

TH Köln
University of Applied Sciences

Fakultät für Wirtschafts- und Rechtswissenschaften

Masterarbeit / Thesis (Drei-Monats-Arbeit)
zur Erlangung
des akademischen Grades Master of Science
im Studiengang Versicherungswesen

„Evaluation eines Gesundheitsprogramms in der
Krankenversicherung mittels Propensity-Score-Matching“

Erstprüfer	Prof. Dr. Jan-Philipp Schmidt
Zweitprüfer	Prof. Dr. Jürgen Strobel
vorgelegt am	30. August 2018
von	Fabienne Sebralla
aus	Luxemburger Str. 316 50937 Köln
Matrikel-Nr.	11118158
Telefon-Nr.	0151-64602966
E-Mail-Adresse	fabiennesebralla@t-online.de

Inhaltsverzeichnis

Abbildungsverzeichnis	I
Tabellenverzeichnis	I
Abkürzungsverzeichnis	III
1 Einleitung	1
2 Gesundheitsprogramme in der Krankenversicherung	3
2.1 Krankheitsbild Depression.....	3
2.2 Disease Management Program (DMP)	5
2.3 Propensity-Score-Matching	7
2.4 Datensätze	11
3 Deskriptive Analyse	13
3.1 Allgemeine Analyse des Datensatzes	13
3.2 Analyse der Leistungsausgaben	16
3.3 Analyse der Leistungsausgaben für Depressionen.....	19
4 Propensity-Score: Basis-Modell	22
4.1 Analyse der Matching-Qualität	22
4.1.1 Auswahl der Matching-Parameter	22
4.1.2 Bewertung der Matching-Qualität.....	26
4.1.2.1 Bewertung anhand von Mittelwerten.....	26
4.1.2.2 Bewertung anhand von Quantil-Quantil-Diagrammen	32
4.2 Analyse der Programm-Qualität	39
5 Propensity-Score: Alternativ-Modell 1	47
6 Propensity-Score: Alternativ-Modell 2	58
6.1 Analyse der Matching-Qualität	58
6.2 Analyse der Programm-Qualität	61
7. Fazit	64
Anhang	67
Literaturverzeichnis	74
Ehrenwörtliche Erklärung	77

Abbildungsverzeichnis

Abb. 1: Verlaufsformen depressiver Erkrankungen.....	4
Abb. 2: DMP im Zeitverlauf	6
Abb. 3: Funktionsweise PS-Matching.....	10
Abb. 4: Programmabbruch nach Jahren	14
Abb. 5: Altersverteilung der DMP-Teilnehmer	15
Abb. 6: Verteilung der Rechnungsbeträge auf Leistungsbereiche	17
Abb. 7: Durchschnittliche Leistungsausgaben der DMP Teilnehmer	18
Abb. 8: Absolute Leistungsausgaben: Angeschriebene Personen	19
Abb. 9: Absolute Leistungsausgaben für F-Diagnosen: Angeschriebene Personen.....	20
Abb. 10: Absolute Leistungsausgaben für F-Diagnosen: Teilnehmer und Nicht-Teilnehmer.....	20
Abb. 11: QQ-Plots Matching-Qualität: Letzte F-Diagnose Ambulant	33
Abb. 12: QQ-Plots Matching-Qualität: Letzte F-Diagnose Stationär	34
Abb. 13: QQ-Plots Matching-Qualität: Leistungen Ambulant	35
Abb. 14: QQ-Plots Matching-Qualität: Leistungen Stationär	36
Abb. 15: QQ-Plots Matching-Qualität: Leistungen F-Diagnosen Stationär	37
Abb. 16: QQ-Plots Matching-Qualität: Anzahl F-Diagnosen Stationär	38
Abb. 17: QQ-Plots Programm-Qualität: Leistungen Stationär.....	40
Abb. 18: QQ-Plots Programm-Qualität: Leistungen Ambulant	41
Abb. 19: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär	43
Abb. 20: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant	44
Abb. 21: QQ-Plots Programm-Qualität: Leistungen Stationär - Alternative 1.....	50
Abb. 22: QQ-Plots Programm-Qualität: Leistungen Ambulant - Alternative 1	51
Abb. 23: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär - Alternative 1	55
Abb. 24: QQ-Plots Matching-Qualität: Leistungen Stationär - Alternative 2.....	60
Abb. 25: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär - Alternative 2.....	62
Abb. 26: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant - Alternative 2	63
Abb. 27: Verteilung Versanddatum der Teilnehmer	67
Abb. 28: Verteilung Anmeldedatum der Teilnehmer	67
Abb. 29: Altersverteilung aller Personen	68
Abb. 30: Durchschnittliche Leistungsausgaben der DMP Nicht-Teilnehmer.....	68
Abb. 31: Durchschnittliche stationäre Leistungsausgaben der DMP Teilnehmer	69
Abb. 32: Durchschnittliche stationäre Leistungsausgaben der DMP Nicht-Teilnehmer	69
Abb. 33: Absolute Leistungsausgaben der DMP-Teilnehmer	70
Abb. 34: Absolute Leistungsausgaben der DMP Nicht-Teilnehmer	70
Abb. 35: Verteilung der Rechnungsbeträge für F-Diagnosen auf Leistungsbereiche	71
Abb. 36: Jitter-Plot und Histogramme des Basis-Modells	73
Abb. 37: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant - Alternative 1	73

Tabellenverzeichnis

Tab. 1: Verteilung Versandjahr und Anmeldejahr DMP	13
Tab. 2: Anmeldung DMP-Teilnehmer nach Jahren	15
Tab. 3: Merkmale für Zwillingsbildung	23
Tab. 4: Beispiel stationäre Kosten.....	25
Tab. 5: Differenz der Mittelwert vor- und nach Matching – Basis-Modell (1).....	27
Tab. 6: Differenz der Mittelwert vor- und nach Matching – Basis-Modell (2).....	31
Tab. 7: Leistungsausgaben nach Anschreiben (Mittelwerte)	39
Tab. 8: Leistungsausgaben F-Diagnosen nach Anschreiben (Mittelwerte).....	42
Tab. 9: Anzahl F-Diagnosen nach Anschreiben (Mittelwerte).....	45
Tab. 10: Leistungsausgaben nach Anschreiben (Mittelwerte) - Alternative 1	48
Tab. 11: Leistungsausgaben F-Diagnosen nach Anschreiben (Mittelwerte) - Alternative 1.....	52
Tab. 12: Anzahl F-Diagnosen nach Anschreiben (Mittelwerte) - Alternative 1	56
Tab. 13: Differenz der Mittelwert vor- und nach Matching – Alternative 2.....	58
Tab. 14: Beurteilung der Programm-Qualität - Alternativ-Modell 2	61
Tab. 15: Mittelwerte vor Matching – Basis-Modell.....	71
Tab. 16: Mittelwerte nach Matching– Basis-Modell (1)	72
Tab. 17: Mittelwerte nach Matching– Basis-Modell (2)	72

Abkürzungsverzeichnis

Abb.	Abbildung
Bzw.	Beziehungsweise
D. h.	Das heißt
Diff.	Differenz
DMP	Disease-Management-Programme
Durch.	Durchschnittlich
E. V.	Eingetragener Verein
Etc.	Et cetera
F-Diagnosen	Spezielle Depressionsdiagnosen anhand derer Personen für das Gesundheitsprogramm eingeladen wurden
F.	Folgende
Ff.	Fortfolgende
Ggf.	Gegebenenfalls
ICD-Code	International Statistical Classification of Diseases and Related Health Problems
Proz.	Prozentual
PS	Propensity-Score
QQ-Plot	Quantil-Quantil-Diagramm
S.	Seite
Tab.	Tabelle
TN	Teilnehmer
U. a.	Unter anderem
Vgl.	Vergleiche
Z. B.	Zum Beispiel

1 Einleitung

Die steigende Lebenserwartung in der Bevölkerung sowie der medizinische Fortschritt führen in der privaten Krankenversicherung in Deutschland zu einem deutlichen Kostenanstieg. Im Jahr 2017 summierten sich die Leistungen auf 27 Milliarden Euro.¹ Zu den häufigsten Erkrankungen zählen Depressionen, mit rund 4,1 Millionen betroffenen Personen.² Ein Ansatz, um die Kosten zu reduzieren und zeitgleich die medizinische Behandlung der Versicherten zu verbessern, ist die Entwicklung von strukturierten Gesundheitsprogrammen, sogenannter Disease Management Programme (DMP). Die Teilnahme an einem solchen Programm soll eine bedarfsgerechte und wirtschaftliche Versorgung sicherstellen, die dazu führt, dass sich der Gesundheitszustand des Betroffenen verbessert und infolge u. a. Krankenhausaufenthalte und Folgeschäden vermieden werden, sodass die Gesamtbehandlungskosten für den Krankenversicherer in Zukunft sinken.³ Zur Überprüfung der Zielerreichung wird eine regelmäßige Evaluation der Qualität sowie der Wirtschaftlichkeit eines Gesundheitsprogramms durchgeführt.

Üblicherweise wird der Erfolg von Therapien in der medizinischen Forschung mittels randomisierter kontrollierter Studien geprüft.⁴ Im Rahmen dessen werden zwei Gruppen gebildet, die sogenannte Interventions- und Kontrollgruppe. Personen der Interventionsgruppe erhalten eine bestimmte Therapie, wohingegen Patienten der Kontrollgruppe diese Therapie nicht erhalten. Die Patienten werden zufällig in beiden Gruppen eingeteilt, sodass bekannte und unbekannte Patientenmerkmale, wie z. B. das Alter oder der Krankheitszustand gleichmäßig auf beide Gruppen verteilt sind. Da es sich bei den Teilnehmern eines DMP nicht um eine Zufallsstichprobe aus der Grundgesamtheit aller möglichen Versicherten mit einer bestimmten Erkrankung handelt, kann eine solche Studie nicht bei der Bewertung von DMPs verwendet werden.

Zur Lösung dieser Problematik wurden verschiedene statistische Verfahren entwickelt. Eine Möglichkeit, um den Erfolg eines DMPs zu messen, stellt die Propensity-Score-Methode dar. Zu Beginn der Evaluation wird sichergestellt, dass die Patientenmerkmale in beiden Gruppen nahezu identisch verteilt sind. Dafür wird für jede Person (sowohl Teilnehmer als auch Nicht-Teilnehmer) anhand der bekannten Personenmerkmale ein sogenannter Propensity-Score berechnet und infolgedessen Teilnehmer und Nicht-Teilnehmer mit gleichem Score gepaart. Sofern beide Gruppen nach Zuordnung miteinander vergleichbar sind, ist die Matching-Qualität gut. Darauf folgend wird die Programm-Qualität bewertet: Anhand von verschiedenen Parametern, wie z. B. den Leistungsausgaben oder der Anzahl an Krankenhausaufent-

¹ Vgl. Gesamtverband der Deutschen Versicherungswirtschaft e. V. (2018), S. 17f.

² Vgl. World Health Organization (2017), S. 8 und 19

³ Vgl. Bundesversicherungsamt (2018), Abruf am 14.07.2018

⁴ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 597

halten wird beurteilt, ob das DMP zu einer Verbesserung des Gesundheitszustands und einer Reduktion der Kosten führte.⁵

In dieser Arbeit wird das DMP für Depressionskranke eines deutschen privaten Krankenversicherungsunternehmens betrachtet. Dieses Programm wird hinsichtlich der kurzfristigen Entwicklung der Leistungsausgaben evaluiert. Das Ziel dieser Arbeit ist es, sinnvolle Merkmale zu finden, anhand derer das Propensity-Score-Matching durchgeführt werden kann, und die zu einer guten Matching-Qualität führen. Des Weiteren soll mit Hilfe des Propensity-Score-Matchings der kurzfristige ökonomische Erfolg des DMP für Depressionskranke beurteilt werden.

In Kapitel 2 werden zunächst das Krankheitsbild Depression, das Disease Management Programm sowie das Propensity-Score-Matching vorgestellt und erklärt. Zudem werden die verfügbaren Datensätze erläutert. Anschließend werden in Kapitel 3 die Datensätze deskriptiv analysiert. Besonderer Fokus wird dabei auf die gesamten Leistungsausgaben sowie auf die Leistungsausgaben für Depressionen gelegt. Das Propensity-Score-Matching wird in Kapitel 4 anhand eines Basis-Modells durchgeführt. Es werden sowohl die Matching- als auch die Programm-Qualität beurteilt. Ausgehend vom Basis-Modell werden in Kapitel 5 und 6 zwei Alternativ-Modelle analysiert. Dabei wird zum einen ein anderer Zeitraum und zum anderen eine größere Grundgesamtheit betrachtet. Abschließend werden die Ergebnisse im Fazit zusammengefasst.

⁵ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 601f.

2 Gesundheitsprogramme in der Krankenversicherung

Zu den häufigsten Erkrankungen in Deutschland zählen Depressionen, die in verschiedenen Ausprägungsformen auftreten können. Aufgrund der Vielzahl an Betroffenen fallen in der Krankenversicherung jährlich hohe Leistungsausgaben an. Folglich entwickelte ein privates Krankenversicherungsunternehmen ein Disease Management Programm (DMP) für Depressionserkrankte. Ziel dieses DMP ist die Steuerung der Betroffenen in eine leitliniengerechte Therapie und folglich die Sicherstellung einer bedarfsgerechten und wirtschaftlichen Versorgung. Die anfänglichen Kosten, die durch die Implementierung eines Gesundheitsprogramms entstehen, sollen in Zukunft durch die Senkung der Gesamtbehandlungskosten aufgrund der besseren Versorgung im Rahmen der Teilnahme am DMP ausgeglichen werden.⁶ Die zur Messung des Erfolgs notwendige Evaluation des Programms erfolgt anhand der Propensity-Score-Methode.

2.1 Krankheitsbild Depression

Bei Depressionen handelt es sich um eine mindestens zwei Wochen anhaltende Phase, in welcher der Betroffene im Denken, Fühlen und Handeln beeinträchtigt wird. Auswirkungen von Depressionen sind insbesondere gedrückte Stimmung, Antriebslosigkeit und negative Gedanken.⁷ Hinzu kommen Symptome wie u. a. verminderte Konzentration und Aufmerksamkeit, Gefühle von Schuld und Wertlosigkeit, Schlafstörungen sowie Suizidgedanken und -handlungen.⁸

Depressionen gehören zu den häufigsten Erkrankungen und werden dennoch oftmals unterschätzt. Von den 322 Millionen weltweit an depressiven Störungen leidenden Menschen leben rund 4,1 Millionen Menschen in Deutschland, was 5,2 % der Bevölkerung entspricht.⁹ Im Vergleich zu Männern sind Frauen in der Regel häufiger von Depressionen betroffen und leiden an einer größeren Anzahl von Symptomen.¹⁰ Dies zeigt sich insbesondere daran, dass die jährlichen Krankheitskosten für Depressionen von weiblichen Betroffenen mit 5,8 Milliarden Euro einen Großteil der gesamten Krankheitskosten in Höhe von 8,7 Milliarden Euro ausmachen.¹¹

In der Regel ist der Krankheitsverlauf von Person zu Person unterschiedlich, wobei unipolare Depressionen, bipolare Depressionen und Dysthymie die häufigsten Erkrankungsformen darstellen:

⁶ Vgl. Bundesversicherungsamt (2018), Abruf am 14.07.2018

⁷ Vgl. Stiftung Deutsche Depressionshilfe (2018f), Abruf am 06.07.2018

⁸ Vgl. Stiftung Deutsche Depressionshilfe (2018a), Abruf am 06.07.2018

⁹ Vgl. World Health Organization (2017), S. 8 und 19

¹⁰ Vgl. Stiftung Deutsche Depressionshilfe (2018b), Abruf am 06.07.2018

¹¹ Vgl. Statistisches Bundesamt (Destatis) (2018), Abruf am 07.07.2018

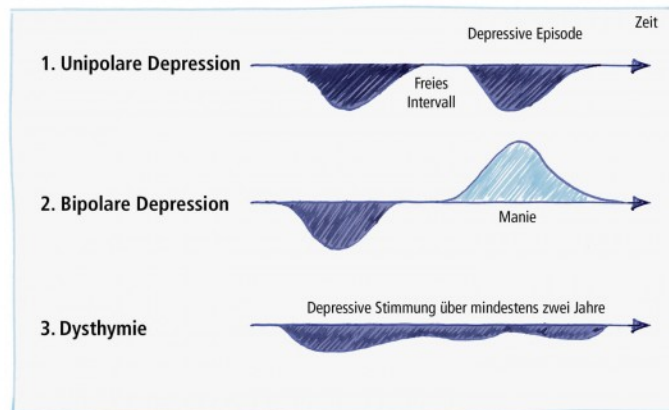


Abb. 1: Verlaufsformen depressiver Erkrankungen¹²

In den meisten Fällen handelt es sich bei depressiven Störungen um unipolare Depressionen. Hierbei leidet der Betroffene an einer depressiven Episode, die zwischen einigen Wochen und mehreren Monaten andauern kann. Nach einer erfolgreichen Behandlung sind die Betroffenen zunächst wieder gesund, sodass die Stimmung weder sehr schlecht noch übermäßig gut ist. Diese Phase wird in der Grafik als „Freies Intervall“ bezeichnet. Häufig kommt es nach einiger Zeit zu einer erneuten depressiven Episode. Das wiederholte Auftreten der depressiven Episode wird als rezidivierende unipolare Depression bezeichnet.

Im Vergleich dazu folgt bei einer bipolaren Depression auf die depressive Episode eine manische Episode, d. h. eine Phase mit übermäßig guter Stimmung. Anschließend verfallen die Betroffenen wieder in eine Depression.

