overline{a}-1)(2\underline{a}-1)\omega\right]}{4\left(\overline{a}-\underline{a}\right)^2v}$$

→ In case ω takes (I) relatively large or (II) intermediate values, the necessary dominance conditions on ℓ imply the SB conditions for contract $\left(\underline{B}^*,\underline{A}^*,\underline{D}^*\right)$. This simplifies the determination of the second-best solutions: For parameter constellations, for which contract $\left(\underline{B}^*,\underline{A}^*,\underline{D}^*\right)$ dominates contract $(\overline{B}^*,\overline{A}^*,\overline{D}^*)$, contract $\left(\underline{B}^*,\underline{A}^*,\underline{D}^*\right)$ is also SB and for parameter constellations, for which contract $(\overline{B}^*,\overline{A}^*,\overline{D}^*)$ is superior, the second-best solutions are defined as follows:

- Implementing course 1 for both types (area [i] in Figure 5-2) is SB optimal only if $\ell\leq\min\left\{\frac{(1-p)h}{p},\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[v+(1-p)(1-\overline{a})h]}{\lambda p\overline{a}}\right\}$.

- Implementing course 2 for both types (area [ii]) is SB optimal only if $\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[v+(p-1)\overline{a}h]}{\lambda p(\overline{a}-1)}\right\}$.

- Implementing course 3 for the $\overline{a}$-types and course 1 or course 2 for the $\underline{a}$-types (areas [iii] and [iv], respectively), depending which one yields the lower ESC, is SB optimal only if $\frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[v+(p-1)\overline{a}h]}{\lambda p(\overline{a}-1)} \geq \ell \geq \frac{\max\left\{\frac{v}{p(2\overline{a}-1)}-\omega,0\right\}+\lambda[v+(1-p)(1-\overline{a})h]}{\lambda p\overline{a}}$.

→ In case ω takes (III) relatively small values, the dominance of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ over $(\overline{B}^*, \overline{A}^*, \overline{D}^*)$ implies that contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ is also SB, only if λ does not fall below a certain threshold. If this holds, the SB solutions may be derived in the same manner as for the relatively large and intermediate values of ω. In case λ falls below the threshold though, the derivation of the SB solutions is more complex. Since fulfilling the necessary dominance constraint on ℓ is not a sufficient condition for contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ to be second-best, the SB conditions, which were defined at the very beginning of this proof, need to be additionally verified.

□

Appendix E

Proof for Lemma 7:

Given pure moral hazard, the expected social cost from implementing the diagnosis and treatment courses in the second-best situation amount to:

- *Course 1 "Blind treatment choice, always T^B"*: $ESC_{1,MH,SB} = p(h+\ell) + \omega$.
- *Course 2 "Blind treatment choice, always T^A"*: $ESC_{2,MH,SB} = h + \omega$.
- *Course 3 "Informed treatment choice"*:

$$\begin{aligned} ESC_{3,MH,SB} &= \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) + \\ &\quad \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]\underline{A}^* + (1-p)\underline{a}\underline{B}^* + \underline{D}^* \\ &= \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) + \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\} \end{aligned}$$

Course 3 dominates Course 1 if and only if:

$ESC_{1,MH,SB} \geq ESC_{3,MH,SB}$

$$p(h+\ell) + \omega \geq v + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) + \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$\ell \geq \frac{v + (1-p)\left(1-\underline{a}\right)h + \max\left\{\frac{v}{p(2\underline{a}-1)} - w, 0\right\}}{p\underline{a}}.$$

Course 3 dominates Course 2 if and only if:

$ESC_{2,MH,SB} \geq ESC_{3,MH,SB}$

$$h + \omega \geq v + \left[p\underline{a} + (1-p)\left(1-\underline{a}\right)\right]h + p\left(1-\underline{a}\right)(h+\ell) + \max\left\{\frac{v}{p(2\underline{a}-1)}, \omega\right\}$$

$$\frac{v + \underline{a}(p-1)h + \max\left\{\frac{v}{p(2\underline{a}-1)} - w, 0\right\}}{-p(1-\underline{a})} \geq \ell$$

Derivation of the threshold for λ:

Given moral hazard only, the optimality conditions for the implementation of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ are defined by Lemma 7.

Given moral hazard and adverse selection, the optimality conditions for the implementation of contract $\left(\underline{B}^*, \underline{A}^*, \underline{D}^*\right)$ for relatively small values of ω are stated in Table 5-1.

Comparing the conditions yields the following insights:

- In case $\lambda \geq \frac{(2\underline{a}-1)\left[[(2\overline{a}-1)-2(\overline{a}-\underline{a})]v-p(2\overline{a}-1)(2\underline{a}-1)\omega\right]}{4(\overline{a}-\underline{a})^2 v}$, the necessary conditions for the size of ℓ from Table 5-1 are obviously always weaker than the conditions for the size of ℓ from Lemma 7.
- In case $\lambda < \frac{(2\underline{a}-1)\left[[(2\overline{a}-1)-2(\overline{a}-\underline{a})]v-p(2\overline{a}-1)(2\underline{a}-1)\omega\right]}{4(\overline{a}-\underline{a})^2 v}$, in the relevant value range for ℓ, that is $\ell \geq \frac{(1-p)h}{p}$ and $\ell \leq \frac{(1-p)h}{p}$, respectively, the necessary conditions for the size of ℓ from Table 5-1 are weaker than the conditions for the size of ℓ from Lemma 7.