Die dritte Verlaufsform depressiver Erkrankungen unterscheidet sich dahingehend, dass die depressive Episode zwar weniger stark ausgeprägt ist, dafür jedoch mindestens über einen Zeitraum von zwei Jahren anhält.¹³

Depressionen können einerseits medikamentös, d. h. durch die Verschreibung von Antidepressiva (Pharmakotherapie) sowie andererseits psychotherapeutisch behandelt werden. In der Regel erfolgt eine Kombination dieser beiden Behandlungsmöglichkeiten.¹⁴

Durch die Einnahme von Antidepressiva werden nicht die Depressionsursachen, wie u. a. Stress im Beruf bzw. Spannungen in der Partnerschaft, beseitigt. Stattdessen sollen durch die Einnahme der Medikamente die Symptome wie Antriebslosigkeit, Hoffnungslosigkeit und Freundlosigkeit abklingen. Da Antidepressiva nicht unmittelbar helfen, sollten sie über einen längeren Zeitraum und zur Vermeidung eines Rückfalls auch nach Besserung der Depression eingenommen werden.¹⁵

Bei einer Psychotherapie hingegen erfolgt die Behandlung der Depression durch Gespräche und Übungen mit einem Psychotherapeuten.¹⁶

¹² Stiftung Deutsche Depressionshilfe (2018e), Abruf am 06.07.2018

¹³ Vgl. Stiftung Deutsche Depressionshilfe (2018e), Abruf am 06.07.2018

¹⁴ Vgl. Stiftung Deutsche Depressionshilfe (2018g), Abruf am 06.07.2018

¹⁵ Vgl. Stiftung Deutsche Depressionshilfe (2018c), Abruf am 06.07.2018

¹⁶ Vgl. Stiftung Deutsche Depressionshilfe (2018d), Abruf am 06.07.2018

2.2 Disease Management Programm (DMP)

Unter einem Disease Management Programm (DMP) wird ein strukturiertes Behandlungsprogramm für chronisch kranke Menschen, mit dem Ziel der Verbesserung der Versorgung, verstanden. Zudem sollen durch die chronische Krankheit bedingte Folgeschäden und Komplikationen vermieden werden, sodass die Gesamtbehandlungskosten in der Zukunft reduziert werden können. An einem solchen Programm können Versicherte freiwillig teilnehmen, sofern durch einen behandelnden Arzt eine entsprechende Diagnose festgestellt wurde. Der Versicherer trägt die Kosten für das DMP, sodass den Versicherten keine Kosten entstehen. Durch ein effektives DMP kann das Versicherungsunternehmen in Zukunft seine Gesamtkosten reduzieren und folglich die Implementierungskosten des DMP ausgleichen. Daher ist eine regelmäßige Evaluation der Qualität sowie der Wirtschaftlichkeit eines Gesundheitsprogramms notwendig.¹⁷

Für Depressionserkrankte existiert derzeit noch kein etabliertes DMP. Aufgrund der hohen Anzahl an Betroffenen und damit zusammenhängenden Kosten ist die Entwicklung eines solchen Programms jedoch sinnvoll. Daher entwickelte ein privates Krankenversicherungsunternehmen ein DMP für Depressionserkrankte. Ziel des DMP war die Steuerung der Betroffenen in eine leitliniengerechte Therapie. Zum DMP wurden Personen eingeladen, bei denen der Krankenversicherer anhand der Abrechnungsunterlagen erkannte, dass sie an einer bestimmte Depressionsform leiden. Jeder medizinischen Diagnose, die durch einen Arzt oder einen Psychotherapeuten gestellt werden kann, wird ein sogenannter ICD-Code (International Statistical Classification of Diseases and Related Health Problems / Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme) zugeordnet. Dabei handelt es sich um eine von der Weltgesundheitsorganisation eingeführte amtliche Diagnoseklassifikation, die der Verschlüsselung von Diagnosen dient.¹⁸ Sämtliche Krankheiten wurden in 22 thematisch zusammengehörende Kapitel eingeordnet, denen wiederum Code-Bereiche zugeordnet wurden. Kapitel 5 ist untergliedert in den Bereich F00-F99 und umfasst „Psychische und Verhaltensstörungen“.¹⁹ Bei Depressionen handelt es sich um eine Form von affektiven Störungen, denen der ICD-Bereich F30-F39 zugeordnet ist. Aus diesem ICD-Bereich selektierte das Krankenversicherungsunternehmen die folgenden Krankheiten, für die das DMP ausgelegt ist:

- F32 = Depressive Episode:

Bei dieser Diagnose leidet der Betroffene an gedrückter Stimmung, Verminderung von Antrieb, Aktivität, Interesse und Konzentration.

¹⁷ Vgl. Bundesversicherungsamt (2018), Abruf am 14.07.2018

¹⁸ Vgl. Deutsches Institut für Medizinische Dokumentation und Information (2018b), Abruf am 13.07.2018

¹⁹ Vgl. Deutsches Institut für Medizinische Dokumentation und Information (2018a), Abruf am 14.07.2018

- F33 = Rezidivierende depressive Störung:
Charakteristisch für eine F33-Diagnose sind wiederholte depressive Episoden unterschiedlicher Schwere.
- F34.1 = Dysthymia:
Bei einer F34.1-Diagnose handelt es sich um eine chronisch depressive Verstimmung, die mehrere Jahre andauert, jedoch nicht schwer genug ist, um in den Bereich der rezidivierenden depressiven Störungen (F33) eingeordnet zu werden.
- F38 = Andere affektive Störungen:
Sofern bei einem Patienten Stimmungsstörungen diagnostiziert werden, die jedoch aufgrund von Ausprägung und Dauer nicht in eine der anderen Kategorien eingeordnet werden können, werden diese der Restkategorie F38 zugeordnet.²⁰

In Rahmen dieser Arbeit werden die aufgelisteten Diagnosen unter der Bezeichnung „F-Diagnosen“ zusammengefasst.

Das DMP wurde Personen angeboten, bei denen im Vorjahr entweder eine dieser F-Diagnosen stationär oder zwei der F-Diagnosen ambulant festgestellt wurden. Eine Besserung der Depressionserkrankung konnte frühestens nach einem halben Jahr Programmteilnahme erreicht werden.

Im Folgenden wird der Zeitablauf des DMPs dargestellt:

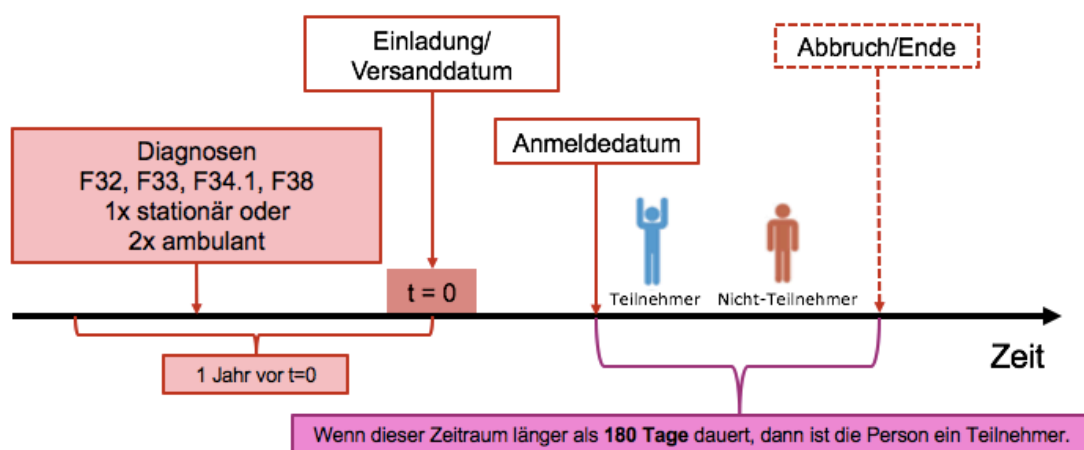


Abb. 2: DMP im Zeitverlauf²¹

Wie bereits erwähnt, werden Personen zum DMP eingeladen, sofern im Vorjahr entweder eine F32, F33, F34.1 bzw. F38 Diagnose stationär oder zwei dieser Diagnosen ambulant festgestellt wurden. Damit zu Evaluationszwecken jeder Person, insbesondere auch den Nicht-Teilnehmern ohne Anmeldedatum, ein eindeutiger Nullpunkt zugeordnet werden kann, wurde das Versanddatum der Einladung zum DMP als $t = 0$ festgelegt. Versicherte, die sich für das Programm anmeldeten und an diesem mehr als 180 Tage teilnahmen, werden im

²⁰ Vgl. Deutsches Institut für Medizinische Dokumentation und Information (2018c), Abruf am 14.07.2018

²¹ Eigene Darstellung

Rahmen der Evaluation als „Teilnehmer“ bezeichnet. Sofern das Programm jedoch im ersten Halbjahr nach Anmeldung abgebrochen wurde, werden diese Personen im weiteren Verlauf nicht weiter in die Auswertung einbezogen, da das Programmziel frühestens nach einem halben Jahr erreicht werden kann. Personen, die eine Einladung erhielten, sich jedoch nicht anmeldeten, werden als „Nicht-Teilnehmer“ klassifiziert.

2.3 Propensity-Score-Matching

Disease Management Programme, die von gesetzlichen Krankenkassen angeboten werden, müssen verpflichtend mindestens im Abstand von drei Jahren evaluiert werden. Nur bei Durchführung einer regelmäßigen Evaluation der strukturierten Behandlungsprogramme ist die Voraussetzung für die Aufrechterhaltung einer DMP-Zulassung gegeben. Eine Bewertung des DMP dient unter anderem dazu, das DMP weiterzuentwickeln. Die Evaluation soll auf Basis allgemein anerkannter wissenschaftlicher Methoden erfolgen.²² Bei Disease Management Programmen, die von privaten Krankenversicherungsunternehmen angeboten werden, ist eine Evaluation nicht vorgeschrieben. Dennoch sollte diese durchgeführt werden, um den Erfolg der Programme zu messen. Dementsprechend soll auch das DMP für Depressionserkrankte evaluiert werden.

Der Erfolg von Therapien wird in der medizinischen Forschung üblicherweise im Rahmen randomisierter kontrollierter Studien geprüft.²³ Dabei wird zu Evaluationszwecken einer Interventionsgruppe eine Kontrollgruppe gegenübergestellt. Der Interventionsgruppe werden die Teilnehmer zugeordnet, die die neue Therapie erhalten. Im Gegensatz dazu wird bei den Personen in der Kontrollgruppe keine bzw. eine andere z. B. die herkömmliche Therapie angewandt. Randomisation bedeutet, dass die Zuteilung der Patienten zu den Therapiegruppen zufällig erfolgt. Jeder Patient hat für jede der Behandlungsmöglichkeiten eine vorgegebene, bekannte Wahrscheinlichkeit. Welche Behandlung er erhält, kann jedoch nicht vorhergesagt werden.²⁴ Dadurch ist garantiert, dass bekannte und unbekannte Patientenmerkmale gleichmäßig auf eine Interventions- und eine Kontrollgruppe verteilt und folglich Aussagen über den Therapieeffekt getroffen werden können. Problematisch an randomisierten kontrollierten Studien ist die fehlende externe Validität. Da es sich bei den Teilnehmern der Studie nicht um eine Zufallsstichprobe aus der Grundgesamtheit aller möglichen Patienten mit der untersuchten Erkrankung handelt, kann das Ergebnis der Studie nicht auf andere Patienten übertragen werden.²⁵

Daher besteht alternativ die Möglichkeit der Durchführung nichtrandomisierter Studien, bei denen die Personen *nicht* zufällig zu den Therapiegruppen zugeteilt werden. Das Problem

²² Vgl. Bundesversicherungsamt (2018), Abruf am 14.07.2018

²³ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 597

²⁴ Vgl. Schumacher, Martin / Schulgen, Gabi (2006), S. 14

²⁵ Vgl. Schumacher, Martin / Schulgen, Gabi (2006), S. 16

dieser Art von Studien ist die fehlende interne Validität. Durch die nicht randomisierte Therapiezuweisung ist es möglich, dass sich die Interventions- und die Kontrollgruppe systematisch im Hinblick auf bekannte und unbekannte Patientenmerkmale unterscheiden. Dadurch lassen sich im Verlauf der Studie ergebende Unterschiede nicht mehr auf die verschiedenen Behandlungen zurückführen.²⁶

Zur Berücksichtigung der Unterschiede zwischen den Gruppen im Rahmen der Auswertung wurden diverse statistische Verfahren entwickelt. Als Standardmethode in nichtrandomisierten Studien gilt das herkömmliche Regressionsmodell, in dem die Auswertung des Therapieeffekts in der Interventions- und der Kontrollgruppe direkt, ohne einen Zwischenschritt, erfolgt, weshalb diese Studienart auch als Ein-Schritt-Verfahren bezeichnet wird. Für die Berechnung des statistischen Modells dient die klinische Zielgröße als abhängige Variable und die Therapie sowie weitere Patientenmerkmale, wie z. B. Alter und Geschlecht, als unabhängige Variablen. Der Einfluss der Therapie auf die Zielgröße wird dabei zeitgleich mit den unabhängigen Variablen geschätzt. Das Ein-Schritt-Verfahren führt jedoch dazu, dass sich beiden Gruppen unter Umständen stark unterscheiden und die Schätzung des Therapieeffekts in diesen Fällen nicht sinnvoll ist.²⁷

Eine Alternative für die statistische Auswertung von nichtrandomisierten Therapiestudien zur Vermeidung interner Validität ist die **Propensity-Score-Methode**. Die Vorgehensweise dieser Methode als Zwei-Schritt-Verfahren ähnelt einer randomisierten kontrollierten Studie: Zunächst soll sichergestellt werden, dass die Patientenmerkmale in beiden Gruppen nahezu identisch sind. Dabei ist zu beachten, dass bei einer nicht-randomisierten Studie nur bekannte und tatsächlich gemessene Patientenmerkmale, wie z. B. Alter, Geschlecht, Leistungsausgaben oder die Medikamentenanzahl, gleichmäßig verteilt werden können, nicht jedoch unbekannte Merkmale. Anschließend kann die Schätzung des Therapieeffekts erfolgen.²⁸

Für die Bildung ähnlicher Gruppen wird zunächst der Propensity-Score (PS) ermittelt. Der PS bezeichnet die Wahrscheinlichkeit, ob ein Patient die zu prüfende Therapie erhält. In einer randomisierten Studie beträgt der PS für jeden Patienten 0,5. In nichtrandomisierten Studien hingegen ist der PS für die einzelnen Patienten unbekannt und bestimmt sich nach dessen individuellen Merkmalen. Daher wird der Propensity-Score für jeden einzelnen Patienten aus den vorliegenden Daten mittels eines logistischen Regressionsmodells (PS-Modell) berechnet. Dabei gilt die zugewiesene Therapie als abhängige Variable und die zu Therapiebeginn bestehenden Patientenmerkmale, wie u. a. Alter, Geschlecht, Leistungsausgaben als unabhängige Variablen. Die unabhängigen Variablen sollten so gewählt werden, dass sie den späteren Therapieerfolg und nicht die Therapiezuweisung vorhersagen. Solche Merkmale ver-

²⁶ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 597

²⁷ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 599-601

²⁸ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 601f.

ringern die Varianz des Behandlungseffekts und erzeugen zudem keinen zusätzlichen Bias.²⁹

Nachdem mittels PS-Modell für jeden Patienten ein PS berechnet wurde, wird dieser PS zur Schätzung des Therapieeffekts, d. h. zur Beurteilung der Programm-Qualität, zur Hilfe genommen. Für die Berücksichtigung des Propensity-Scores können vier verschiedene Methoden verwendet werden, die jeweils sowohl Vor- als auch Nachteile besitzen:

- PS-Matching: 1:1-Matching anhand berechneter PS
- „Inverse probability of treatment weighting“: Zuordnung des Kehrwerts des PS als statistisches Gewicht
- Stratifizierung: Einteilung des Datensatzes in gleich große Teile anhand PS
- Regressionsadjustierung für den PS: Regressionsmodell mit interessierenden Zielgrößen als abhängige Variablen und Therapieeffekt sowie PS als unabhängige Variable.³⁰

Im Rahmen dieser Arbeit wird das PS-Matching verwendet. Diese Methode wird oftmals aufgrund der Möglichkeit zur expliziten Darstellung der erhobenen Merkmale von behandelten und unbehandelten Patienten genutzt. Dadurch ist es möglich, zu prüfen, ob eine Ähnlichkeit in der Verteilung der erhobenen Merkmale zwischen den Patienten besteht.³¹

Beim PS-Matching handelt es sich um ein „1:1-Matching“, bei dem jedem behandelten Patienten ein unbehandelter Patient zugeteilt wird. Üblicherweise erfolgt die Zuteilung anhand identischer PS. Es besteht zudem die Möglichkeit, Grenzen zu definieren, sodass der PS der beiden Patienten minimal innerhalb der definierten Grenzen voneinander abweicht. Sofern für eine Person kein Matching-Partner gefunden werden kann, wird diese Person ausgeschlossen und die Anzahl der Personen in der Gruppe verringert sich, wodurch sich die Varianz vergrößert. Bei Verwendung einer der drei anderen Methoden wird grundsätzlich jeweils der gesamte Datensatz berücksichtigt, jedoch führt die Methode des PS-Matchings dazu, dass die Interventions- und die Kontrollgruppe möglichst ähnlich sind und folglich ein Bias vermieden wird. Die Güte des PS-Modells wird anhand der Balance der Merkmale in den beiden gebildeten Gruppen gemessen. Das bedeutet, dass anhand der Vergleichbarkeit der beiden Gruppen beurteilt wird, wie gut das Modell ist. Nur wenn die Matching-Qualität gut ist, kann im Anschluss die Auswertung erfolgen.³²

²⁹ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 597f

³⁰ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 597f.

³¹ Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 600

³² Vgl. Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016), S. 598ff.



Abb. 3: Funktionsweise PS-Matching³³

Die Abbildung verdeutlicht die Schritte der Propensity-Score-Methode. Zu Beginn der Analyse liegen Daten zu einer Gruppe von Personen vor. Ein Teil dieser Personen wurde mit einer bestimmten Methode behandelt (blau) und ein anderer Teil wurde *nicht* mittels dieser Methode behandelt (rot). Im Fall dieser Arbeit handelt es sich um eine Grundgesamtheit von 12.065 Personen. Von diesen haben sich 678 Personen zum Gesundheitsprogramm angemeldet und wurden nach dieser Methode behandelt (blau = 678), 11.173 Personen hingegen haben sich *nicht* angemeldet und wurden folglich *nicht* nach dieser Methode behandelt (rot = 11.173).

Für jede Person aus der Grundgesamtheit wird zunächst auf Basis der vorhandenen Merkmale mittels eines logistischen Regressionsmodells ein Propensity-Score berechnet. Die Merkmale und die Methode nach denen der PS berechnet wird, wird als PS-Modell bezeichnet. Anschließend erfolgt das PS-Matching, d. h. die Zuordnung von Teilnehmern und Nicht-Teilnehmern mit identischem bzw. ähnlichem PS. Sofern kein geeigneter Matching-Partner gefunden werden kann, wird die Person ausgeschlossen (schwarzes Kreuz). Für die gematchten Paare erfolgt schließlich die Auswertung. Im Rahmen der Auswertung wird die Programm-Qualität anhand der Leistungsausgaben nach Programmteilnahme beurteilt.³⁴

Die Analyse der Daten sowie das Propensity-Score-Matching werden mit Hilfe der Software R durchgeführt. Bei R handelt es sich um eine frei verfügbare Software, in der statistische Berechnungen durchgeführt sowie Grafiken erstellt werden können, wobei „R“ sowohl das Programm als auch die Befehlssprache bezeichnet.³⁵ Die Standardfunktionen von R können durch verschiedene Zusatzkomponenten, sogenannte Pakete, um spezielle Funktionen zur Datenanalyse erweitert werden. Zur Durchführung des Propensity-Score-Matching wird das „Matchit“-Paket benötigt.

³³ Kuss, Oliver / Blettner, Maria / Börgemann, Jochen (2016), S. 600

³⁴ Vgl. Kuss, Oliver / Blettner, Maria / Börgemann, Jochen (2016), S. 599

³⁵ Vgl. The R Foundation (2018), Abruf am 21.07.2018

2.4 Datensätze

Für die Evaluation des DMP stellte das Krankenversicherungsunternehmen insgesamt vier Datensätze mit Daten der zum DMP eingeladenen Personen zur Verfügung. Der Datenabzug erfolgte im März X6 und enthält die Daten aller Personen, die eine Einladung zum Gesundheitsprogramm erhielten und zu diesem Zeitpunkt privat vollversichert waren. Lediglich der Abzug der Diagnosedaten erfolgte zeitversetzt im April X7.

Datei: Stammdaten:

Die Datei „Stammdaten“ besteht aus 12.063 Datensätzen mit 36 Variablen und enthält allgemeine Informationen zu den Personen, wie u. a. das Alter und das Geschlecht, die Versicherungsart (Vollversicherung oder Beihilfe), das Datum des Versicherungsbeginns sowie ggf. das Datum des Versicherungsendes aufgrund von Stornierung oder Tod und den vom Versicherer errechneten Komorbiditätsindex. Als Komorbidität wird vor allem in der Psychologie das Auftreten zusätzlicher Erkrankungen im Rahmen einer definierten Grunderkrankung bezeichnet.³⁶ Das Geburts- sowie das Sterbedatum wurde jeweils auf den 01.01. des jeweiligen Jahres verfremdet.

Datei: Leistungsdaten:

Die Datei „Leistungsdaten“ besteht aus 2.903.358 Datensätzen mit zehn Variablen und umfasst sämtliche von den zum DMP eingeladenen Personen eingereichte Leistungsdaten im Zeitraum von Januar X-1 bis Ende März X6. Neben dem Behandlungsbeginn und -ende werden pro eingereichter Leistung sowohl der Rechnungs- als auch der Leistungsbetrag angegeben. Das Jahr der Behandlung muss dabei nicht zwingend mit dem Jahr, in dem die Rechnung beim Krankenversicherer eingereicht wird, übereinstimmen. Der Rechnungsbetrag umfasst alle Kosten, die im Behandlungszeitraum angefallen sind, wohingegen der Leistungsbetrag nur die Höhe angibt, die vom Krankenversicherer nach Abzug des Selbstbehaltes bzw. von Kürzungen erstattet wird.

Datei: Grunddaten:

Die Datei „Grunddaten“ besteht aus 12.072 Datensätzen mit fünf Variablen. In diesem Datensatz ist vermerkt, wann die Person zum DMP eingeladen wurde sowie ob und wenn ja, wann sie sich zum DMP angemeldet hat. Hinzu kommt die Angabe des Abbruchdatums, sofern das Programm vorzeitig abgebrochen wurde.

Datei: Diagnosedaten:

Die Datei „Diagnosedaten“ besteht aus 5.116.493 Datensätzen mit 22 Variablen und enthält alle Krankheiten, die bei den zum DMP eingeladenen Personen im Zeitraum von Januar X-1

³⁶ Vgl. Stangl, W. (2018), Abruf am 18.06.2018

bis Ende März X7 diagnostiziert wurden. Von den 22 Variablen sind die Informationen zur Diagnose (ICD-Code), zum Behandlungsbeginn und -ende und in welchem Bereich die Diagnose gestellt wurde (ambulante oder stationäre Diagnose bzw. Krankentagegeld), relevant.

Jeder Person wurde eine Identifikationsnummer zugeordnet. Alle Datensätze enthalten eine Spalte mit dieser Identifikationsnummer, sodass anhand derer die einzelnen Datensätze miteinander verknüpft werden können.

3 Deskriptive Analyse

Vor Durchführung der Evaluation des Gesundheitsprogramms werden die Datensätze zunächst deskriptiv analysiert. Dafür werden zu Beginn allgemeine Parameter aus den Daten der Personen, die zum Programm eingeladen wurden, betrachtet. Dazu zählen unter anderem das Geschlecht und Alter, aber auch programmrelevante Informationen, wie z. B. die Anzahl und Ausprägung der Depressionserkrankung. Zudem wird erläutert, wie der Versand der Einladungen, die Anmeldung zum Programm und der Abbruch des Programms verteilt sind. Im Anschluss werden die gesamten Leistungsausgaben sowie die Ausgaben für Depressionskrankheiten im Zeitraum vor Anschreiben analysiert. Die Tabellen und Abbildungen in diesem und in den darauffolgenden Kapiteln basieren auf eigenen Berechnungen.

3.1 Allgemeine Analyse des Datensatzes

Zum Gesundheitsprogramm wurden in der Zeit von Oktober X1 bis April X5 insgesamt 12.065 Versicherte eingeladen. In der Regel sind Frauen zwar häufiger von Depressionen betroffen, dennoch wurden zum DMP mehr männliche als weibliche Personen eingeladen, was darauf zurückzuführen ist, dass die Versichertenstruktur des Krankenversicherers mehr männliche als weibliche Personen aufweist. Insgesamt wurden Einladungen an 6.529 Männer und 5.536 Frauen verschickt. Für das DMP meldeten sich jedoch nur 892 Personen an, was lediglich 7,4 % der eingeladenen Personen entspricht. Die meisten Einladungen wurden im Jahr X3 versandt, wohingegen sich die meisten Personen bereits im Jahr X2 anmeldeten:

Jahr	Einladungen ³⁷	Anmeldungen ³⁸	Prozentual
X1	612	51	8,33 %
X2	3.993	431	10,79 %
X3	6.121	328	5,36 %
X4	1.338	82	6,13 %
X5	1	-	
Summe	12.065	892	7,39 %

Tab. 1: Verteilung Versandjahr und Anmeldejahr DMP

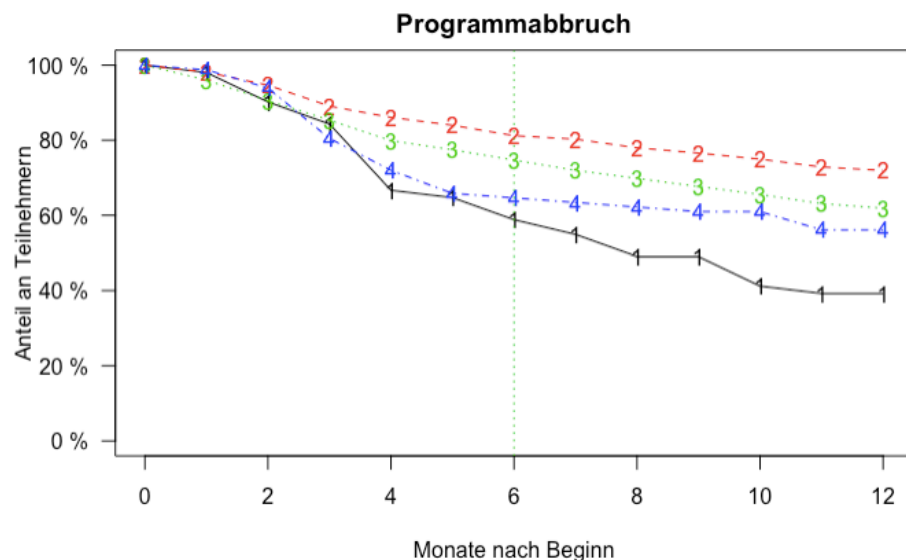
Zu Beginn des DMP, insbesondere in Jahr X2, meldeten sich überdurchschnittlich viele Personen für das Programm an, wohingegen sich die Anzahl der Anmeldungen in den Jahren X3 und X4 trotz vieler Einladungen stark verringerte. Die vergleichsweise hohen Anmeldequoten zum Zeitpunkt der Programmeinführung ergeben sich daraus, dass zunächst Versicherte mit einer langen Krankheitsgeschichte angeschrieben wurden und diese sehr affin für die Teilnahme an einem Gesundheitsprogramm waren.

³⁷ Vgl. Abb. 27: Verteilung Versanddatum der Teilnehmer

³⁸ Vgl. Abb. 28: Verteilung Anmeldedatum der Teilnehmer

Erkennbar ist diese Affinität auch bei Betrachtung der Dauer zwischen Versanddatum der Einladungen und Anmeldedatum zum DMP. Über alle fünf Jahre hinweg lagen zwischen diesen Zeitpunkten im Durchschnitt 42 Tage. Im Jahr X2 hingegen dauerte es durchschnittlich nur 28 Tage bis sich die Personen für das Gesundheitsprogramm anmeldeten.

Ein Erfolg des DMP und damit eine Verbesserung der Depression ist frühestens nach einer sechsmonatigen Teilnahme möglich. Bei vorzeitigem Abbruch ist eine Verbesserung der Krankheit nicht wahrscheinlich. Daher werden Personen, die das Programm innerhalb der ersten sechs Monate abbrechen, in der Auswertung nicht berücksichtigt. Das Gesundheitsprogramm konnte bei diesen Personen nicht wirken, sodass der Einschluss dieser Personen in die Auswertung zu einer Verzerrung führen würde. Von den 892 Personen, die sich insgesamt anmeldeten, beendeten 214 Personen das Programm vorzeitig. Dies entspricht 24 % der angemeldeten Personen. Bei Betrachtung der Abbruchraten in den einzelnen Jahren werden große Unterschiede ersichtlich:



1 = X1 2 = X2 3 = X3 4 = X4

Abb. 4: Programmabbruch nach Jahren

Personen, die sich im Jahr X1 zum Programm anmeldeten, brechen dieses am häufigsten vorzeitig ab, was vermutlich damit zusammenhängt, dass sich in diesem Jahr lediglich 51 Personen anmeldeten und sich das Programm noch im Anfangsstadium befand. Die größte Programmtreue haben Teilnehmer aus dem Jahr X2. Wie bereits erwähnt, handelt es sich hier um Personen, die bereits lange an Depressionen erkrankt und folglich affin für das Gesundheitsprogramm sind. Allgemein ist jedoch zu erkennen, dass die Anzahl an Personen, die das Programm vorzeitig abbrechen, hoch ist. Die Abbruchrate in Höhe von 24 % führte dazu, dass das Programm im Jahr X5 in dieser Form eingestellt und stattdessen in einer modifizierten Form angeboten wurde.

Aufgrund des Ausschlusses der 214 Personen, die das Programm vorzeitig beendeten, werden von den 12.065 eingeladenen Personen nur 678 Personen als DMP-Teilnehmer gewertet. Dies entspricht lediglich 5,6 % der angeschriebenen Personen.

Jahr	DMP-Teilnehmer
X1	30
X2	350
X3	245
X4	53
Summe	678

Tab. 2: Anmeldung DMP-Teilnehmer nach Jahren

Von den 678 Teilnehmern nahmen die meisten im Jahr X2 und X3 am DMP teil. Zu Beginn und kurz vor Einstellung des DMP im Jahr X5 hingegen war die Teilnehmerquote sehr gering.

Zum DMP wurden mehr männliche als weibliche Personen eingeladen. Zudem meldeten sich prozentual ebenfalls mehr Männer als Frauen an, sodass von den 678 DMP-Teilnehmern 382 Personen männlich und 296 Personen weiblich sind.

Die meisten DMP-Teilnehmer sind zwischen 40 und 70 Jahren alt (insgesamt 576 der 678 Personen). Lediglich 40 Personen sind unter 40 Jahren und 62 Personen über 70 Jahren, sodass das Durchschnittsalter über alle Jahre hinweg 56 Jahre beträgt.

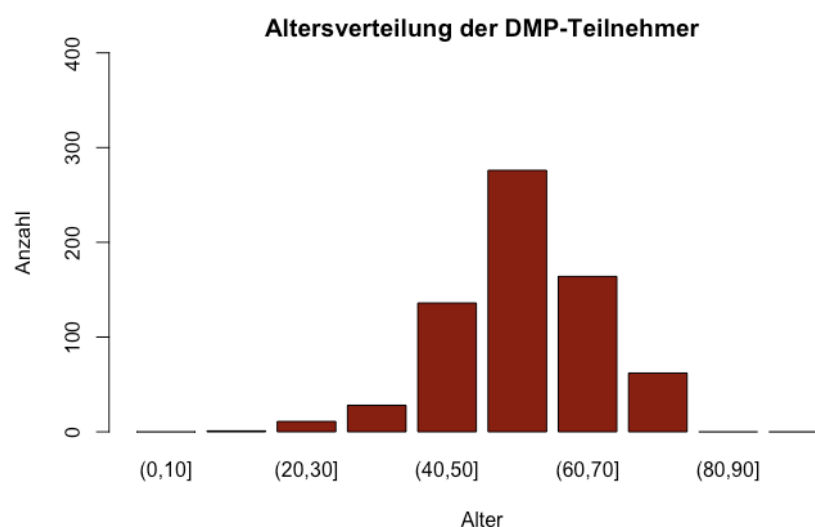


Abb. 5: Altersverteilung der DMP-Teilnehmer

Eine ähnliche Altersverteilung ist ebenfalls bei den 12.065 angeschriebenen Personen erkennbar.³⁹

³⁹ Vgl. Abb. 29: Altersverteilung aller Personen

Aus der Datei „Depressionsdaten“ konnte in 236.286 Fällen eine F-Diagnose einer Rechnung in der Datei „Leistungsdaten“ zugeordnet werden. Dabei waren F-Diagnosen in den Bereichen Stationär, Ambulant, Krankenhaustagegeld, Krankentagegeld, Kur, Pflege und Zahn möglich. Am häufigsten wurden F-Diagnosen ambulant gestellt (rund 180.000 Diagnosen), gefolgt von stationären Feststellungen (rund 19.000 Diagnosen). Bei 9.882 Versicherten wurden im Betrachtungszeitraum ambulant bis zu 20 F-Diagnosen, wohingegen stationär häufig keine bzw. maximal bis zu 10 F-Diagnosen gestellt wurden.

Eine mögliche Behandlungsform für Depressionen ist die Einnahme von Antidepressiva. Die Datenlieferung enthält für jede Person die Anzahl der verschriebenen Antidepressiva im Jahr bevor diese zum DMP eingeladen wurde. Die Hälfte aller angeschriebenen Personen bekam im Betrachtungszeitraum keine Antidepressiva verschrieben. Von den Personen, denen Antidepressiva verordnet wurden, erhielten die meisten Versicherten bis zu fünf dieser Medikamente verschrieben. Eine regelmäßige Verordnung ist ein Zeichen für eine funktionierende Behandlungsform. Bei vereinzelt Personen wurden bis zu 20 Antidepressiva verschrieben, was eher als Zeichen für einen komplizierten Behandlungsverlauf zu deuten ist.

Das Krankenversicherungsunternehmen berechnet für jeden angeschriebenen Versicherten einen Komorbiditätsindex, den sogenannten Charlson-Komorbiditätsindex. Dabei werden insgesamt 19 Nebenerkrankungen, wie u. a. Demenz, Lebererkrankungen und AIDS anhand der Schwere der Krankheit in sechs verschiedene Stufen eingeteilt. Die Summe daraus ergibt anschließend den Komorbiditätsindex.⁴⁰ Bei 80 % der Personen hat der Index den Wert 0 und bei weiteren 10 % den Wert 1. Der Maximalwert liegt bei 23.

3.2 Analyse der Leistungsausgaben

Im Folgenden werden zunächst die Leistungsausgaben für alle Krankheiten betrachtet, die von den zum DMP eingeladenen Personen eingereicht wurden:

Die meisten im Zeitraum von Januar X-1 bis Ende März X6 eingereichten Rechnungsbeträge sowohl für Teilnehmer als auch für Nicht-Teilnehmer fallen im ambulanten Bereich an (Teilnehmer: 54 %; Nicht-Teilnehmer: 58 %), gefolgt von Leistungen für stationäre Behandlungen (Teilnehmer: 35 %; Nicht-Teilnehmer: 30 %) und Zahnbehandlungen (Teilnehmer: 10 %; Nicht-Teilnehmer: 11 %). Die Kosten für Pflege und Kur hingegen sind vernachlässigbar:

⁴⁰ Vgl. Deutsche Krebsgesellschaft e.V. (2018), S. 23

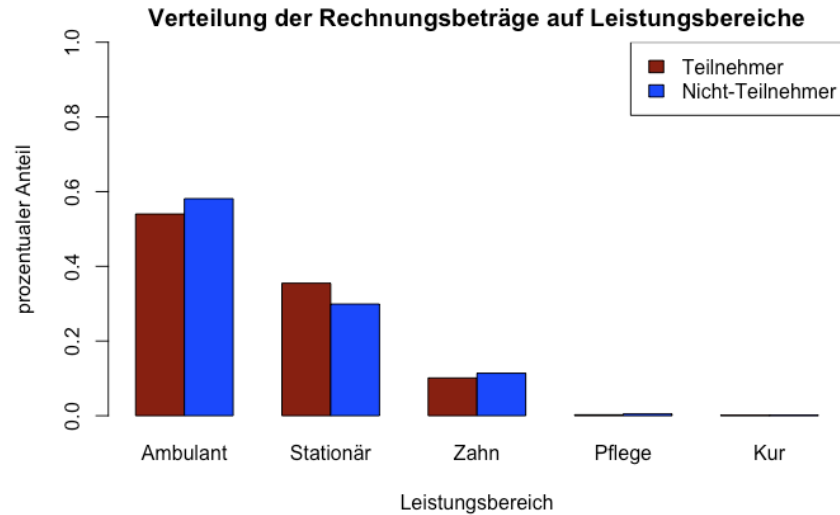


Abb. 6: Verteilung der Rechnungsbeträge auf Leistungsbereiche

Die gesamten Leistungsausgaben der Teilnehmer sowie der Nicht-Teilnehmer sind ähnlich verteilt. Über den Zeitverlauf haben Teilnehmer im Vergleich zu Nicht-Teilnehmern prozentual geringere Kosten im ambulanten Bereich sowie für Zahnbehandlungen. Die stationären Kosten hingegen sind bei den Nicht-Teilnehmern geringer.

Durchschnittliche Leistungsausgaben:

In den 300 Tagen vor Einladung zum DMP steigen die Kosten der Teilnehmer stetig an und jeder Teilnehmer reicht durchschnittlich Rechnungen in Höhe von 10.363 Euro ein. In den 300 Tagen nach Anschreiben hingegen sinkt die Höhe der eingereichten Rechnungen und die durchschnittlichen Teilnehmerkosten sinken auf 9.268 Euro. Der Unterschied im Jahr vor und im Jahr nach Einladung zum DMP in Höhe von 1.094 Euro ist statistisch signifikant. Als Signifikanzniveau wird $\alpha \leq 0,05$ festgelegt. Das Signifikanzniveau bezeichnet die Irrtumswahrscheinlichkeit. Sofern der p-Wert kleiner als das festgelegte Signifikanzniveau ist, ist das Ergebnis statistisch signifikant.⁴¹

Die folgende Abbildung zeigt den Verlauf der durchschnittlichen Rechnungsbeträge der 678 Teilnehmer. Wie erwähnt, ist der Zeitpunkt $t = 0$ jeweils das Anschreibdatum. Die schwarzen Punkte in der Grafik kennzeichnen jeweils die Höhe der Rechnungsbeträge in Zeiträumen von 60 Tagen.