Hence, rearranging the necessary conditions for λ from Table 5-1, equating the latter to the conditions for the size of ℓ from Lemma 7 and subsequently solving for λ gives the threshold.

Rearranging the necessary conditions for λ from Table 5-1 yields:

$$\lambda \leq X_6 \quad \leftrightarrow \quad \frac{(\lambda-1)\left(v+(1-p)(1-\underline{a})h\right)-\frac{\lambda 2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{(\lambda-1)p\underline{a}} \leq \ell$$

$$\lambda \leq X_5 \quad \leftrightarrow \quad \frac{(\lambda-1)\left(v+(p-1)\underline{a}h\right)-\frac{\lambda 2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{(1-\lambda)p(1-\underline{a})} \leq \ell$$

Equating and solving for λ gives:

$$\frac{(\lambda-1)\left(v+(1-p)(1-\underline{a})h\right)-\frac{\lambda 2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{(\lambda-1)p\underline{a}} = \frac{v+(1-p)(1-a)h+\max\left\{\frac{v}{p(2a-1)}-w,0\right\}}{pa} \quad \leftrightarrow$$

$$-\frac{\lambda 2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)} = (\lambda-1)\left(\max\left\{\frac{v}{p(2a-1)}-w,0\right\}\right) \quad \leftrightarrow$$

$$\lambda = \frac{\max\left\{\frac{v}{p(2a-1)}-w,0\right\}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{\max\left\{\frac{v}{p(2a-1)}-w,0\right\}+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$$

$$\frac{(\lambda-1)\left(v+(p-1)\underline{a}h\right)-\frac{\lambda 2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{(1-\lambda)p(1-\underline{a})} = \frac{v+a(p-1)h+\max\left\{\frac{v}{p(2a-1)}-w,0\right\}}{-p(1-a)} \quad \leftrightarrow$$

$$\max\left\{\frac{v}{p(2a-1)}-w,0\right\}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)} = \lambda\left(\max\left\{\frac{v}{p(2a-1)}-w,0\right\}+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}\right) \quad \leftrightarrow$$

$$\lambda = \frac{\max\left\{\frac{v}{p(2a-1)}-w,0\right\}-\frac{2v(\overline{a}-\underline{a})}{p(2\underline{a}-1)(2\overline{a}-1)}}{\max\left\{\frac{v}{p(2a-1)}-w,0\right\}+\frac{2v(\overline{a}-\underline{a})}{(2\underline{a}-1)}}$$

Appendix F

The First-Best:

The first-best analyses follow the same pattern as for the operationalization by means of diagnostic accuracy. The ESC for the diagnosis and treatment courses amount to

- $ESC_{1,v} = p(h + \ell) + \omega$
- $ESC_{2,v} = h + \omega$
- $ESC_{3,v} = v + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) + \omega, \qquad \text{with } \frac{\partial ESC_{3,v}}{\partial v} > 0$

The ESC for the informed treatment choice (course 3) decrease in ability since more able physicians incur less effort cost.

The pairwise comparison of the ESC by type gives the first-best solutions:

- Course 1 "Blind treatment choice, always T^B" if $\ell \leq \min\left\{\frac{1-p}{p}h,\ \frac{v+(1-p)(1-a)h}{pa}\right\}$,
- Course 2 "Blind treatment choice, always T^A" if $\ell \geq \max\left\{\frac{1-p}{p}h,\ \frac{v+(p-1)ah}{p(a-1)}\right\}$ and
- Course 3 "Informed treatment choice" if $\frac{v+(p-1)ah}{p(a-1)} \geq \ell \geq \frac{v+(1-p)(1-a)h}{pa}$,

Whereby, in this case, $v \in \{\underline{v}, \overline{v}\}$ with $\underline{v} > \overline{v}$.

The single policy contracts:

In case the payer offers contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ to the physicians:

- $\overline{v}$-types' expected utilities amount to:

  $EU_1^{\overline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} + (1-p)\overline{B}^{*,v} = \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

  $EU_2^{\overline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} + \overline{A}^{*,v} = \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

  $EU_3^{\overline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} - \overline{v} + (1-p)a\overline{B}^{*,v} + [pa + (1-p)(1-a)]\overline{A}^{*,v}$

  $= \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

- $\underline{v}$-types' expected utilities amount to:

  $EU_1^{\underline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} + (1-p)\overline{B}^{*,v} = \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

  $EU_2^{\underline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} + \overline{A}^{*,v} = \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