⁴¹ Vgl. Grimmer, Arnd (2014), S. 185

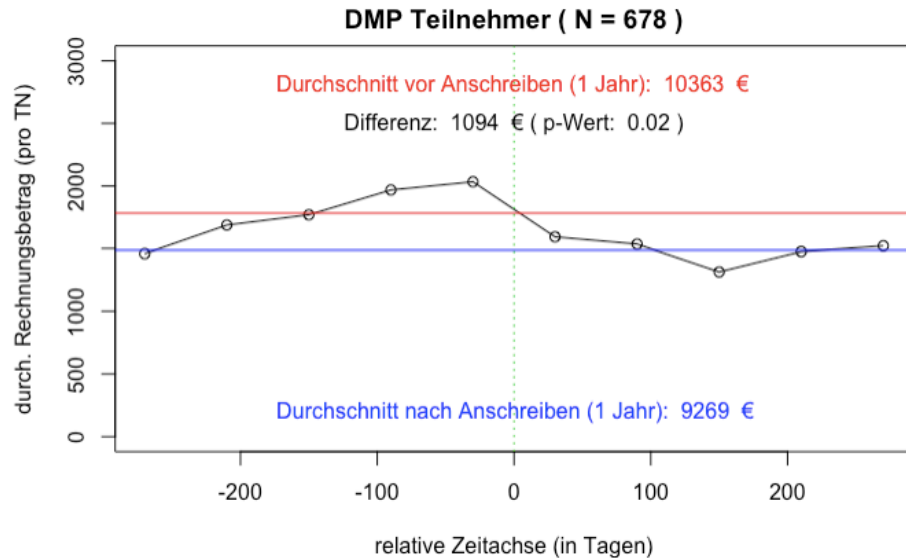


Abb. 7: Durchschnittliche Leistungsausgaben der DMP Teilnehmer

Der gleiche Verlauf ist ebenfalls bei den Nicht-Teilnehmern erkennbar. Die Kosten sinken in dieser Gruppe von durchschnittlich 8.992 Euro vor Einladung auf 7.603 Euro im Jahr nach Einladung zum DMP.⁴² Daher kann aus der Kostensenkung der Teilnehmer im Jahr nach DMP-Teilnahme nicht geschlussfolgert werden, dass das DMP erfolgreich war.

Zum Gesundheitsprogramm werden Personen eingeladen, bei denen im Jahr vor Einladung Depressionen diagnostiziert wurden. Durchschnittlich liegt die letzte Depressionsdiagnose im ambulanten Bereich 148 Tage und im stationären Bereich 94 Tage vor Anschreiben. Dementsprechend werden in dieser Zeit hohe Leistungsausgaben für stationäre Aufenthalte, psychische Behandlungen, Antidepressiva und ähnliches verursacht, weshalb die Kosten im Zeitraum vor Anschreiben steigen. Der Rückgang der Kosten im Jahr nach Einladung kann mithilfe von *Abb. 1: Verlaufsformen depressiver Erkrankungen* erklärt werden. Nach einer depressiven Phase folgt häufig eine gesunde Phase, in der die Person ggf. noch Antidepressiva einnimmt, jedoch keine intensive Behandlung mehr benötigt. Folglich sinken in dieser Phase der Erkrankung die Kosten, solange bis die nächste depressive Phase beginnt.

Absolute Leistungsausgaben:

Insgesamt verursachen die 12.065 Personen in den 300 Tagen bevor sie zum DMP eingeladen werden in allen Leistungsbereichen absolute Kosten in Höhe von 109 Millionen Euro. In den 300 Tagen nach Einladung zum Gesundheitsprogramm hingegen belaufen sich die Kosten lediglich auf 93 Millionen Euro.

⁴² Vgl. Abb. 30: Durchschnittliche Leistungsausgaben der DMP Nicht-Teilnehmer

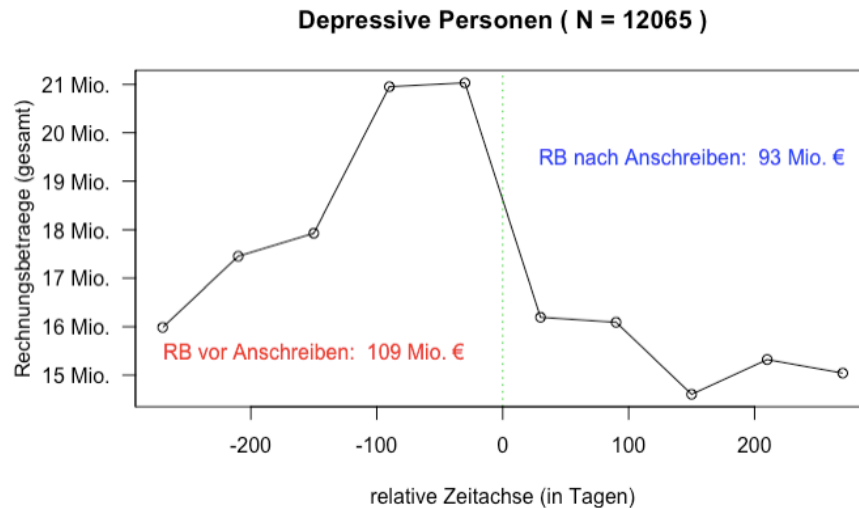


Abb. 8: Absolute Leistungsausgaben: Angeschriebene Personen

Ein ähnlicher Leistungsverlauf ist ebenfalls bei den Teilnehmern sowie bei den Nicht-Teilnehmern erkennbar.⁴³ In den 300 Tagen vor Anschreiben verursachen die 678 Teilnehmer Kosten in Höhe von 7 Millionen Euro, wohingegen in den 300 Tagen nach Einladung 6,3 Millionen Euro Kosten anfallen. Die Kosten der 11.173 Nicht-Teilnehmer hingegen belaufen sich in den 300 Tagen vor Einladung auf 100,5 Millionen Euro und in den 300 Tagen danach auf 84,9 Millionen Euro. Die Summe der Kosten der Teilnehmer und der Kosten der Nicht-Teilnehmer ergibt nicht die Gesamtkosten aller depressiven Personen, da in diesen Gesamtkosten zusätzlich die Kosten der Programmabbrecher enthalten sind.

3.3 Analyse der Leistungsausgaben für Depressionen

Bei 8,1 % der Rechnungsbelege aus der Datei „Leistungsdaten“ (rund 236.000 der 2,9 Millionen Rechnungsbelege) handelt es sich um Rechnungen für F-Diagnosen. Die Verteilung der Leistungsausgaben für Depressionsdiagnosen unterscheidet sich von der Verteilung der gesamten Leistungsausgaben. Für Depressionsdiagnosen fallen mehr Kosten stationär, als ambulant an. In allen weiteren Bereichen (Kur, Krankentagegeld, Pflege etc.) fallen kaum bzw. gar keine Depressionskosten an. Die geringen Kosten im Bereich des Krankentagegelds sind darauf zurückzuführen, dass nur wenige Personen einen Krankentagegeldtarif versichert haben. Die Leistungsausgaben sind bei Teilnehmern und Nicht-Teilnehmern ähnlich verteilt. Nicht-Teilnehmer haben stationär prozentual leicht höhere Kosten als Teilnehmer, dafür jedoch im ambulanten Bereich geringere Kosten.⁴⁴

Insgesamt fallen in allen Bereichen in den 300 Tagen vor Einladung Rechnungsbeträge für Depressionsdiagnosen in Höhe von 15 Millionen Euro an. In den 300 Tagen nach Anschreiben reduzieren sich die Rechnungsbeträge um vier Millionen auf elf Millionen Euro:

⁴³ Vgl. Abb. 33: Absolute Leistungsausgaben der DMP-Teilnehmer und Abb. 34: Absolute Leistungsausgaben der DMP Nicht-Teilnehmer

⁴⁴ Vgl. Abb. 35: Verteilung der Rechnungsbeträge für F-Diagnosen auf Leistungsbereiche

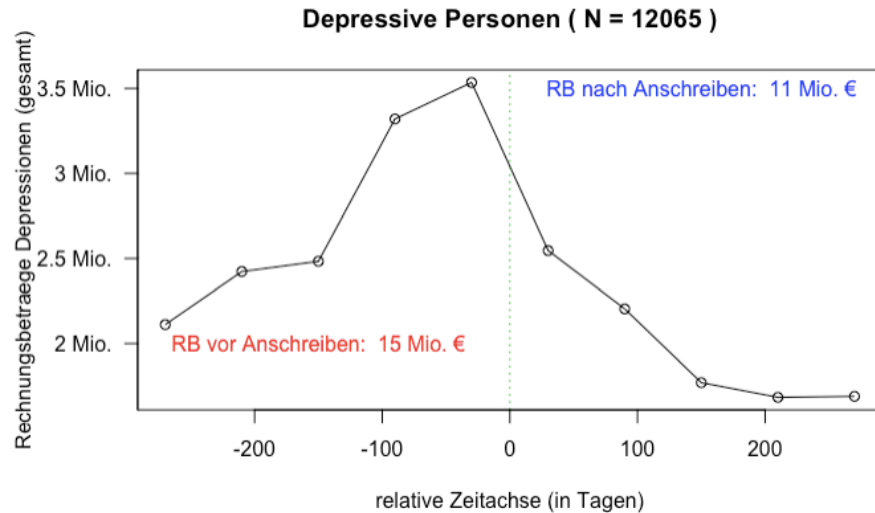


Abb. 9: Absolute Leistungsausgaben für F-Diagnosen: Angeschriebene Personen

Werden lediglich die Rechnungsbeträge der F-Diagnosen betrachtet, so ist der typische Depressionsverlauf bei allen angeschriebenen Personen noch stärker erkennbar, als bei Betrachtung der gesamten Leistungsausgaben.⁴⁵ Zwischen 30 und 60 Tagen nach Einladung zum DMP ist das Absinken des Grafen leicht gemindert. Diese Dämpfung lässt sich mit Hilfe des Kostenverlaufs der DMP Teilnehmer im Jahr nach Anschreiben erklären:

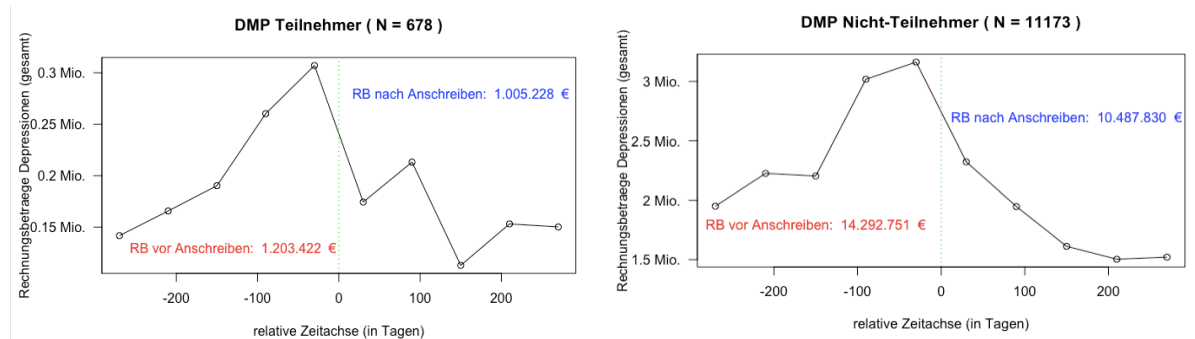


Abb. 10: Absolute Leistungsausgaben für F-Diagnosen: Teilnehmer und Nicht-Teilnehmer

Die Depressionskosten der Teilnehmer fallen zunächst im Monat nach Anschreiben, steigen jedoch im darauffolgenden Monat wieder an. Im Gegensatz dazu sinken die Kosten der Nicht-Teilnehmer nach Anschreiben stetig. Grund dafür ist der Krankheitsverlauf der Depression sowie die Ausgestaltung des DMP. Vom Versanddatum der Einladung bis zur Anmeldung zum DMP vergehen durchschnittlich 42 Tage. In dieser Zeit sinken die Kosten der Teilnehmer analog zu den Kosten der Nicht-Teilnehmer. Nach Anmeldung zum DMP steigen die Rechnungsbeträge in der Gruppe der Teilnehmer, da jeder Teilnehmer nun die Leistungen des DMP, wie u. a. Psychologengespräche, Arztbesuche etc. in Anspruch nimmt. Diese Leistungen sollen die Depression auf langfristige Sicht heilen, sind kurzfristig jedoch mit Kosten verbunden. Nicht-Teilnehmer erhalten nicht die besondere Behandlung des DMP und

⁴⁵ Vgl. Abb. 8: Absolute Leistungsausgaben: Angeschriebene Personen

werden folglich nicht in die leitliniengerechte Therapie gesteuert. Daher sinken die Rechnungsbeträge der Nicht-Teilnehmer im Jahr nach Einladung zum DMP stetig, werden jedoch in Zukunft vermutlich wieder stark ansteigen, sobald eine erneute depressive Episode beginnt.

Anhand der deskriptiven Analyse kann dementsprechend nicht beurteilt werden, ob das DMP erfolgreich war. Daher wird im Folgenden zur Evaluation des Programms das Propensity-Score-Matching durchgeführt.

4 Propensity-Score: Basis-Modell

Ziel eines DMP ist die Verbesserung des Gesundheitszustands der Teilnehmer und damit zusammenhängend die langfristige Reduktion der Leistungsausgaben. Zur Beurteilung der Zielerreichung wird in diesem Kapitel zunächst ein Basis-Modell für das Propensity-Score-Matching hergeleitet. Ausgehend vom Basis-Modell können im Anschluss verschiedene Abweichungen analysiert werden.

Im Basis-Modell werden alle Personen betrachtet, die im Jahr X3 angeschrieben wurden und in diesem Jahr vollständig versichert waren, d. h. Personen, die im Jahr X3 verstorben sind bzw. die aufgrund von Stornierung des Vertrags nicht mehr im Bestand waren, werden nicht in die Auswertung einbezogen. Als Teilnehmer zählen Personen, die sich im Jahr X3 für das DMP anmeldeten und mindestens ein halbes Jahr teilnahmen. Die Kontrollgruppe im Basis-Modell besteht aus 5.673 und die Interventionsgruppe aus 239 Personen.

Bevor der Erfolg des DMP analysiert werden kann (Güte des Programms), muss zunächst die Güte des PS-Modells (Matching-Qualität) beurteilt werden. Nur wenn eine Vergleichbarkeit der Teilnehmer und Nicht-Teilnehmer erzeugt wird, ist das Propensity-Score-Matching für die Evaluation geeignet.

4.1 Analyse der Matching-Qualität

Die Gruppe der Teilnehmer und die Gruppe der Nicht-Teilnehmer unterscheiden sich in ihren Merkmalen und sollen durch das Matching im Vorfeld der eigentlichen Auswertung vergleichbar gemacht werden. Deshalb ist es unerlässlich zu beurteilen, wie sehr sich die beiden Gruppen durch das Matching angenähert haben (Güte des PS-Modells). Dabei ist es bis zur Erreichung einer zufriedenstellenden Matching-Qualität notwendig, verschiedene Methoden, z. B. durch Einbeziehung weiterer Parameter, auszuprobieren. Für die Beurteilung der Güte stehen numerische sowie grafische Methoden zur Verfügung.⁴⁶

4.1.1 Auswahl der Matching-Parameter

Zur Evaluation des Gesundheitsprogramms mittels der Propensity-Score-Methode muss zunächst der PS für jede Person geschätzt werden. Die Teilnahme bzw. Nicht-Teilnahme am Gesundheitsprogramm gilt als abhängige Variable, die zu Therapiebeginn bestehenden Patientenmerkmale hingegen als unabhängige Variablen.

Als unabhängige Variablen wurden folgende Parameter genutzt:

⁴⁶ Vgl. Ho, Daniel E., et al. (2011), S. 4, Abruf am 31.07.2018

Abkürzung	Erläuterung
ALTER	Alter der Person
GESCHLECHT	Geschlecht der Person
V_LEIST_ANTIDEP_COUNT	Anzahl an verschriebenen Antidepressiva
COMORBIDITY	Vom Krankenversicherer berechneter Komorbiditätsindex
LAST_FDIAG_AMB	Zeitraum zwischen der letzten ambulanten F-Diagnose und dem Datum der Einladung
LAST_FDIAG_STAT	Zeitraum zwischen der letzten stationären F-Diagnose und dem Datum der Einladung
LEIST_AMB_360	Ambulante Leistungen in den 360 Tagen vor Einladung
LEIST_STAT_360	Stationäre Leistungen in den 360 Tagen vor Einladung
LEIST_FDIAG_AMB_360	Ambulante Leistungen für F-Diagnosen in den 360 Tagen vor Einladung
LEIST_FDIAG_STAT_360	Stationäre Leistungen für F-Diagnosen in den 360 Tagen vor Einladung
ANZ_FDIAG_AMB_360	Anzahl der ambulanten F-Diagnosen in den 360 Tagen vor Einladung
ANZ_FDIAG_STAT_360	Anzahl der stationären F-Diagnosen in den 360 Tagen vor Einladung

Tab. 3: Merkmale für Zwillingsbildung

Die Auswahl der Parameter erfolgte in Zusammenarbeit mit dem Krankenversicherungsunternehmen und wird im Folgenden erläutert:

Alter und Geschlecht:

Die Häufigkeit sowie die Ausprägung depressiver Symptome unterscheiden sich zwischen den Geschlechtern sowie zwischen einzelnen Altersgruppen.⁴⁷ Folglich ist es notwendig sowohl das Alter als auch das Geschlecht bei der Berechnung des PS zu berücksichtigen. Da das Geburtsjahr auf den 01.01. des jeweiligen Jahres verfremdet wurde, erfolgt die Betrachtung des Alters jeweils als ganze Zahl.

Anzahl an verschriebenen Antidepressiva (V LEIST ANTIDEP COUNT):

Depressionen können medikamentös oder psychotherapeutisch behandelt werden. Sofern der Patient mit Antidepressiva behandelt wird, sollte er die Medikamente über einen längeren Zeitraum von mindestens sechs Monaten einnehmen.⁴⁸ Durch die Berücksichtigung der Anzahl an Antidepressiva, die im Jahr vor Einladung zum Gesundheitsprogramm verschrieben

⁴⁷ Vgl. Busch, M. A., Maske, U. E., Ryl, L., Schlack, R., & Hapke, U. (2013), S. 735

⁴⁸ Vgl. Stiftung Deutsche Depressionshilfe (2018c), Abruf am 06.07.2018

wurde, werden Personen zueinander geordnet, die bislang auf die gleiche Weise behandelt wurden.

Komorbiditätsindex (COMORBIDITY):

Der Komorbiditätsindex wurde vom Krankenversicherer für jede Person berechnet und wird ebenfalls als Matchingkriterium genutzt, da er Aufschluss über den Krankheitsstatus der Person gibt.

Letzte F-Diagnose Ambulant und Stationär (LAST FDIAG AMB/STAT):

Wie zu Beginn erläutert, folgt bei Depressionserkrankten häufig nach der depressiven eine gesunde Phase. Daher wurde für jede Person der Behandlungsbeginn der letzten in den Abrechnungsunterlagen dokumentierten Depressionsdiagnose im ambulanten sowie im stationären Bereich ermittelt. Infolgedessen sollten Personen, die sich in der gleichen Phase befinden, zueinander geordnet werden. Die Differenz zwischen Anschreibedatum und Behandlungsbeginn der letzten diagnostizierten Depression ergibt das Merkmal LAST_FDIAG.

Leistungen Ambulant und Stationär (LEIST AMB / STAT 360):

Die Datei „Leistungsdaten“ enthält sämtliche eingereichte Leistungsdaten im Zeitraum von Januar X-1 bis Ende März X6. Für jede der zum DMP eingeladenen Personen wurden die eingereichten Leistungsdaten im ambulanten sowie im stationären Bereich in den 360 Tagen vor Einladung summiert und zur Berechnung des PS genutzt. Dabei wurden nicht die Leistungsbeträge, sondern die Rechnungsbeträge summiert, da diese Aufschluss über die gesamten Behandlungskosten geben.

Für jede Rechnung enthält die Datei „Leistungsdaten“ das Datum an dem die Rechnung eingereicht wurde, das Datum des Behandlungsbeginns sowie des Behandlungsendes. Häufig sammeln Privatpatienten die Rechnungen während des Jahres und reichen sie gebündelt am Jahresende ein. Daher wird bei der Auswertung jeweils das Datum des Behandlungsbeginns betrachtet. Um eine Vergleichbarkeit der verschiedenen Jahre herzustellen, wird für jeden Rechnungsbeleg ein relatives Datum berechnet:

$$\text{Relatives Datum} = \text{Datum Behandlungsbeginn} - \text{Versanddatum der Einladung}$$

Das relative Datum kann zur Analyse der Leistungskosten vor- und nach Einladung zum DMP genutzt werden.

Leistungen für F-Diagnosen Ambulant und Stationär (LEIST FDIAG AMB / STAT 360):

Insbesondere im stationären Bereich ist es möglich, dass eine Person hohe stationäre Leistungen einreicht, die jedoch nicht vollständig auf Depressionen zurückzuführen sind:

Beispiel	Stationäre Kosten	Kosten für stationäre F-Diagnosen	Anzahl stationäre F-Diagnosen
1	123.316,8	56,31	1
2	100.211,3	85.617,66	72

Tab. 4: Beispiel stationäre Kosten

Beispielperson 1 verursacht in einem Jahr hohe stationäre Kosten. Die stationären Kosten für F-Diagnosen sind jedoch sehr gering. Im Vergleich dazu hat Beispielperson 2 ebenfalls hohe Kosten im stationären Bereich, die jedoch überwiegend auf F-Diagnosen zurückgeführt werden können.

Aufgrund dessen wurden für jede Person die eingereichten Rechnungsbeträge für F-Diagnosen im ambulanten sowie im stationären Bereich in den 360 Tagen vor Einladung summiert und zur Berechnung des PS genutzt.

Anzahl F-Diagnosen Ambulant und Stationär (ANZ FDIAG AMB / STAT 360):

Neben den Leistungsausgaben wurde auch die Anzahl an ambulanten bzw. stationären F-Diagnosen nach Bereichen berechnet und für die 360 Tage vor Versanddatum summiert. Dabei ist möglich, dass innerhalb eines Zeitraums mehrere Diagnosen gleichzeitig gestellt werden. Beispielsweise kann bei einer Person zeitgleich ambulant eine F32.1 (Mittelgradige depressive Episode) sowie eine F33.9-Diagnose (Rezidivierende depressive Störung) diagnostiziert werden.⁴⁹ Personen wurden jedoch nur zum DMP eingeladen, sofern im Vorjahr eine stationäre oder zwei ambulante Depressionsdiagnosen zu unterschiedlichen Zeitpunkten in den Abrechnungsunterlagen dokumentiert wurden. Daher werden mehrere Diagnosen im gleichen Behandlungszeitraum als eine Diagnose gewertet.

Der Propensity-Score wird mittels eines logistischen Regressionsmodells (Logit-Modell) berechnet und gibt die Wahrscheinlichkeit an, ob ein Patient die zu prüfende Therapie erhält. Für die Durchführung des Matching stehen sechs Methoden zur Verfügung:

- Exact Matching
- Subclassification
- Nearest neighbor Matching
- Optimal Matching
- Full Matching
- Genetic Matching.

Als Standardmethode gilt das „nearest neighbor matching“, bei der jedem Teilnehmer ein Nicht-Teilnehmer mit möglichst vergleichbarem PS, d. h. mit einer möglichst geringen Diffe-

⁴⁹ Vgl. Deutsches Institut für Medizinische Dokumentation und Information (2018c), Abruf am 14.07.2018

renz zugeordnet wird. Beispielweise wird eine Person mit einem PS von 0,02985487 mit einer Person, für die ein PS in Höhe von 0,02985321 berechnet wurde, gematcht. Die beiden Scores sind nicht exakt identisch, aber unterscheiden sich erst in der sechsten Nachkommastelle. Es handelt sich jeweils um eine 69-jährige Frau mit drei ambulanten F-Diagnosen, jedoch keiner stationären F-Diagnose im Jahr vor Anschreiben. Beide Personen unterscheiden sich geringfügig in der Höhe der Leistungsausgaben im Jahr vor Anschreiben. Unter anderem betragen die ambulanten Leistungen für F-Diagnosen bei einer Person 291,27 Euro und bei der anderen Person 626,16 Euro. Im Falle eines exakten Matchings werden nur Personen gemacht, die einen identischen PS aufweisen. Bei Nutzung des exakten Matchings besteht jedoch die Gefahr der Fallreduzierung und damit verbunden einer Vergrößerung der Varianz, da es möglich ist, dass zu einigen Propensity Scores kein identischer Score gefunden werden kann.⁵⁰

4.1.2 Bewertung der Matching-Qualität

Nach Durchführung des Matchings muss die Güte des PS-Modells beurteilt werden. Für die Beurteilung stehen einerseits numerische Methoden zur Verfügung, wie u. a. Mittelwerte und die Differenzen der Mittelwerte. Andererseits ist ebenfalls eine grafische Beurteilung mittels sogenannter Quantil-Quantil-Diagramme möglich.⁵¹ Nicht einbezogen werden sollte ein t-Test der Mittelwertdifferenzen, da dieser zu einer falschen Schlussfolgerung führen würde.⁵²

4.1.2.1 Bewertung anhand von Mittelwerten

Für das Matching wird eine Kombination aus „nearest neighbor matching“ und exaktem Matching genutzt. Aufgrund der unterschiedlichen Depressionsausprägungen in verschiedenen Altersklassen und zwischen den Geschlechtern werden die Parameter „Alter“ und „Geschlecht“ exakt („exact“) gematcht, sodass jeweils Personen im gleichen Alter und mit gleichem Geschlecht zueinander geordnet werden. Für alle weiteren Merkmale wird die „nearest neighbor matching“-Methode genutzt. Die PS der Teilnehmer und Nicht-Teilnehmer sollen dabei einen möglichst geringen Abstand aufweisen. Als Abstandsmaß wird ein logistisches Regressionsmodell (Logit-Funktion) verwendet.

Nach Durchführung des Matchings erstellt R als Ergebnis des Befehls „summary (m.out)“ eine Zusammenfassung der Parameterwerte des ursprünglichen Datensatzes sowie des neu gematchten Datensatzes mit Hilfe deren eine numerische Beurteilung der Güte des PS-Modells möglich ist. Die Zusammenfassung enthält u. a. die Mittelwerte je Gruppe sowie die Differenz der Mittelwerte der beiden Gruppen, die Standardabweichung der Kontrollgruppe und die Quantil-Differenzen zwischen den beiden Gruppen (Minimum, Maximum, Median). Zudem wird angegeben, um wie viel Prozent sich die Werte verbessert haben. Sofern das

⁵⁰ Vgl. Ho, Daniel E., et al. (2011), S. 6f., Abruf am 31.07.2018

⁵¹ Vgl. Ho, Daniel E., et al. (2011), S. 3f., Abruf am 31.07.2018

⁵² Vgl. Imai Kosuke, King Gary, Stuart Elizabeth A. (2008), S. 481f., Abruf am 06.08.2018

Matching gut war, sollten sich die Abweichungen der Mittelwerte von Teilnehmern und Zwillingen reduziert haben.⁵³

Im Rahmen des Basis-Modells kann allen 239 Teilnehmer ein Zwilling zugeordnet werden, sodass alle Teilnehmer in der Auswertung betrachtet werden. Im Folgenden wird die Differenz der Mittelwerte der beiden Gruppen vor und nach Matching abgebildet. Dabei werden die Mittelwertdifferenzen aus der Differenz zwischen durchschnittlichem Mittelwert der Teilnehmer und durchschnittlichem Mittelwert der Nicht-Teilnehmer bzw. nach Matching der Zwillinge bestimmt:

$$\begin{aligned} \text{Differenz der Mittelwerte} = & \text{ Durchschnittlicher Mittelwert Teilnehmer} \\ & - \text{ Durchschnittlicher Mittelwert Nicht-Teilnehmer/Zwillinge} \end{aligned}$$

Merkmal	Differenz der Mittelwerte		proz. Änderung
	vor Matching	nach Matching	
ALTER	- 0,06	0,00	100,00
GESCHLECHT	0,09	0,00	100,00
V_LEIST_ANTIDEP_COUNT	0,40	- 0,06	84,17
COMORBIDITY	- 0,04	0,07	- 88,78
LAST_FDIAG_AMB	34,72	0,63	98,18
LAST_FDIAG_STAT	- 29,86	- 17,26	42,20
LEIST_AMB_360	341,16	118,21	65,35
LEIST_STAT_360	2.535,63	1.415,10	44,19
LEIST_FDIAG_AMB_360	42,65	- 19,02	55,41
LEIST_FDIAG_STAT_360	918,31	162,35	82,32
ANZ_FDIAG_AMB_360	0,63	- 0,08	88,09
ANZ_FDIAG_STAT_360	0,67	0,33	49,93

Tab. 5: Differenz der Mittelwert vor- und nach Matching – Basis-Modell (1)

Differenz der Mittelwerte vor Matching:⁵⁴

Die Mittelwerte der Teilnehmer vor Matching berechnet sich anhand der Werte der 239 Programmteilnehmer im Jahr X3, die Werte der Nicht-Teilnehmer anhand der 5.673 Nicht-Teilnehmer.

Vor Durchführung des Matchings unterscheiden sich die Teilnehmer und Nicht-Teilnehmer strukturell in allen Merkmalen. Die Gruppe der Nicht-Teilnehmer ist durchschnittlich geringfügig älter (- 0,06) und enthält mehr Frauen und weniger Männer als die Gruppe der Teilneh-

⁵³ Vgl. Ho, Daniel E., et al. (2011), S. 8f., Abruf am 31.07.2018

⁵⁴ Vgl. Tab. 15: Mittelwerte vor Matching – Basis-Modell

mer. Durchschnittlich wurden den Teilnehmern mehr Antidepressiva verschrieben (+ 0,4). Nichtsdestotrotz haben die Teilnehmer im Vergleich zu den Nicht-Teilnehmern einen geringeren Komorbiditätsindex (- 0,04).

Die letzte ambulante F-Diagnose wurde bei den Teilnehmern durchschnittlich 102 Tage vor Versanddatum der DMP-Einladung festgestellt, bei Nicht-Teilnehmern hingegen 137 Tage vor Anschreibedatum, d. h. der Zeitpunkt der letzten ambulanten F-Diagnose liegt bei Teilnehmern näher am Anschreibedatum als bei den Nicht-Teilnehmern (+ 34,72). Der Zeitpunkt der letzten stationären F-Diagnose hingegen liegt bei den Nicht-Teilnehmern näher am Versanddatum der Einladung (- 29,86).

Insbesondere die Leistungsausgaben der Teilnehmer sind höher als die der Nicht-Teilnehmer. Eine besonders große Differenz zwischen den Mittelwerten liegt bei den stationären Leistungen in den 360 Tagen vor Einladung zum DMP vor. Teilnehmer des Programms verursachen in der Zeit vor Einladung durchschnittlich deutlich höhere stationäre Rechnungsbeträge (+ 2.536 Euro) als Nicht-Teilnehmer. Aber auch die durchschnittlichen ambulanten Rechnungsbeträge sind in der Gruppe der Teilnehmer höher als in der Gruppe der Nicht-Teilnehmer (+ 341,16 Euro). Diese Unterschiede können auch in *Abb. 7: Durchschnittliche Leistungsausgaben der DMP Teilnehmer* sowie *Abb. 30: Durchschnittliche Leistungsausgaben der DMP Nicht-Teilnehmer* erkannt werden.

Die Mittelwertdifferenzen der Depressionskosten sind ähnlich der Differenzen der Gesamtkosten. Teilnehmer verursachen im ambulanten Bereich durchschnittlich 42,65 Euro mehr. Analog der Gesamtkosten ist die Differenz bezogen auf die Depressionskosten im stationären Bereich mit 918,31 Euro in der Gruppe der Teilnehmer deutlich größer als in der Gruppe der Nicht-Teilnehmer.

Die höheren Depressionskosten der Teilnehmer hängen vermutlich auch mit der durchschnittlich höheren Anzahl an F-Diagnosen (Ambulant: + 0,63; Stationär: + 0,67) zusammen.

Eine Auswertung und Beurteilung der Programm-Qualität anhand der ungematchten Daten würde aufgrund der Verschiedenheit der Gruppen zu einer Verzerrung führen. Aus diesem Grund ist ein Matching notwendig.

Differenz der Mittelwerte nach Matching:⁵⁵

Das Matching impliziert, dass den Teilnehmern Nicht-Teilnehmer mit gleichem bzw. ähnlichem PS zugeordnet werden, die sogenannten Zwillinge, sodass sich die Mittelwerte stark annähern. Da im Matching-Modell alle Teilnehmer berücksichtigt werden, verändern sich die Mittelwerte der Teilnehmer nicht. Lediglich die Mittelwerte der Kontrollgruppe unterscheiden sich nach dem Matching von den Mittelwerten vor Durchführung des Matchings, da die Werte nun ausschließlich als Durchschnitt der 239 zugeordneten Nicht-Teilnehmern berechnet

⁵⁵ Vgl. Tab. 16: Mittelwerte nach Matching– Basis-Modell (1)

werden. Die 5.434 Nicht-Teilnehmer, die keinem Teilnehmer zugeordnet werden konnten, werden bei der Mittelwertbetrachtung nach Matching nicht berücksichtigt.

In Folge des exakten Matchings der Parameter „Alter“ und „Geschlecht“ besteht nun kein Unterschied mehr zwischen der Gruppe der Teilnehmer und der Gruppe der Zwillinge. In beiden Gruppen sind die Personen nun durchschnittlich 55,87 Jahre alt und jede Gruppe enthält 151 Männer und 88 Frauen.

Die deutlichste Annäherung der Mittelwerte ist anhand des Parameters „LAST_FDIAG_AMB“ mit einer Verbesserung um 98,18 % zu beobachten. Die letzte ambulante F-Diagnose wurde bei den Teilnehmern durchschnittlich 101,97 Tage und bei den Zwillingen 102,6 Tage vor Anschreiben in einer Rechnung aufgeführt, was zu einer Mittelwertdifferenz von lediglich 0,63 Tagen führt.

Reduziert haben sich ebenfalls die Mittelwertdifferenzen der Anzahl an verschriebenen Antidepressiva im Jahr vor Einladung zum DMP (- 0,06) sowie die Differenz des Komorbiditätsindex (0,07), sodass die beiden Gruppen in Bezug auf diese Kriterien vergleichbar geworden sind. In der Gruppe der Zwillinge ist die Anzahl an verschriebenen Antidepressiva vor Einladung zum DMP geringfügig höher, der Komorbiditätsindex hingegen ist im Vergleich zu den Teilnehmern geringer.

Die Anzahl der ambulanten F-Diagnosen im Jahr vor Anschreiben ist nach dem Matching fast identisch: Während bei den Teilnehmern im ambulanten Bereich durchschnittlich 3,78 Depressionsdiagnosen dokumentiert wurden, liegt die Anzahl bei den Zwillingen bei 3,86. Dies wirkt sich ebenfalls auf die Ausgaben für Depressionen im ambulanten Bereich aus. Infolge der höheren Anzahl an ambulanten F-Diagnosen sind ebenfalls die ambulanten Depressionskosten der Zwillinge geringfügig höher (- 19,02 Euro).

Die Mittelwertdifferenz der Anzahl an stationären F-Diagnosen konnte nicht ganz so stark reduziert werden wie im ambulanten Bereich und liegt nach dem Matching weiterhin bei 0,33. Die stationären Depressionskosten hingegen sind mit einer Änderung in Höhe von 82,32 % vergleichbarer geworden, wenn auch mit einer weiterhin bestehenden Mittelwertdifferenz von 162,35 Euro.

Das Matching führt jedoch bei den Gesamtausgaben im ambulanten Bereich zu keiner vollständigen Vergleichbarkeit der beiden Gruppen. Nach Matching beträgt die Mittelwertdifferenz weiter 118,21 Euro und konnte um lediglich 65,35 % verbessert werden.

Zudem gibt es weitere Parameter bei denen sich die Mittelwerte nicht besonders gut angenähert haben, sodass die beiden Gruppen in Bezug auf diese Kriterien nicht vergleichbar sind. Die Mittelwerte der beiden Gruppen weichen insbesondere mit Blick auf die gesamten stationären Kosten in den 360 Tagen vor Anschreiben stark voneinander ab. Die Differenz der Mittelwerte in Höhe von 1.415,10 Euro und die prozentuale Verbesserung von lediglich 44,19 % sind kein Zeichen für ein gutes Matching. Ähnlich ist es bei Betrachtung der letzten stationären F-Diagnose. Nach Matching beträgt der Unterschied zwischen den Teilnehmern

und den Zwillingen trotz der Zuordnung mehr als einen halben Monat (- 17,26). Die Mittelwerte haben sich um lediglich 42,2 % angenähert.

Abgesehen vom exakten Matching der Parameter „Alter“ und „Geschlecht“ sollten weitere zehn Parameter mittels der „nearest neighbor matching-Methode“ vergleichbar gemacht werden. Insgesamt liegt die prozentuale Verbesserung, nur einmal über 90 %. In drei Fällen liegt sie sogar unter 50 %. Demzufolge ist das Matching, insbesondere mit Blick auf die stationären Parameter, noch nicht optimal, sodass durch Abwandlung der Matching-Methode versucht werden sollte, die Mittelwertdifferenz zu verringern und so die beiden Gruppen vergleichbarer zu machen.

Eine Möglichkeit zur Verbesserung des Matchings ist die Berücksichtigung weiterer bzw. weniger Parameter. Zudem kann die genutzte Matching-Methode verändert werden. Im soeben dargestellten Matching wurde eine Kombination aus der „nearest neighbor matching-Methode“ und einem exakten Matching für die Parameter „Geschlecht“ und „Alter“ verwendet. Durch das exakte Matching des Parameters „Alter“ werden Personen zueinander geordnet, die genau gleich alt sind. Die Auswahl der Methode für die Variable basiert auf der Tatsache, dass Unterschiede in der Ausprägung einer Depression abhängig vom Alter vorliegen. Diese Unterschiede sind jedoch eher zwischen Altersgruppen, als zwischen den einzelnen Jahren vorzufinden. Eine 20-jährige Person sollte nicht mit einer 70-jährigen Person gematcht werden, jedoch sollte die Zuordnung einer 53-jährigen Person mit einer 54-jährigen Person kein Problem darstellen.

Das Matching wird daher erneut durchgeführt, sodass nun lediglich die Variable „Geschlecht“ exakt gematcht wird. Für alle weiteren Parameter wird die „nearest neighbor matching-Methode“ verwendet.

Im Folgenden werden die Differenzen der Mittelwerte vor und nach Matching sowie die prozentuale Veränderung der Mittelwertdifferenzen des neuen Matchings dargestellt. Die Differenzen vor Matching entsprechen den Werten, die zuvor erläutert wurden, weshalb auf eine erneute Analyse verzichtet wird.