  $EU_3^{\underline{v}}(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \overline{D}^{*,v} - \underline{v} + (1-p)a\overline{B}^{*,v} + [pa + (1-p)(1-a)]\overline{A}^{*,v}$

  $= \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\} - (\underline{v} - \overline{v})$

In case contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is offered:

- $\overline{v}$-types' expected utilities amount to:

  $EU_1^{\overline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} + (1-p)\underline{B}^{*,v} = \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$

  $EU_2^{\overline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} + \underline{A}^{*,v} = \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$

  $EU_3^{\overline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} - \overline{v} + (1-p)a\underline{B}^{*,v} + [pa + (1-p)(1-a)]\underline{A}^{*,v}$

  $= \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\} + (\underline{v} - \overline{v})$

- $\underline{v}$-types' expected utilities amount to:

  $EU_1^{\underline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} + (1-p)\underline{B}^{*,v} = \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$

  $EU_2^{\underline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} + \underline{A}^{*,v} = \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$

  $EU_3^{\underline{v}}\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = \underline{D}^{*,v} - \underline{v} + (1-p)a\underline{B}^{*,v} + [pa + (1-p)(1-a)]\underline{A}^{*,v}$

  $= \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

The pairwise comparison of the expected utilities by type shows that under contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ the $\overline{v}$-types are indifferent between courses and, thus, pick the informed treatment choice course. The $\underline{v}$-types generate a higher expected utility from the blind courses than from the informed treatment choice course. In fact, they are indifferent between the blind courses, and, hence, pick the blind course that yields the lower ESC.[12] Under contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$, the $\overline{v}$-types strictly prefer the informed treatment choice course and the $\underline{v}$-types are indifferent between courses and, hence, choose the first-best course.

The Second-Best:

The second-best solutions are determined by, firstly, identifying the dominant single policy contract. The ESC for the single policy contracts are given by:

$$ESC(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) = \lambda[\overline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell)] + (1-\lambda)\min\{p(h+\ell), h\} + \max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$$

$$ESC\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) = [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) + \lambda\left[\overline{v} + \left(\underline{v} - \overline{v}\right)\right] + (1-\lambda)\underline{v} + \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$$

[12] As the treatment constraints (compare ICs (4) and (5)) do not depend on effort cost v, they also hold when a contract is offered to a type, it originally was not designed for. This would be the $\underline{v}$-types and the $\bar{v}$-types under contracts $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ and $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$, respectively. Thus, comparing the expected utilities from the different courses by type is sufficient to determine whether a contract is incentive-compatible.

Contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ dominates contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ if and only if:

$$ESC(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}) \geq ESC\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right) \quad (17)$$

Plugging in for $ESC(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ and $ESC\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ into (17) and rewriting yields

$\lambda[\overline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\}] \geq$
$\underline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\} + \max\left\{\frac{\underline{v}}{p(2a-1)}, \omega\right\}$
$-\max\left\{\frac{\overline{v}}{p(2a-1)}, \omega\right\}$

Again, the dominance constraint may be simplified by distinguishing between three value ranges of the physicians' reservation utility ω:

(I) In case ω takes relatively large values, $\omega \geq \frac{\underline{v}}{p(2a-1)}$, which is in case both ability-types' PCs bind:

$\lambda[\overline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\}] \geq$
$\underline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\}$

- For $l \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting into the simplified constraint gives
  $\lambda[\overline{v} + (p-1)ah + p(1-a)\ell] \geq \underline{v} + (p-1)ah + p(1-a)\ell$
  - For $\ell > \frac{\overline{v}+(p-1)ah}{p(a-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{\underline{v}+(p-1)ah+p(1-a)\ell}{\overline{v}+(p-1)ah+p(1-a)\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\ell = \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} \leq \ell < \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} > \ell$, both sides are negative. Thus, $\frac{\underline{v}+(p-1)ah}{p(a-1)} > \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$. Further, $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell}{\overline{v}+(p-1)ah+p(1-a)\ell}$.
- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting into the simplified constraint gives
  $\lambda[\overline{v} + (1-p)(1-a)h - pa\ell] \geq \underline{v} + (1-p)(1-a)h - pa\ell$

- For $\ell < \frac{\overline{v}+(1-p)(1-a)h}{pa}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{\underline{v}+(1-p)(1-a)h-pa\ell}{\overline{v}+(1-p)(1-a)h-pa\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\ell = \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} \geq \ell > \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} < \ell$, both sides are negative. Thus, $\frac{\underline{v}+(1-p)(1-a)h}{pa} < \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $\left(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}\right)$. Further, $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell}{\overline{v}+(1-p)(1-a)h-pa\ell}$.

(II) In case ω takes intermediate values, $\frac{\underline{v}}{p(2a-1)} > \omega \geq \frac{\overline{v}}{p(2a-1)}$, which is in case only the $\overline{v}$-types' PC binds:

$\lambda[\overline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\}] \geq$
$\underline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\} + \frac{\underline{v}}{p(2a-1)} - \omega$

- For $l \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting into the simplified constraint gives
  $\lambda[\overline{v} + (p-1)ah + p(1-a)\ell] \geq \underline{v} + (p-1)ah + p(1-a)\ell + \frac{\underline{v}}{p(2a-1)} - \omega$
  - For $\ell > \frac{\overline{v}+(p-1)ah}{p(a-1)}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(p-1)ah+p(1-a)\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\ell = \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} \leq \ell < \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} > \ell$, both sides are negative. Thus, $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} > \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $\left(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v}\right)$.
    Further, $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(p-1)ah+p(1-a)\ell}$.
- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting into the simplified constraint gives
  $\lambda[\overline{v} + (1-p)(1-a)h - pa\ell] \geq \underline{v} + (1-p)(1-a)h - pa\ell + \frac{\underline{v}}{p(2a-1)} - \omega$

- For $\ell < \frac{\overline{v}+(1-p)(1-a)h}{pa}$, both sides of the (in)equality condition are strictly positive.

  Thus, $\lambda \geq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(1-p)(1-a)h-pa\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

- For $\ell = \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} \geq \ell > \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} < \ell$, both sides are negative. Thus, $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} < \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$.

  Further, $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}}{p(2a-1)}-\omega}{\overline{v}+(1-p)(1-a)h-pa\ell}$.

(III) In case ω takes relatively small values, $\frac{\overline{v}}{p(2a-1)} > \omega$, which is in case neither ability-types' PC binds:

$\lambda[\overline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\}] \geq$

$\underline{v} + [pa + (1-p)(1-a)]h + p(1-a)(h+\ell) - \min\{p(h+\ell), h\} + \frac{\underline{v}-\overline{v}}{p(2a-1)}$

- For $l \geq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = h$. Inserting, into the simplified constraint gives

  $\lambda[\overline{v} + (p-1)ah + p(1-a)\ell] \geq \underline{v} + (p-1)ah + p(1-a)\ell + \frac{\underline{v}-\overline{v}}{p(2a-1)}$

  - For $\ell > \frac{\overline{v}+(p-1)ah}{p(a-1)}$, both sides of the (in)equality condition are strictly positive.

    Thus, $\lambda \geq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(p-1)ah+p(1-a)\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

  - For $\ell = \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}-\overline{v}}{p^2(a-1)(2a-1)} \leq \ell < \frac{\overline{v}+(p-1)ah}{p(a-1)}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.

  - For $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}-\overline{v}}{p^2(a-1)(2a-1)} > \ell$, both sides are negative. Thus, $\frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}-\overline{v}}{p^2(a-1)(2a-1)} > \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$. Further, $\lambda \leq \frac{\underline{v}+(p-1)ah+p(1-a)\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(p-1)ah+p(1-a)\ell}$.

- For $\ell \leq \frac{(1-p)h}{p}$: $\min\{p(h+\ell), h\} = p(h+\ell)$. Inserting, into the simplified constraint gives

$\lambda[\overline{v} + (1-p)(1-a)h - pa\ell] \geq \underline{v} + (1-p)(1-a)h - pa\ell + \frac{\underline{v}-\overline{v}}{p(2a-1)}$

- For $\ell < \frac{\overline{v}+(1-p)(1-a)h}{pa}$, both sides of the (in)equality condition are strictly positive. Thus, $\lambda \geq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(1-p)(1-a)h-pa\ell} \leftrightarrow \lambda > 1$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\ell = \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left side is zero. Thus, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}-\overline{v}}{p^2a(2a-1)} \geq \ell > \frac{\overline{v}+(1-p)(1-a)h}{pa}$, the left (right) side is negative (non-negative). Thus, $\lambda \leq 0$, $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ never dominates.
- For $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}-\overline{v}}{p^2a(2a-1)} < \ell$, both sides are negative. Thus, $\frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}-\overline{v}}{p^2a(2a-1)} < \ell$ is a necessary condition for the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$.Further, $\lambda \leq \frac{\underline{v}+(1-p)(1-a)h-pa\ell+\frac{\underline{v}-\overline{v}}{p(2a-1)}}{\overline{v}+(1-p)(1-a)h-pa\ell}$.

Secondly, it needs to be examined whether the dominant contract is also second-best optimal. As for the operationalization by means of diagnostic accuracy, I compare the conditions, under which the implementation of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is second-best to the necessary conditions for ℓ, under which contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ dominates contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$. While the dominance conditions were derived above, the conditions for the second-best solutions, given the payer offers contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$, are defined as follows:

- Implementing course 1 for both types (area $[v$.i]) is SB optimal only if $\ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa}\right\}$
- Implementing course 2 for both types (area $[v$.ii]) is SB optimal only if $\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)}\right\}$
- Implementing course 3 for both types (area $[v$.v]) is SB optimal only if $\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} \geq \ell \geq \frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa}$