Merkmal	Differenz der Mittelwerte		proz. Änderung
	vor Matching	nach Matching	
ALTER	- 0,06	0,41	- 596,13
GESCHLECHT	0,09	0,00	100,00
V_LEIST_ANTIDEP_COUNT	0,40	- 0,06	85,23
COMORBIDITY	- 0,04	0,14	- 277,56
LAST_FDIAG_AMB	34,72	- 0,98	97,18
LAST_FDIAG_STAT	- 29,86	1,31	95,60
LEIST_AMB_360	341,16	180,91	46,97
LEIST_STAT_360	2.535,63	- 17,83	99,30
LEIST_FDIAG_AMB_360	42,65	- 22,94	46,21
LEIST_FDIAG_STAT_360	918,31	78,56	91,45
ANZ_FDIAG_AMB_360	0,63	- 0,05	91,40
ANZ_FDIAG_STAT_360	0,67	0,11	83,10

Tab. 6: Differenz der Mittelwert vor- und nach Matching – Basis-Modell (2)

Differenz der Mittelwerte nach Matching:⁵⁶

Nach Durchführung des Matchings haben sich die Mittelwerte der Teilnehmer und der Zwillinge stark angenähert. Im Unterschied zur ersten Durchführung des Matchings sind keine extrem hohen Mittelwertdifferenzen mehr zu erkennen.

Eine prozentuale Verschlechterung besteht beim Parameter „Alter“. Die Teilnehmer sind im Durchschnitt 55,9 Jahre und ihre zugeordneten Zwillinge 55,5 Jahre alt. Prozentual ist die Verschlechterung zwar bemerkenswert (- 596,13 %), der Unterschied der Mittelwerte hingegen ist jedoch so gering (0,41), dass er die Auswertung nicht beeinflusst.

Aufgrund des weiterhin durchgeführten exakten Matchings des Merkmals „Geschlecht“ besteht keine Differenz zwischen den beiden Gruppen. Allen männlichen wurden männliche Personen und allen weiblichen wurden weibliche Personen zugeordnet.

Die Mittelwertdifferenzen der Parameter „Anzahl der verschriebenen Antidepressiva“ sowie „Korbiditätsindex“ entwickeln sich analog zum ersten Matching. Die Anzahl der verschriebenen Antidepressiva vor Einladung zum DMP ist in der Gruppe der Zwillinge geringfügig höher (- 0,06), der Korbiditätsindex hingegen ist im Vergleich zu den Teilnehmern geringer (0,14). Die Mittelwertdifferenz bezogen auf den Korbiditätsindex ist nicht besonders hoch, sodass trotz einer prozentualen Änderung von - 277,56 % eine Vergleichbarkeit zwischen den beiden Gruppen gegeben ist.

⁵⁶ Vgl. Tab. 17: Mittelwerte nach Matching– Basis-Modell (2)

Deutlich angenähert mit Verbesserungen über 95 % hat sich das Merkmal „LAST_FIAG“, sodass in beiden Gruppen die letzte ambulante sowie die letzte stationäre F-Diagnose zeitlich ungefähr gleich lange her ist (Ambulant: - 0,98; Stationär: 1,31). Insbesondere bezogen auf die letzte stationäre F-Diagnose konnte im Vergleich zum ersten Matching eine deutliche Verbesserung erreicht werden (von - 17,26 auf 1,31).

Die Mittelwertdifferenz der ambulanten Gesamtkosten konnte durch das Matching reduziert werden (180,91 Euro), wenn auch nicht ganz so stark, wie im ersten Matching. Jedoch konnte die hohe Differenz in den stationären Leistungsausgaben besonders stark reduziert werden. Vor Matching verursachten die Teilnehmer stationär durchschnittlich 2.536 Euro mehr Kosten als die Nicht-Teilnehmer. Nach Matching liegt der Unterschied zwischen den Mittelwerten bei lediglich - 17,83 Euro. Folglich sind die Leistungsausgaben der Zwillinge nach Matching geringfügig höher als die der Teilnehmer. Dies entspricht einer prozentualen Änderung in Höhe von 99,3 %. Auch bei diesem Parameter ist die prozentuale Änderung verglichen mit dem ersten Matching (44,19 %) deutlich besser.

Bei Betrachtung der Kosten, die nur für F-Diagnosen anfallen, ist eine Ähnlichkeit zu den Gesamtkosten beobachtbar. Die ambulanten Depressionskosten nähern sich mit einer prozentualen Änderung von 46,21 % weniger an, als die stationären Depressionskosten mit einer Änderung von 91,45 %. Der Unterschied der Mittelwerte der ambulanten Kosten kann von 42,65 Euro auf - 22,94 Euro und der stationären Kosten von 918,31 Euro auf 78,56 Euro reduziert werden.

Auch im Hinblick auf die Anzahl der F-Diagnosen sind beide Gruppen nach dem Matching vergleichbar (Ambulant: - 0,05; Stationär: 0,11).

Die Differenzen der Mittelwerte haben sich deutlich verbessert und sind nach Matching nur noch gering. Von den elf Parametern, die mittels der „nearest neighbor matching-Methode“ vergleichbar gemacht werden sollten, liegt die prozentuale Änderung bei fünf Parametern über 90 %. In zwei Fällen hat sie sich prozentual zwar deutlich verschlechtert, jedoch ist die Differenz der Mittelwerte dennoch annehmbar. Daher kann bei Analyse der Matching-Qualität auf Basis der Mittelwerte geschlussfolgert werden, dass die Güte des PS-Modells gut ist.

4.1.2.2 Bewertung anhand von Quantil-Quantil-Diagrammen

Mittelwerte werden häufig durch einzelne hohe Werte verzerrt. Daher ist neben einer numerischen Beurteilung der Güte des PS-Modells ebenfalls eine grafische Beurteilung sinnvoll. Mithilfe des Befehls „plot ()“ können drei verschiedene Grafikarten erstellt werden.⁵⁷ Zur Betrachtung der Verteilung der Propensity-Scores in der Gruppe der Teilnehmer sowie der

⁵⁷ Vgl. Ho, Daniel E., et al. (2011), S. 9f., Abruf am 31.07.2018

Nicht-Teilnehmer kann ein „Jitter-Plot“ erstellt werden.⁵⁸ Zudem kann die Dichte der Verteilung der Propensity-Scores sowohl von den ungematchten als auch von den gematchten Daten der Teilnehmer und Nicht-Teilnehmer bzw. Zwillingen mithilfe von Histogrammen bewertet werden.⁵⁹ Des Weiteren besteht die Möglichkeit für jeden Parameter ein Quantil-Quantil-Diagramm (QQ-Plots) zu erstellen und diese zur Analyse der Verteilung der Parameter zu nutzen. Dabei werden die Quantile der Teilnehmer (x-Achse) mit den Quantilen der Nicht-Teilnehmer (y-Achse) verglichen und als Punkte im Diagramm abgetragen. Für die Auswertung „nach Matching“ erfolgt der Vergleich mit den entsprechenden Zwillingen (y-Achse). Sofern identische Quantile vorliegen, liegen die Punkte auf der Winkelhalbierenden (45 Grad Linie). Weichen die Quantile hingegen voneinander ab, liegen die Punkte ober- bzw. unterhalb der Winkelhalbierenden.⁶⁰

Letzte F-Diagnose Ambulant und Stationär (LAST_FDIAG_AMB/STAT):

Die Analyse der Mittelwerte ergab, dass vor Durchführung des Matchings der Zeitpunkt der letzten ambulanten F-Diagnose bei Teilnehmern näher am Anschreibedatum liegt als bei den Nicht-Teilnehmern (Mittelwertdifferenz: 34,72 Tage). Dieser Unterschied ist ebenfalls bei Betrachtung der QQ-Plots erkennbar. Vor Matching liegen sämtliche Punkte in der Teilnehmerhälfte, d. h. die letzte ambulante F-Diagnose der Nicht-Teilnehmern liegt länger in der Vergangenheit zurück.

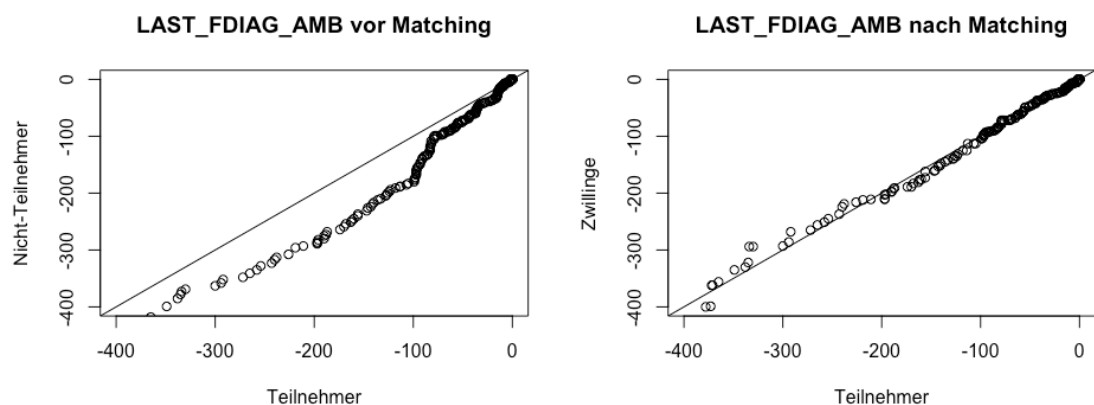


Abb. 11: QQ-Plots Matching-Qualität: Letzte F-Diagnose Ambulant

Durch das Matching konnte eine starke Verbesserung der Vergleichbarkeit erzielt werden. Im oberen Bereich der Verteilung liegen die Punkte vollständig auf der Winkelhalbierenden. In den längeren Zeiträumen vor Anschreiben weichen die Quantile der Teilnehmer und ihrer Zwillinge zwar leicht ab, dennoch sind diese Punkte deutlich näher an der 45 Grad Linie als vor Durchführung des Matchings. Ab einem Zeitraum von 200 Tagen vor Anschreiben liegen

⁵⁸ Vgl. Abb. 36: Jitter-Plot und Histogramme des Basis-Modells

⁵⁹ Vgl. Abb. 36: Jitter-Plot und Histogramme des Basis-Modells

⁶⁰ Vgl. Wollschläger, Daniel (2012), S. 463f.

die Quantile leicht in der Hälfte der Zwillinge, wodurch die Mittelwertdifferenz von - 0,98 Tagen zu Stande kommt.

Wie bereits in Kapitel 3.1 Allgemeine Analyse des Datensatzes erläutert, wurde bei mehreren der angeschriebenen Personen im Jahr vor Einladung zum DMP keine stationäre Depression diagnostiziert. Folglich enthalten die QQ-Plots der letzten stationären F-Diagnose weniger Punkte als die QQ-Plots der letzten ambulanten F-Diagnose, da weniger Quantile abgetragen werden konnten.

Im Unterschied zur letzten ambulanten F-Diagnose liegt der Zeitpunkt der letzten stationären F-Diagnose bei den Nicht-Teilnehmern vor Matching näher am Versanddatum der Einladung. Die Differenz der Mittelwerte von - 29,86 Tagen ist darauf zurückzuführen, dass sämtliche Punkte in der Nicht-Teilnehmerhälfte vorzufinden sind und diese Punkte weit entfernt von der Winkelhalbierenden liegen.

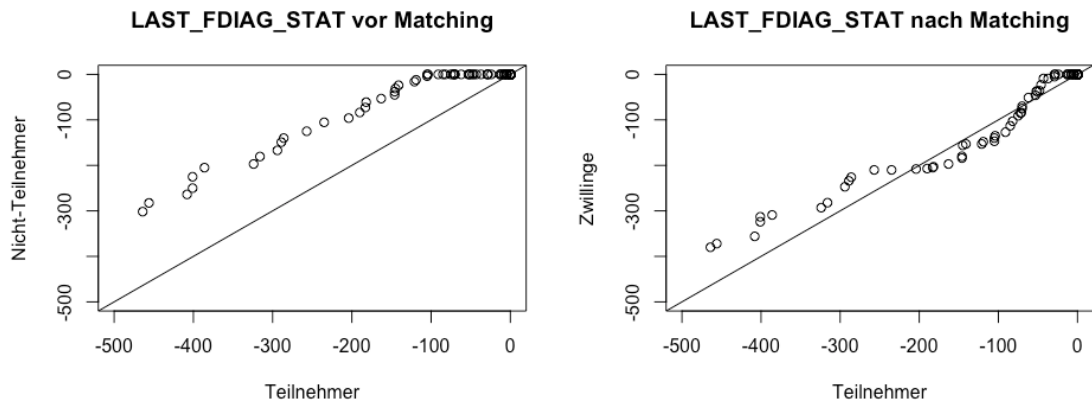


Abb. 12: QQ-Plots Matching-Qualität: Letzte F-Diagnose Stationär

Die Annäherung durch das Matching konnte im Vergleich zum Merkmal der letzten ambulanten F-Diagnose nicht optimal erreicht werden. Nach Matching beträgt die Mittelwertdifferenz 1,31 Tage. Bezogen auf die letzte stationäre F-Diagnose haben sich im Zeitraum bis 100 Tage vor Anschreiben die Teilnehmer und deren Zwillinge stark angenähert. Die abgetragenen Quantile liegen auf bzw. nahe der Winkelhalbierenden. In den 200 bis 500 Tagen vor Anschreiben liegen die Punkte geringfügig weiter von der Winkelhalbierenden entfernt, jedoch im Vergleich zum QQ-Plot vor Matching dennoch näher dran.

Das Matching der letzten F-Diagnose muss nicht vollständig identisch sein. Eine Person, bei der die letzte stationäre F-Diagnose 300 Tage vor Anschreiben liegt, ist mit einer Person, bei der die letzte stationäre F-Diagnose 305 Tage vor Anschreiben liegt, trotz des Unterschieds dennoch vergleichbar, sodass trotz Abweichung von der Winkelhalbierenden dennoch von einer guten Matching-Qualität in Bezug auf das Merkmal „Letzte F-Diagnose Stationär“ ausgegangen werden kann.

Folglich kann nach Analyse der Mittelwerte und der QQ-Plots festgehalten werden, dass die beiden Gruppen in Bezug auf die letzte ambulante sowie auch stationäre F-Diagnose vergleichbar geworden sind.

Leistungen Ambulant und Stationär (LEIST_AMB / STAT_360):

Die Betrachtung von QQ-Plots der ambulanten Leistungsausgaben ist weniger aussagekräftig als der QQ-Plots der stationären Parameter. Im ambulanten Bereich liegt eine Vielzahl ähnlicher Rechnungsbeträge vor, sodass bereits vor Matching der überwiegende Teil der Quantile auf der Winkelhalbierenden liegt. Lediglich im Bereich ab 10.000 Euro weichen die Quantile leicht voneinander ab. In diesem Bereich verursachen Teilnehmer höhere Leistungsausgaben als Nicht-Teilnehmer, da die Punkte in der Teilnehmerhälfte, d. h. unterhalb der Winkelhalbierenden liegen. Dies war bereits anhand der Mittelwertdifferenz in Höhe von 341,16 Euro ersichtlich.

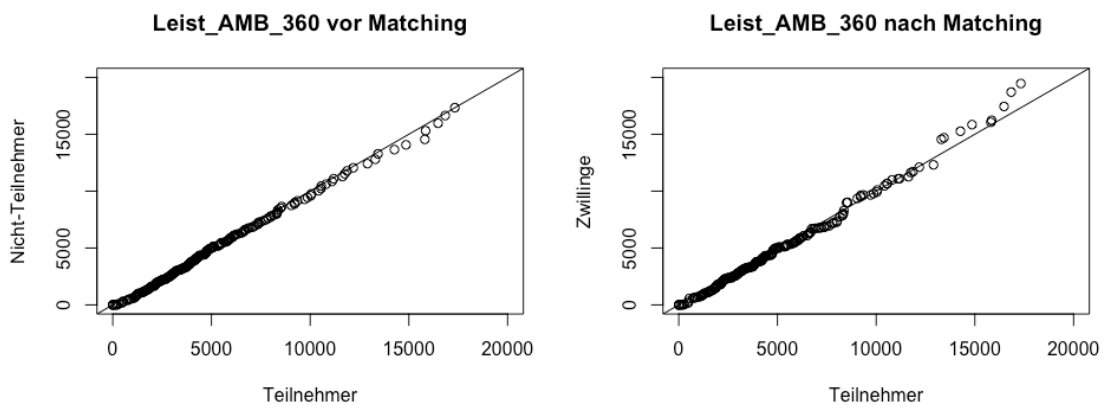


Abb. 13: QQ-Plots Matching-Qualität: Leistungen Ambulant

Nach Matching hat sich die Punktelinie sogar verschlechtert. Im Bereich bis 10.000 Euro liegen die Punkte weiterhin auf der 45 Grad Linie. Im oberen Bereich ab 15.000 Euro hingegen haben sich die Punkte weiter als im QQ-Plot vor Matching von der 45 Grad Linie entfernt. Zwillinge verursachen in diesem Bereich höhere Leistungsausgaben als die Teilnehmer. Dieses Beispiel zeigt, dass eine Kombination aus numerischer und grafischer Beurteilung notwendig ist. Die Verschiebung der Punkte auf die Seite der Zwillinge führt zu einer Reduktion der Mittelwertdifferenz, sodass bei alleiniger Betrachtung der Mittelwerte auf eine Verbesserung der Vergleichbarkeit zu schließen ist, wohingegen die grafische Beurteilung auf keine Verbesserung hinweist.

Durch die Verschlechterung des QQ-Plots nach Matching sind zueinander geordneten Personen im oberen Teil der Verteilung in Bezug auf die ambulanten Gesamtkosten nicht unmittelbar vergleichbar. Nichtsdestotrotz ist zu beachten, dass zum einen der Großteil der Teilnehmer mit dem Großteil der Zwillinge in Bezug auf die ambulanten Leistungen ver-

gleichbar ist und zum anderen für das Matching 12 Parameter genutzt werden. Sofern die ambulanten Gesamtkosten als einziger Parameter kein optimales Matching erzeugen, kann dennoch von einer guten Matching-Qualität des gesamten PS-Modells ausgegangen werden.

Die QQ-Plots der stationären Leistungsausgaben unterscheiden sich hingegen von den QQ-Plots der ambulanten Leistungsausgaben. Das QQ-Plot vor Matching zeigt einen deutlichen Unterschied zwischen den Leistungsausgaben der Teilnehmer und denen der Nicht-Teilnehmer. Dieser Unterschied konnte bereits an der Mittelwertdifferenz in Höhe von 2.535,63 Euro erkannt werden. Im QQ-Plot liegen sämtliche Punkte in der Teilnehmerhälfte, was bedeutet, dass die Ausgaben der Teilnehmer in allen Quantilen höher sind als die der Nicht-Teilnehmer.