(I) In case ω takes relatively large values, $\omega \geq \frac{\underline{v}}{p(2a-1)}$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} \geq \frac{\underline{v}+(p-1)ah}{p(a-1)} \quad \leftrightarrow$$

$$\underline{v}+(p-1)ah \leq \underline{v}+(p-1)ah \quad \leftrightarrow$$

$$h \geq h$$

- For all $\ell < \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa} \leq \frac{\underline{v}+(1-p)(1-a)h}{pa} \quad \leftrightarrow$$

$$\underline{v}+(1-p)(1-a)h \leq \underline{v}+(1-p)(1-a)h \quad \leftrightarrow$$

$$h \geq h$$

(II) In case ω takes intermediate values, $\frac{\underline{v}}{p(2a-1)} > \omega \geq \frac{\overline{v}}{p(2a-1)}$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} \geq \frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}}{p^2(a-1)(2a-1)} - \frac{\omega}{p(a-1)} \quad \leftrightarrow$$

$$p(2a-1)\underline{v} + p(2a-1)(p-1)ah + \frac{p(2a-1)\underline{v}}{p(2a-1)} - p(2a-1)\omega \leq$$

$$p(2a-1)\underline{v} + p(2a-1)(p-1)ah + \underline{v} - p(2a-1)\omega \quad \leftrightarrow$$

$$h \geq h$$

- For all $\ell < \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(1-p)(1-a)h+max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa} \leq \frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}}{p^2a(2a-1)} - \frac{\omega}{pa} \quad \leftrightarrow$$

$$p(2a-1)\underline{v} + p(2a-1)(1-p)(1-a)h + \frac{p(2a-1)\underline{v}}{p(2a-1)} - p(2a-1)\omega \leq$$

$$p(2a-1)\underline{v} + p(2a-1)(1-p)(1-a)h + \underline{v} - p(2a-1)\omega \quad \leftrightarrow$$

$$h \leq h$$

(III) In case ω takes relatively small values, $\frac{\overline{v}}{p(2a-1)} > \omega$:

- For all $\ell \geq \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(p-1)ah+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{p(a-1)} \geq \frac{\underline{v}+(p-1)ah}{p(a-1)} + \frac{\underline{v}-\overline{v}}{p^2(a-1)(2a-1)} \quad \leftrightarrow$$

$$p(2a-1)\underline{v} + p(2a-1)(p-1)ah + \frac{p(2a-1)\underline{v}}{p(2a-1)} - p(2a-1)\omega \leq$$

$$p(2a-1)\underline{v} + p(2a-1)(p-1)ah + \underline{v} - \overline{v} \quad \leftrightarrow$$

$$p(2a-1)\omega \geq \overline{v} \quad \leftrightarrow$$

$$\omega \geq \frac{\overline{v}}{p(2a-1)}$$

- For all $\ell < \frac{(1-p)h}{p}$, the SB constraint for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is not stronger than the dominance constraint on ℓ if:

$$\frac{\underline{v}+(1-p)(1-a)h+\max\left\{\frac{\underline{v}}{p(2a-1)}-\omega,0\right\}}{pa} \leq \frac{\underline{v}+(1-p)(1-a)h}{pa} + \frac{\underline{v}-\overline{v}}{p^2a(2a-1)} \quad \leftrightarrow$$

$$p(2a-1)\underline{v} + p(2a-1)(1-p)(1-a)h + \frac{p(2a-1)\underline{v}}{p(2a-1)} - p(2a-1)\omega \leq$$

$$p(2a-1)\underline{v} + p(2a-1)(1-p)(1-a)h + \underline{v} - \overline{v} \quad \leftrightarrow$$

$$p(2a-1)\omega \geq \overline{v} \quad \leftrightarrow$$

$$\omega \geq \frac{\overline{v}}{p(2a-1)}$$

→ In case ω takes (I) relatively large or (II) intermediate values, the dominance of contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ over $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ implies that $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ is also always SB. For parameter constellations, for which contract $(\overline{B}^{*,v}, \overline{A}^{*,v}, \overline{D}^{*,v})$ is superior, the second-best solutions are defined by:

- Implementing course 1 for both types (area [v.i] in Figure 5-7) is SB optimal only if

$$\ell \leq \min\left\{\frac{(1-p)h}{p}, \frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(1-p)(1-a)h]}{\lambda pa}\right\}$$

- Implementing course 2 for both types (area [v.ii]) is SB optimal only if

$$\ell \geq \max\left\{\frac{(1-p)h}{p}, \frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(p-1)ah]}{\lambda p(a-1)}\right\}$$

- Implementing course 3 for the $\overline{v}$-types and course 1 or course 2 for the $\underline{v}$-types (areas [v.iii] and [v.iv], respectively), depending on which one yields the lower ESC, is SB optimal only if

$$\frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(p-1)ah]}{\lambda p(a-1)} \geq \ell \geq \frac{\max\left\{\frac{\overline{v}}{p(2a-1)}-\omega,0\right\}+\lambda[\overline{v}+(1-p)(1-a)h]}{\lambda pa}$$

→ In case ω takes (III) relatively small values, the SB conditions for contract $\left(\underline{B}^{*,v}, \underline{A}^{*,v}, \underline{D}^{*,v}\right)$ are always stronger than the necessary dominance constraint on ℓ. Hence, when determining the second-best solutions, it always needs to be verified whether the dominating contract is indeed second-best.

Summary

Summary

In view of the continuously rising number of magnetic resonance imaging (MRI) scans that are provided across OECD countries and the associated increase in expenditures, this dissertation generates new insights on the provision of (advanced) diagnostic tests and their contribution to improved treatment choice. With the help of microeconomic models, the dissertation examines optimal financial incentives for physicians' diagnostic (testing) and treatment choice and highlights the challenges, which regulators face in creating incentive-compatible payment schemes. Furthermore, it presents empirical evidence on whether MRI scans of the female breast are provided in adherence to guideline-based reimbursement conditions in Germany and analyzes the effects on cost and outcomes.

In particular, a first study examines optimal financial incentives for physician's sequential diagnostic testing and treatment choice. Based on a principal-agent model, the study analyzes the use of expensive advanced diagnostic tests, such as MRI scans of the back, to improve cost-effective treatment choice. A first-best analysis reveals that it can be optimal to use an advanced diagnostic test only conditional on the outcome of a less expensive, yet imperfect, pretest. Then, contracts, which implement optimal sequential diagnostic testing and treatment choice given a situation with double moral hazard from hidden action and hidden information, are determined. Results show that a mixed payment contract, consisting of a capitation payment, conditional fee-for-service (FFS) payments and a fee for the utilization of the advanced diagnostic test, provides optimal incentives for the physician. Yet, implementing the optimal diagnosis and treatment course can create information rents for the physicians, which may yield alternative diagnosis and treatment courses superior and eventually leads to a rent-efficiency trade-off. By contrast, a less restrictive contract, comprising only a capitation payment and unconditional FFS payments, as common in numerous OECD countries, is not incentive compatible. This important finding indicates that existing payment contracts do not seem to put sufficient responsibility for ordering expensive diagnostic tests on the attending physician.

A second study investigates the provision of MRI scans of the female breast in Germany. The Federal Joint Committee (G-BA) enacts under which conditions outpatient MRI scans of the breast for breast cancer treatment are reimbursed by the social health insurance and how providers are paid for the service - which is FFS in this case. However, considering the prevailing non-RCT standard evidence on the effect of these MRI scans and the observation

that, inter alia, due to asymmetric information, payment schemes may provide suboptimal incentives, providers may disregard the reimbursement conditions. In a first step, using administrative data from Techniker Krankenkasse, the study investigates whether outpatient MRI scans of the female breast are provided non-adherent to the reimbursement conditions. In a second step, it examines the effect of the non-adherent provision on direct medical cost and patient outcomes. To allow for causal interpretation, propensity score matching and a difference-in-difference approach are applied. Results suggest that more than one third of all breast MRI scans that were performed in 2016 may be classified as non-adherent to reimbursement conditions. Women who received a non-adherent MRI scan on average cause significantly higher direct medical cost in the follow-up period than women in a control group, yet, they also benefit from increased overall survival. Hence, considering cost and outcomes, it is not straightforward whether the overprovision of MRI scans is indeed ineffective and should be targeted by regulators. In brief, the study concludes that further research, essentially based on data from complementary sources such as clinical data and cause of death statistics, should be conducted to support regulators in making informed decisions on the enforcement of the reimbursement conditions.

A third study examines payment contracts to promote optimal diagnostic and treatment choice among physicians with heterogeneous diagnostic ability. In the principal-agent framework, heterogeneities in diagnostic ability pose a special challenge to regulators. Considering a treatment choice model with endogenous diagnostic effort, the study examines how a public insurer should optimally contract with providers with heterogeneous ability. Hence, a situation with simultaneous double moral hazard from hidden action and hidden information as well as adverse selection from hidden knowledge is assumed. A first-best analysis reveals that it may turn out suboptimal to incentivize all ability-types to exert costly diagnostic effort. While single policy mixed payment contracts, consisting of a capitation payment and conditional FFS payments, are found to promote optimal incentives, they also entail costly information rent payments to the physicians. Therefore, in the second-best situation, particularly depending on the share of the high-ability types in the market, the parameter constellations for which all ability-types should be incentivized are further restricted. In contrast to standard principal-agent theory, a menu of policies that aims at inducing self-selection among the physicians does not improve on this outcome. Although medical expert associations frequently campaign for the special importance of a correct diagnosis to effective health care and technological innovations allow for an increasing number of diagnostic tests, considering the social cost of care provision

(given heterogeneous physician ability) may paint a more differentiated picture and yield the provision of these tests suboptimal.

Overall, this dissertation identifies limitations of existing payment schemes and regulatory instruments. However, it also highlights the challenges, which regulators face when creating optimal incentives for physicians' diagnostic and treatment choice and evaluating existing policies. Essentially, these insights could be used to encourage further, targeted theoretical and empirical research on the interaction of physicians' diagnostic and treatment choice and to possibly adjust payment schemes and regulations.