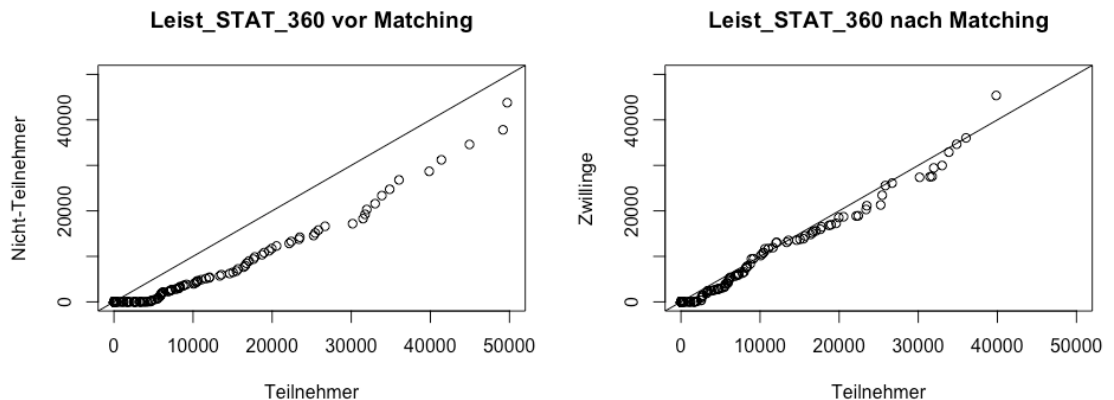


Abb. 14: QQ-Plots Matching-Qualität: Leistungen Stationär

In Folge des Matching haben sich beiden Gruppen deutlich angenähert. Der Großteil der Punkte liegt auf der Winkelhalbierenden. Lediglich im oberen Teil der Verteilung stimmen die Quantile von Teilnehmern und Zwillingen nicht exakt überein und die Teilnehmer verursachen ab einer Höhe von 20.000 Euro geringfügig mehr Kosten als die zugeordneten Zwillinge. Dennoch konnte eine starke Verbesserung zwischen den Grafiken erzielt werden. Das letzte Quantil hingegen liegt deutlich entfernt von der Winkelhalbierenden in der Hälfte der Zwillinge. Damit kann auch die Mitteldifferenz in Höhe von - 17,83 Euro erklärt werden.

Die Analyse der Mittelwerte und der QQ-Plots ergibt, dass die beiden Gruppen in Bezug auf die ambulanten Gesamtkosten nicht optimal vergleichbar geworden sind. Der große Unterschied zwischen den beiden Gruppen im Hinblick auf die stationären Gesamtkosten konnte jedoch deutlich reduziert werden.

Im Folgenden wird aufgrund der geringeren Aussagekraft bezogen auf die ambulanten Leistungen für F-Diagnosen und die Anzahl der ambulanten F-Diagnosen auf die Betrachtung der QQ-Plots der ambulanten Parameter verzichtet.

Leistungen für F-Diagnosen Stationär (LEIST_FDIAG_STAT_360):

Das folgende QQ-Plot stellt einen Auszug aus *Abb. 14: QQ-Plots Matching-Qualität: Leistungen Stationär* dar und ähnelt daher dieser Abbildung. Von den stationären Leistungen für alle Krankheiten werden in den folgenden QQ-Plots nur die Leistungen abgetragen, die für Depressionsdiagnosen anfallen. Teilnehmer haben vor Durchführung des Matchings größtenteils höhere Leistungsausgaben als Nicht-Teilnehmer mit einer Mittelwertdifferenz von 918,31 Euro. Der Abstand der Quantile zur Winkelhalbierenden ist deutlich. Lediglich das höchste Quantil weicht ab und ist bei den Nicht-Teilnehmern höher.

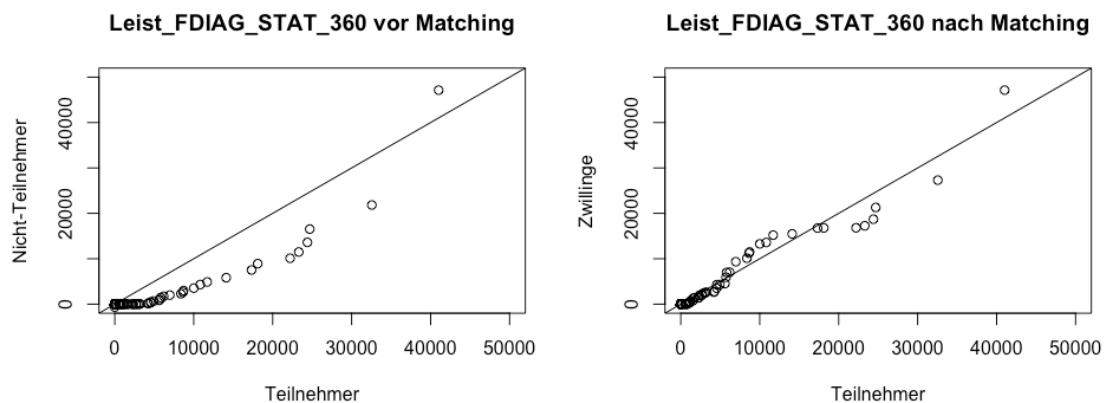


Abb. 15: QQ-Plots Matching-Qualität: Leistungen F-Diagnosen Stationär

Analog zu den stationären Gesamtkosten führt das Matching dazu, dass im Bereich der kleinen Beträge die Quantile auf der Winkelhalbierenden liegen und folglich bei Teilnehmern und Zwillingen gleich sind. Ab einem Wert von 10.000 Euro ändert sich die Verteilung jedoch. Zunächst liegen einige Quantile oberhalb der 45 Grad Linie, was beutet, dass die Zwillinge höhere Kosten als die Teilnehmer verursachen. Anschließend liegen die Punkte jedoch unterhalb der 45 Grad Linie. Dennoch konnte eine deutliche Annäherung an die Winkelhalbierende erzielt werden, wodurch sich die Mittelwertdifferenz auf 78,56 Euro verringert und eine Verbesserung der Vergleichbarkeit zwischen den beiden Gruppen erzielt werden konnte.

Dementsprechend kann nach Analyse der Mittelwerte und der QQ-Plots festgehalten werden, dass Teilnehmer und Zwillinge in Bezug auf die stationären Kosten für F-Diagnose vergleichbar geworden sind.

Anzahl F-Diagnosen Stationär (ANZ_FDIAG_STAT_360):

Vor Durchführung des Matchings hatten Teilnehmer im stationären Bereich eine höhere Anzahl an F-Diagnosen als Nicht-Teilnehmer, was einerseits durch die Mittelwertdifferenz in Höhe von 0,67 sowie andererseits im QQ-Plot durch die Lage der Punkte in der Hälfte der Teilnehmer erkennbar ist. Die meisten abgetragenen Quantile liegen in einem Bereich bis zu 10 F-Diagnosen. Nur wenige Punkte liegen oberhalb von 10 F-Diagnosen. Da die Anzahl von Personen mit stationären F-Diagnosen im Vorjahr des Anschreibens gering ist, konnten in den unten dargestellten QQ-Plots analog der QQ-Plots der stationären Depressionsausgaben nur wenige Quantile abgetragen werden. Sofern nur wenige Werte vorhanden sind, ist es besonders wichtig, neben der Betrachtung der Mittelwerte auch eine grafische Beurteilung durchzuführen, da einzelne Ausreißer die Mittelwerte verzerren können.

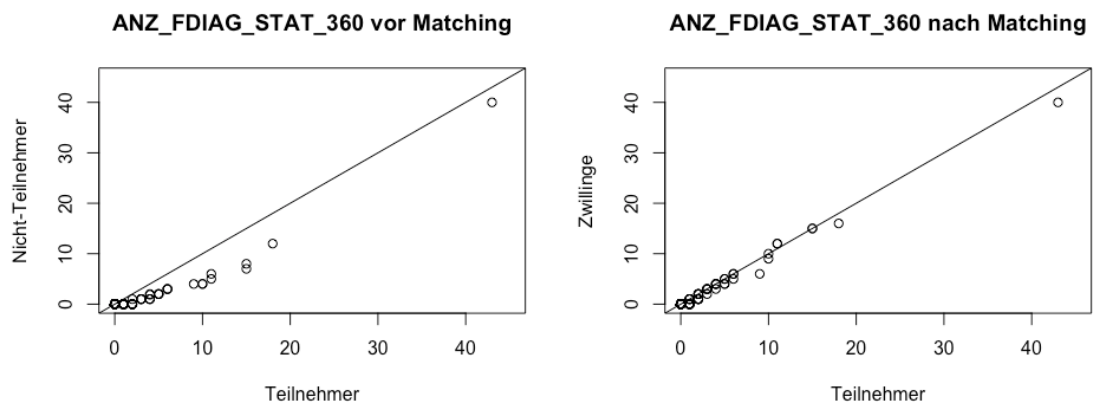


Abb. 16: QQ-Plots Matching-Qualität: Anzahl F-Diagnosen Stationär

Nach Matching haben sich die Mittelwerte prozentual um 83,10 % angenähert, sodass die Mittelwertdifferenz noch 0,11 beträgt. Diese Annäherung kann auch bei Betrachtung der QQ-Plots erkannt werden. Der Großteil der Punkte liegt nach dem Matching auf der Winkelhalbierenden bzw. nahe der Winkelhalbierenden.

Die Betrachtung der Mittelwerte sowie der QQ-Plots ergibt, dass Teilnehmer und Zwillinge durch das Matching in Bezug auf die Anzahl der stationären F-Diagnosen im Jahr vor Anschreiben vergleichbar geworden sind.

Fazit Matching-Qualität Basis-Modell:

Neben der Annäherung der Differenzen der Mittelwerte kann auch anhand der QQ-Plots erkannt werden, dass die Gruppe der Teilnehmer und die Gruppe der Zwillinge nach dem Matching vergleichbar geworden sind. Teilnehmer und Zwillinge sind zwar nicht vollständig identisch, insbesondere im Hinblick auf die ambulanten Gesamtkosten, dies wäre jedoch

auch nur bei einem exakten Matching möglich. Die durch das Matching erzeugte Vergleichbarkeit reicht aus, um zusammenfassend die Güte des PS-Modells als gut einzuschätzen.

4.2 Analyse der Programm-Qualität

Zur Analyse der Programm-Qualität (Güte des Programms) werden die gleichen Parameter genutzt wie auch bei der Analyse der Programmqualität:

- Allgemeine Leistungsausgaben (Gesamt, Stationär, Ambulant)
- Leistungsausgaben für F-Diagnosen (Gesamt, Stationär, Ambulant)
- Anzahl der F-Diagnosen (Gesamt, Stationär, Ambulant).

Im ersten Jahr nach Programmteilnahme ist mit einer Erhöhung der Leistungsausgaben der Teilnehmer zu rechnen, da eine leitliniengerechte Therapie je nach Schweregrad eine intensivere medizinische Betreuung erfordert. Daher werden jeweils die Leistungsausgaben bzw. die Anzahl der F-Diagnosen für das erste und für das zweite Jahr separat ausgewiesen. Die Zeiträume sind dabei wie folgt:

Jahr 1 nach Anschreiben	=	Zeitraum 0 - 360 Tage nach Anschreiben
Jahr 2 nach Anschreiben	=	Zeitraum 360 - 720 Tage nach Anschreiben

Analog zur Analyse der Matching-Qualität werden zur Beurteilung der Güte des Programms sowohl numerische als auch grafische Methoden genutzt. Die Tabellen geben jeweils die Mittelwerte der Teilnehmer und der Zwillinge sowie die Differenz zwischen den Mittelwerten und den p-Wert als Ergebnis eines t-Tests an.

Leistungsausgaben:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	10.338	10.312	26	0,508
	2. Jahr	9.548	8.659	889	0,764

Stationär	1 Jahr	3.767	3.963	- 196	0,413
	2. Jahr	4.060	2.636	1.424	0,939

Ambulant	1 Jahr	5.633	5.322	311	0,657
	2. Jahr	4.725	4.985	- 260	0,336

Tab. 7: Leistungsausgaben nach Anschreiben (Mittelwerte)

Teilnehmer haben sowohl im ersten als auch im zweiten Jahr nach Programmteilnahme durchschnittlich höhere, jedoch nicht signifikant höhere Gesamtkosten (+ 26 Euro / + 889 Euro) als deren Zwillinge. Ebenfalls nicht signifikant höher sind die stationären Leistungsausgaben im zweiten Jahr (+ 1.424 Euro) sowie die ambulanten Leistungsausgaben im ersten Jahr nach Programmteilnahme (+ 311 Euro). Lediglich die stationären Kosten im ersten Jahr (- 196 Euro) sowie die ambulanten Kosten im zweiten Jahr nach Teilnahme (- 260 Euro) sind geringer, jedoch ebenfalls nicht signifikant.

Im zweiten Jahr nach Programmteilnahme bzw. Anschreiben sind die Kosten in allen Bereich sowohl bei Teilnehmern als auch bei deren Zwillingen geringer als im ersten Jahr. Lediglich die stationären Kosten der Teilnehmer bilden eine Ausnahme. Im ersten Jahr ist der Mittelwert der stationären Leistungen der Teilnehmer geringer als der Mittelwert dieser Leistungen der Zwillinge (- 196 Euro). Im zweiten Jahr liegen die durchschnittlichen Kosten der Teilnehmer (4.060 Euro) jedoch deutlich über den Kosten der Zwillinge (2.636 Euro). Die Mittelwertdifferenz in Höhe von 1.424 Euro begründet sich zum einen darin, dass der Mittelwert der Teilnehmer vom ersten zum zweiten Jahr leicht ansteigt und zum anderen darin, dass der Mittelwert der Zwillinge im gleichen Zeitraum erheblich sinkt.

Im Folgenden werden zur Analyse der Mittelwertveränderung die QQ-Plots der stationären Leistungen im ersten und zweiten Jahr nach Programmteilnahme/Anschreiben dargestellt:

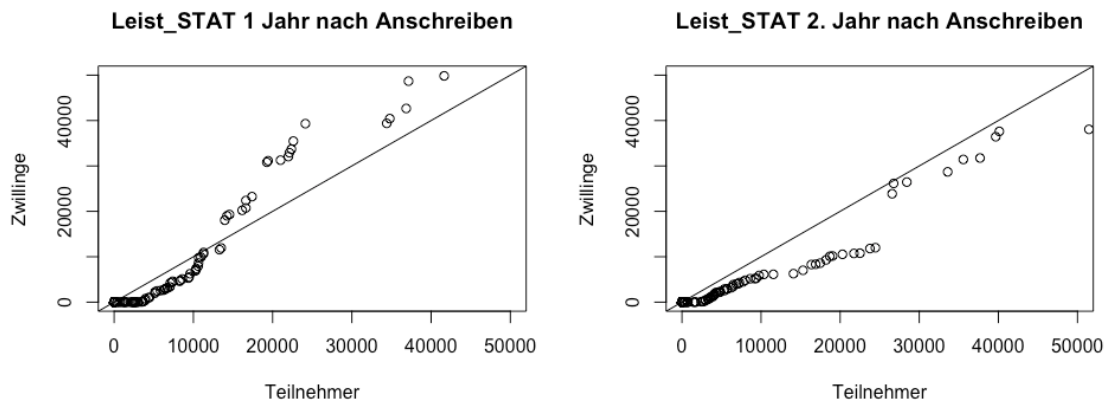


Abb. 17: QQ-Plots Programm-Qualität: Leistungen Stationär

In Jahr 1 nach Programmteilnahme sind die Leistungsausgaben der Teilnehmer im Bereich bis 12.000 Euro höher als die der Zwillinge und liegen daher unterhalb der 45 Grad Linie. Die Kosten der Zwillinge hingegen sind im Bereich ab 15.000 Euro höher als die der Teilnehmer, was auch durch Betrachtung der Maxima deutlich wird: Das Maximum der Teilnehmer liegt bei 50.699 Euro, wohingegen das Maximum der Zwillinge bei 78.056 Euro liegt. Zusammengefasst bedeutet dies, dass der Großteil der Teilnehmer in einem Bereich bis zu 12.000 Euro höhere stationäre Kosten als der Großteil der Zwillinge verursacht. Einzelne Zwillinge verursachen jedoch sehr hohe Leistungsausgaben. Aufgrund dieser hohen Werte haben die Zwi-

linge verglichen mit den Teilnehmern durchschnittlich größere Leistungsausgaben im stationären Bereich (Teilnehmer: 3.767 Euro; Zwillinge: 3.963 Euro; Differenz: - 196 Euro).

Im zweiten Jahr verändert sich die Verteilung der Leistungsausgaben wider Erwarten zu Lasten der Teilnehmer. Im QQ-Plot des zweiten Jahres liegen sämtliche Quantile in der Teilnehmerhälfte, d. h. die Leistungsausgaben der Teilnehmer sind generell höher als die Leistungsausgaben der Zwillinge. Selbst die Quantile der Leistungsausgaben ab einer Höhe von 30.000 Euro liegen nun unterhalb der Winkelhalbierenden, jedoch nicht weit von dieser entfernt. Folglich sind die Extremwerte in den Leistungsausgaben der Teilnehmer nun geringfügig höher als die der Zwillinge. Das Maximum der Teilnehmer hat sich im Vergleich zum ersten Jahr auf 104.837 Euro fast verdoppelt, wodurch es zum Anstieg des Mittelwertes kommt. Das Maximum der Zwillinge hingegen ist ungefähr gleich geblieben, sodass durch die geringeren Extremwerte der Mittelwert der Zwillinge massiv um 1.327 Euro im Vergleich zum Vorjahr sinkt. Folglich kommt es zu einem beträchtlichen Anstieg der Mittelwertdifferenz auf 1.424 Euro, sodass im stationären Bereich nach numerischer sowie grafischer Betrachtung der Leistungsausgaben auf keine Verbesserung des Gesundheitszustands der Teilnehmer infolge der Teilnahme am DMP zu schließen ist.

Das Ergebnis der ambulanten Leistungen hingegen ist wie erwartet. In Jahr 1 sind die durchschnittlichen Leistungsausgaben der Teilnehmer höher (5.655 Euro) als die der Zwillinge (5.322 Euro), da aufgrund der Programmteilnahme Behandlungskosten anfallen. Sie sind ebenfalls höher als im Jahr vor der Programmteilnahme.⁶¹ Im zweiten Jahr nach Programmteilnahme sinken die durchschnittlichen ambulanten Kosten der Teilnehmer dann deutlich um 908 Euro. Im Vergleich dazu sinken die Kosten der Zwillinge nur um 337 Euro, sodass die Kosten der Teilnehmer durchschnittlich 260 Euro geringer sind als die der Zwillinge. Im Folgenden werden zur Analyse der Mittelwertveränderung ebenfalls die QQ-Plots der ambulanten Leistungsausgaben im ersten sowie im zweiten Jahr dargestellt:

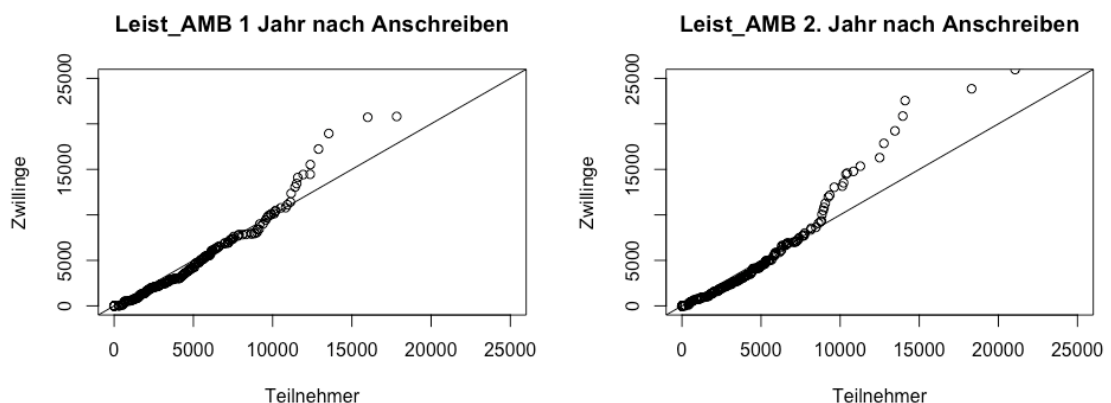


Abb. 18: QQ-Plots Programm-Qualität: Leistungen Ambulant

⁶¹ Vgl. Tab. 17: Mittelwerte nach Matching-Basis-Modell (2)

Die Quantile der Teilnehmer und der Zwillinge im unteren Bereich der Verteilung bis ca. 10.000 Euro stimmen sowohl im ersten als auch im zweiten Jahr weitestgehend überein, sodass sie auf bzw. nahe der Winkelhalbierenden liegen. Abweichungen zwischen den Leistungsausgaben der Teilnehmer und der Zwillinge kommen erst ab einer Höhe von 10.000 Euro vor. Die Gruppe der Zwillinge hat in diesem Bereich in beiden Jahren deutlich höhere Kosten. Nichtsdestotrotz ist der durchschnittliche Mittelwert der Teilnehmer im ersten Jahr höher, was insbesondere am Maximum liegt: Das Maximum der Teilnehmer hat einen Wert von 118.740 Euro, wohingegen das Maximum der Zwillinge lediglich 59.418 Euro beträgt. Im zweiten Jahr hingegen sinkt das Maximum der Teilnehmer auf einen Betrag von 48.251 Euro und das Maximum der Zwillinge steigt auf 80.135 Euro, was dazu führt, dass nun der Mittelwert der Teilnehmer geringer ist.