Zusammenfassung

In den vergangenen Jahren ist die Anzahl an erbrachten Magnetresonanztomographie (MRT)-Untersuchungen in Deutschland, und auch in anderen OECD-Ländern, stetig angestiegen und mit steigenden Gesundheitsausgaben einhergegangen. Vor diesem Hintergrund beschäftigt sich die vorliegende Dissertation mit dem Einsatz von (moderner) Diagnostik und deren Beitrag zu einer effektiven Therapiewahl des Arztes. Mit Hilfe mikroökonomischer Modelle untersucht die Dissertation optimale finanzielle Anreize für den Einsatz von modernen diagnostischen Tests und der anschließenden Therapiewahl eines Arztes und zeigt die Herausforderungen auf, denen sich Sachverwalter bei der Schaffung anreizkompatibler Honorierungssysteme gegenübersehen. Darüber hinaus wird mit Hilfe empirischer Analysen untersucht, inwiefern MRT-Untersuchungen der Mamma im Rahmen der Brustkrebsbehandlung in Deutschland unter Einhaltung leitliniengerechter Richtlinien erbracht werden und anschließend die Auswirkungen auf die medizinischen Kosten und auf Patientenresultate analysiert.

Im Rahmen einer ersten Studie werden optimale finanzielle Anreize für die aufeinanderfolgende diagnostische Test- und Therapiewahl eines Arztes untersucht. Basierend auf einem Prinzipal-Agenten-Modell, analysiert die Studie den Einsatz teurer, moderner diagnostischer Tests, wie beispielweise MRT-Untersuchungen des Rückens, zur Förderung einer kosteneffektiven Therapiewahl des Arztes. Eine first-best Analyse zeigt, dass es effizient sein kann, einen teuren modernen diagnostischen Test nur unter der Bedingung einzusetzen, dass das Ergebnis eines kostengünstigeren, aber weniger präzisem konventionellen Vortests vorliegt. Anschließend werden Honorierungssysteme hergeleitet, die vor dem Hintergrund einer Situation mit doppeltem „Moral Hazard", resultierend aus den nicht-beobachtbaren Handlungen und dem nicht-beobachtbaren Informationsstand des Arztes, zu einer effizienten sequenziellen diagnostischen Test- und Therapiewahl des Arztes führen. Die Ergebnisse

zeigen, dass ein gemischtes Honorierungssystem, bestehend aus einer Kopfpauschale, bedingten Einzelleistungszahlungen und einer Gebühr für die Nutzung des modernen diagnostischen Tests, optimale Anreize für den Arzt schafft. Die Implementierung des optimalen diagnostischen Test- und Therapiepfads kann jedoch Informationsrentenzahlungen an die Ärzte erforderlich machen. Dies führt wiederum dazu, dass alternative diagnostische Test- und Therapiepfade effizienter werden und schließlich ein Abwägen zwischen Informationsrentenzahlungen an den Arzt und allokativer Effizienz erfolgen muss. Hingegen ist ein weniger restriktives Honorierungssystem, welches ausschließlich eine Kopfpauschale und unbedingte Einzelleistungszahlungen umfasst, wie es auch in zahlreichen OECD-Ländern üblich ist, nicht anreizkompatibel. Dieses wichtige Ergebnis deutet darauf hin, dass bestehende Honorierungssysteme den behandelnden Ärzten offenbar nicht ausreichend Kostenverantwortung für die Anordnung moderner diagnostischer Tests übertragen.

Eine zweite Studie untersucht den Einsatz von MRT-Untersuchungen der Mamma in Deutschland. Der Gemeinsame Bundesausschuss (G-BA) beschließt in Form von Richtlinien, für welche medizinischen Indikationen ambulante MRT-Untersuchungen der Mamma im Rahmen der Brustkrebsbehandlung zu Lasten der gesetzlichen Krankenversicherung (GKV) erbracht werden dürfen (und damit erstattungsfähig sind) und wie Radiologen für die Erbringung der Leistung vergütet werden - in diesem Fall über eine separate Gebührenordnungsposition. Aufgrund der unklaren Evidenzlage zur Effektivität dieser MRT-Untersuchungen (es sind bisher keine RCT-Studien vorhanden) und der Tatsache, dass die Untersuchung als Einzelleistung vergütet wird, aber keine Prüfverfahren stattfinden, haben Radiologen einen Anreiz, MRTs zu Lasten der GKV abzurechnen, die eigentlich nicht erstattungsfähig sind. Mit Hilfe von Routinedaten der Techniker Krankenkasse wird in einem ersten Schritt untersucht, inwiefern nicht-erstattungsfähige ambulante MRT-Untersuchungen der Mamma erbracht werden. In einem zweiten Schritt wird untersucht, welchen Effekt die Erbringung dieser nicht-erstattungsfähigen MRTs auf die direkten medizinischen Kosten der Versorgung und auf Patientenresultate hat. Um eine kausale Interpretation zu ermöglichen, werden Propensity Score Matching und ein Differenzen-in-Differenzen-Ansatz angewendet. Die Ergebnisse deuten darauf hin, dass mehr als ein Drittel aller MRT-Untersuchungen der Mamma, die im Jahr 2016 abgerechnet wurden, nicht erstattungsfähig waren und damit hätten nicht zu Lasten der GKV abgerechnet werden dürfen. Frauen, die eine nicht-erstattungsfähige MRT-Untersuchung erhielten, verursachten im Durchschnitt signifikant höhere direkte medizinische Kosten in der Nachbeobachtungszeit als Frauen in einer Kontrollgruppe. Sie profitierten aber gleichzeitig von einem längeren Gesamtüberleben. In Anbetracht der

Auswirkungen auf die Kosten der Versorgung und die Patientenresultate lässt sich daher nicht eindeutig feststellen, ob die Überversorgung mit MRT-Untersuchungen der Mamma tatsächlich ineffektiv ist und das Problem damit von Sachverwaltern aufgegriffen werden sollte. Die Studie kommt zu dem Schluss, dass weitere Analysen, insbesondere basierend auf Daten aus ergänzenden Quellen wie klinischen Studien oder Todesursachenstatistiken, durchgeführt werden sollten, um zu evaluieren, ob die Erbringung der Untersuchungen gemäß den Richtlinien des (G-BA) durchgesetzt werden sollte.

Eine dritte Studie untersucht, wie ärztliche Honorierungssysteme idealerweise ausgestaltet werden sollten, um eine effiziente Diagnose- und Therapiewahl unter Ärzten mit heterogenen diagnostischen Fähigkeiten zu bewirken. In dem vorliegenden Prinzipal-Agenten-Modell können Ärzte (die Agenten) kostenintensive, nicht-beobachtbare Anstrengungen unternehmen, um einen Patienten zu diagnostizieren und anschließend zwischen zwei Therapieoptionen wählen. Es wird eine Situation mit doppeltem „Moral Hazard“ und adverser Selektion angenommen, wobei Letztere daraus resultiert, dass der Krankenversicherer (der Prinzipal) die diagnostischen Fähigkeiten der Ärzte nicht beobachten kann. Eine first-best Analyse zeigt, dass es sich als ineffizient erweisen kann, sowohl für Ärzte mit guten diagnostischen Fähigkeiten als auch für Arzte mit schlechten diagnostischen Fähigkeiten Anreize für diagnostische Anstrengungen zu setzen. Gemischte Honorierungssysteme, bestehend aus einer Kopfpauschale und bedingten Einzelleistungsvergütungen, führen zwar zu einer optimalen Allokation der Anreize, können aber auch Informationsrentenzahlungen an die Ärzte mit sich bringen. Eine second-best Analyse weist darauf hin, dass, insbesondere in Abhängigkeit des Anteils an Ärzten mit guten diagnostischen Fähigkeiten, die Parameterkonstellationen, unter denen für sämtliche Ärzte Anreize für diagnostische Anstrengung gesetzt werden sollten, weiter eingeschränkt werden. Entgegen den Erkenntnissen der klassischen Prinzipal-Agenten-Theorie, führt die Bereitstellung einer Auswahl an Honorierungssystemen, welche darauf abzielt, dass Ärzte basierend auf ihren Präferenzen das System selektieren, was für sie gedacht ist, nicht zu einer effizienteren Versorgung. Obgleich medizinische Fachgesellschaften die Relevanz der korrekten Diagnosestellung für eine effektive Gesundheitsversorgung betonen und der medizinisch-technische Fortschritt eine zunehmende Auswahl an diagnostischen Möglichkeiten hervorbringt, kann sich die Bereitstellung in Anbetracht der sozialen Kosten der Versorgung als suboptimal erweisen.

Abschließend lässt sich festhalten, dass in dieser Dissertation Schwächen bestehender Honorierungssysteme und Regulierungsinstrumente identifiziert werden, aber auch Herausforderungen aufgezeigt werden, denen sich Sachverwalter bei der Bestimmung

optimaler Anreize und der Evaluierung von Maßnahmen stellen müssen. Diese Erkenntnisse können genutzt werden, um weitere, zielgerichtete theoretische und empirische Forschung zum Zusammenspiel von diagnostischer Test- und Therapiewahl eines Arztes zu fördern und gegebenenfalls eine Anpassung der Honorierungssysteme und Regulierungsansätze zu erwägen.

List of Publications

1) “Optimal Financial Incentives for Physician’s Sequential Diagnostic Testing and Treatment Choice” (with Mathias Kifmann), *Working Paper*, 2021

2) “The Provision of Breast MRIs in Adherence to Guideline-Based Reimbursement Conditions and the Impact on Cost and Outcomes: Evidence from Germany” (with Dirk Horenkamp-Sonntag, Udo Schneider and Mathias Kifmann), *Working Paper*, 2021

3) “Contracts to Promote Optimal Diagnostic and Treatment Choice among Physicians with Heterogeneous Diagnostic Ability”, Working Paper, 2021