Zusammenfassend kann festgehalten werden, dass die Kosten der Teilnehmer sowohl im ersten als auch im zweiten Jahr nach Programmteilnahme überwiegend höher sind als die Kosten der Zwillinge. Der erwartete Kostenrückgang im zweiten Jahr wird nur im ambulanten Bereich erzielt. Leistungsausgaben im unteren Bereich der Verteilung sind bei Teilnehmern und Zwillingen identisch, jedoch verursacht die Gruppe der Zwillinge enorm höhere Leistungsausgaben in einem Bereich ab 10.000 Euro. Folglich kann aus diesen Daten geschlossen werden, dass durch das DMP die ambulanten Leistungen in einem hohen Bereich reduziert werden können. Im stationären Bereich ist jedoch auf keine Verbesserung des Gesundheitszustands durch das DMP zu schließen. Zur Validierung dieser Aussagen müssen die Leistungsausgaben eines längeren Zeitraums betrachtet werden.

Um auszuschließen, dass die höheren Kosten lediglich aufgrund anderer Krankheiten verursacht wurden, werden im Folgenden die Leistungsausgaben für F-Diagnosen betrachtet.

Leistungsausgaben für F-Diagnosen:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	1.986	1.367	619	0,924
	2. Jahr	1.655	818	837	0,991

Stationär	1 Jahr	1.217	790	427	0,847
	2. Jahr	1.078	392	686	0,98

Ambulant	1 Jahr	769	577	192	0,987
	2. Jahr	576	427	149	0,974

Tab. 8: Leistungsausgaben F-Diagnosen nach Anschreiben (Mittelwerte)

Im Unterschied zu den Leistungsausgaben für alle Krankheiten sind bei Betrachtung der Leistungsausgaben für Depressionen die durchschnittlichen Leistungsausgaben der Teilnehmer in beiden Jahren und in allen Bereichen deutlich höher als die der Zwillinge. Alle Werte, ausgenommen der Gesamtausgaben im ersten Jahr (p -Wert = 0,924) sowie der stationären Leistungsausgaben im ersten Jahr (p -Wert = 0,847) sind signifikant höher.

Die Kosten im zweiten Jahr nach Programmteilnahme bzw. Anschreiben sind in allen Bereichen verglichen mit den Kosten im ersten Jahr sowohl bei Teilnehmern als auch bei deren Zwillingen geringer. Nichtsdestotrotz steigen im zweiten Jahr die Mittelwertdifferenzen der Gesamtkosten (von 619 Euro auf 837 Euro) sowie der Kosten im stationären Bereich (von 427 Euro auf 686 Euro) an, da die Mittelwerte der Zwillinge im zweiten Jahr deutlich stärker sinken als die der Teilnehmer. Die Teilnahme am Gesundheitsprogramm soll zu einer Kostensenkung im zweiten Jahr führen. Im stationären Bereich ist diese Kostensenkung zwar bei den Teilnehmern zu erkennen, jedoch sinken die Kosten der Zwillinge ebenfalls und das in größerem Maß. Daher werden im Folgenden die QQ-Plots der stationären Leistungsausgaben für F-Diagnosen dargestellt:

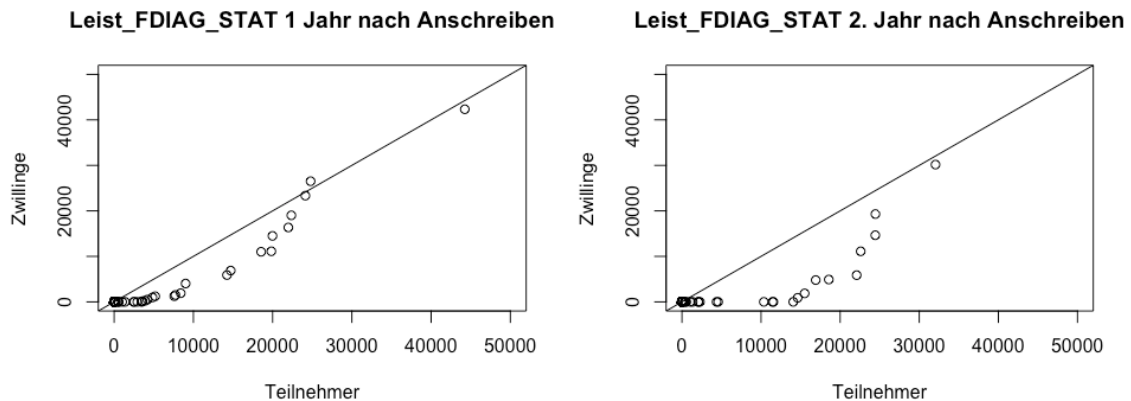


Abb. 19: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär

Der Verlauf der Quantile ist mit Abb. 17: QQ-Plots Programm-Qualität: Leistungen Stationär kaum vergleichbar. In den QQ-Plots der stationären Leistungen für F-Diagnosen sind aufgrund der geringen Anzahl an stationären Depressionsleistungen weitaus weniger Quantile abgetragen als in den QQ-Plots aller stationären Leistungen. Im ersten Jahr fallen bei lediglich 30 Teilnehmern und 20 Zwillingen Leistungen für Depressionen im stationären Bereich an. Diese Anzahl reduziert sich weiter für das zweite Jahr auf 23 Teilnehmer und neun Zwillinge mit Rechnungsbeträgen für stationäre F-Diagnosen.

Trotz der geringen Anzahl an Quantilen ist erkennbar, dass die Kosten der Teilnehmer im ersten und im zweiten Jahr deutlich höher sind als die Kosten der Zwillinge, da fast alle Quantile in der Teilnehmerhälfte liegen. Insbesondere im zweiten Jahr sind die Punkte weit von der Winkelhalbierenden entfernt. Die Maximalwerte von Teilnehmern und Zwillingen sind ähnlich:

Jahr 1:	Teilnehmer: 44.250 Euro	Zwillinge: 42.335 Euro
Jahr 2:	Teilnehmer: 32.055 Euro	Zwillinge: 30.169 Euro,

sodass die höheren Mittelwerte der Teilnehmer nicht auf Ausreißerwerte zurückzuführen sind. Insofern kann zusammenfassend festgehalten werden, dass das DMP zu keiner Reduktion der Leistungsausgaben für stationäre F-Diagnosen führen konnte.

Aus der Analyse der ambulanten Leistungen für alle Leistungsbereiche konnte geschlossen werden, dass sich die ambulanten Leistungen in einem hohen Bereich durch das DMP reduziert haben und infolge der Mittelwert der Teilnehmer im zweiten Jahr geringer als der Mittelwert der Zwillinge ist.

Die Mittelwerte der ambulanten Leistungen für F-Diagnosen sind hingegen sowohl im ersten als auch im zweiten Jahr in der Gruppe der Teilnehmer höher als in der Gruppe der Zwillinge, sodass anhand der numerischen Analyse keine Reduktion der Leistungsausgaben für F-Diagnosen von Teilnehmer durch das DMP hergeleitet werden kann. Zur grafischen Beurteilung werden nachfolgend die QQ-Plots dargestellt:

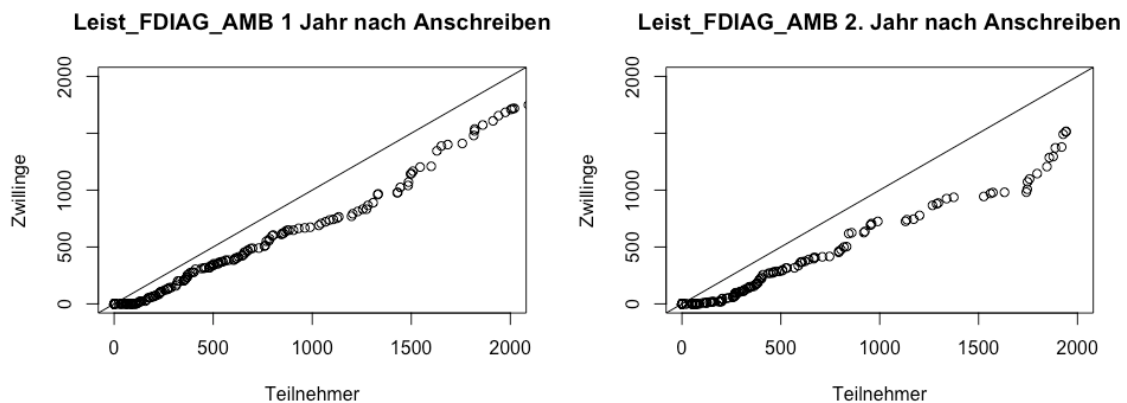


Abb. 20: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant

Anders als bei allen anderen grafischen Analysen von ambulanten Leistungen, liegen die Punkte der ambulanten Leistungen für F-Diagnosen im ersten und zweiten Jahr vollständig in der Teilnehmerhälfte. Im zweiten Jahr entfernen sich die abgetragenen Quantile noch weiter von der Winkelhalbierenden als im ersten Jahr. Die identifizierten Reduktionen der gesamten ambulanten Leistungsausgaben im höheren Bereich durch das DMP gehen nicht mit einer Reduktion der Leistungsausgaben für Depressionen einher.

Die Analyse der Leistungsausgaben für Depressionserkrankungen ergibt, dass die Leistungsausgaben sowohl im stationären als auch im ambulanten Bereich durch das DMP in den zwei Jahren nach Programmteilnahme nicht reduziert werden konnten. Die Kosten der Teilnehmer liegen konstant oberhalb der Kosten der Zwillinge. Aus der Analyse der ambulanten Leistungsausgaben für alle Krankheiten wurde geschlossen, dass durch die Teilnahme

am DMP die ambulanten Leistungen in einem hohen Bereich reduziert werden können. Dieses Ergebnis kann durch die Analyse der ambulanten Leistungen für F-Diagnosen nicht bestätigt werden.

Neben den Leistungsausgaben wird ebenfalls die Anzahl der F-Diagnosen zur Beurteilung der Programm-Qualität herangezogen.

Anzahl F-Diagnosen:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	7	5	2	1
	2. Jahr	5	3	2	0,999

Stationär	1 Jahr	1	1	0	0,8
	2. Jahr	1	0	1	0,866

Ambulant	1 Jahr	4	3	1	0,992
	2. Jahr	3	2	1	0,992

Tab. 9: Anzahl F-Diagnosen nach Anschreiben (Mittelwerte)

Das Krankenversicherungsunternehmen erkennt eine F-Diagnose anhand des Abrechnungsbelegs. Folglich ist eine F-Diagnose immer mit einer Leistungsausgabe verbunden, weshalb die Auswertungen der Leistungsausgaben für F-Diagnosen ähnlich der Auswertung der Anzahl an F-Diagnosen ist. Teilnehmer haben in allen Bereichen eine höhere Anzahl an F-Diagnosen als ihre Zwillinge. Im ersten Jahr ist die Anzahl der stationären F-Diagnosen identisch, was jedoch aus der Rundung resultiert. Bei Betrachtung der ungerundeten Zahlen kann ebenfalls ein erhöhter Wert der Teilnehmer identifiziert werden (Teilnehmer: 0,78; Zwillinge: 0,57).

Die Mittelwerte der Zwillinge sind u. a. geringer, da bei weniger Personen eine F-Diagnose gestellt wird. Beispielsweise haben im zweiten Jahr mehr Personen aus der Gruppe der Zwillinge keine stationäre bzw. ambulante Depression diagnostiziert bekommen als aus der Gruppe der Teilnehmer:

Keine stationäre F-Diagnose:	Teilnehmer: 215	Zwillinge: 230
Keine ambulante F-Diagnose:	Teilnehmer: 82	Zwillinge: 105

Folglich haben die Teilnehmer eine höhere Anzahl an F-Diagnosen als ihre vergleichbaren Zwillinge.

Fazit Programm-Qualität Basis-Modell:

Ziel des DMP ist eine Steuerung der Betroffenen in eine leitliniengerechte Therapie und damit einhergehend die Reduktion der Leistungsausgaben ab dem zweiten Jahr. Die numerische sowie grafische Analyse ergab jedoch, dass die Kosten der Teilnehmer im zweiten Jahr lediglich in den ambulanten Leistungsausgaben für alle Krankheiten geringer sind als die Ausgaben der Zwillinge. Diese Reduktion konnte jedoch nicht auf eine Senkung der Leistungsausgaben für Depressionen zurückgeführt werden. Alle weiteren Leistungsausgaben sowie die Anzahl der F-Diagnosen sind in der Gruppe der Zwillinge geringer, sodass aus den vorliegenden Daten auf keine Reduktion der Leistungsausgaben im zweiten Jahr nach Programmteilnahme geschlossen werden kann.

5 Propensity-Score: Alternativ-Modell 1

Wie zuvor erläutert, müssen Personen mindestens ein halbes Jahr am Gesundheitsprogramm teilnehmen, damit dieses zu einer Reduktion der Leistungsausgaben führt. Während der Teilnahme werden die Personen besonders umfassend betreut. Diese Steuerung in eine leitliniengerechte Behandlung ist mit einer Erhöhung der Leistungsausgaben verbunden. Im Basis-Modell erfolgt die Auswertung der Leistungsausgaben sowie der Anzahl der Depressionsdiagnosen ab dem Anschreibedatum. Für jede Person, die zur Teilnahme am Gesundheitsprogramm eingeladen wurde, wurde das Versanddatum als Zeitpunkt $t = 0$ definiert. Die Kosten der Teilnehmer im ersten Jahr sind jedoch durch die erhöhten Ausgaben für die Programmteilnahme geprägt, sodass die Auswertung des ersten Jahres mit Beginn $t = 0$ nur eingeschränkt aussagekräftig ist. Deshalb wird im Folgenden ein Alternativ-Modell mit einer zeitlichen Abweichung betrachtet.

Das Matching des Alternativ-Modells 1 entspricht dem Matching des Basis-Modells, sodass keine erneute Beurteilung der Güte des PS-Modells durchgeführt wird. Lediglich die Berechnung zur Beurteilung der Programm-Qualität unterscheidet sich. Im Unterschied zum Basis-Modell werden nun die Leistungsausgaben sowie die Anzahl der F-Diagnosen um ein halbes Jahr versetzt berechnet:

Jahr 1 nach Anschreiben	=	Zeitraum 180 - 540 Tage nach Anschreiben
Jahr 2 nach Anschreiben	=	Zeitraum 540 - 900 Tage nach Anschreiben

Durch die zeitliche Verschiebung sollte in der Auswertung der Teilnehmer nun kein Kostenanstieg im ersten Jahr zu erkennen sein, da die Zeit der Programmteilnahme nun nicht mehr in der Auswertung berücksichtigt wird. Vielmehr sollten die Kosten im ersten Jahr bereits geringer sein als die der Zwillinge, da die Teilnehmer durch das DMP in eine leitliniengerechte Behandlung gesteuert wurden.

Im Folgenden werden zur Beurteilung der Programmqualität die gleichen Parameter:

- Allgemeine Leistungsausgaben (Gesamt, Stationär, Ambulant)
- Leistungsausgaben für F-Diagnosen (Gesamt, Stationär, Ambulant)
- Anzahl der F-Diagnosen (Gesamt, Stationär, Ambulant),

wie im Basis-Modell genutzt. Es erfolgt jeweils eine numerische und eine grafische Auswertung sowie ein Vergleich mit den Ergebnissen des Basis-Modells.

Leistungsausgaben:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	10.674	9.495	1.179	0,83
	2. Jahr	8.691	7.939	752	0,732
Stationär	1 Jahr	4.301	3.480	821	0,829
	2. Jahr	3.441	2.401	1.040	0,884
Ambulant	1 Jahr	5.444	5.041	403	0,696
	2. Jahr	4.433	4.509	- 76	0,449

Tab. 10: Leistungsausgaben nach Anschreiben (Mittelwerte) - Alternative 1

Jahr 1:

Entgegen der Annahme sind die Kosten der Teilnehmer, mit Ausnahme der ambulanten Leistungen im zweiten Jahr (- 76 Euro), immer höher als die der Zwillinge, wenn auch nicht signifikant. Durch die zeitliche Verschiebung der Auswertung sollte der natürliche Kostenanstieg in der Gruppe der Teilnehmer während der Teilnahme am DMP unterbunden werden. Nichtsdestotrotz sind die Gesamtkosten und die stationären Kosten der Teilnehmer im ersten Jahr sogar höher als die Kosten im Basis-Modell⁶²:

Teilnehmer Jahr 1:

<u>Gesamt:</u>	Basis-Modell	10.338 Euro	vs.	10.674 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	3.767 Euro	vs.	4.301 Euro	Alternativ-Modell

Im Unterschied zu den gesamten und den stationären Leistungsausgaben sind die ambulanten Kosten der Teilnehmer im ersten Jahr im Alternativ-Modell rund 190 Euro niedriger als im Basis-Modell:

Teilnehmer Jahr 1:

<u>Ambulant:</u>	Basis-Modell	5.633 Euro	vs.	5.444 Euro	Alternativ-Modell
------------------	--------------	------------	-----	------------	-------------------

Im Vergleich dazu sind die Leistungsausgaben der Zwillinge im Alternativ-Modell in allen Bereichen geringer als im Basis-Modell:

⁶² Vgl. Tab. 7: Leistungsausgaben nach Anschreiben (Mittelwerte)

Zwillinge Jahr 1:

<u>Gesamt:</u>	Basis-Modell	10.312 Euro	vs.	9.495 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	3.963 Euro	vs.	3.480 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	5.322 Euro	vs.	5.041 Euro	Alternativ-Modell

Die ambulanten Kosten der Teilnehmer sind zwar im Alternativ-Modell geringer als im Basis-Modell, jedoch sind ebenfalls die ambulanten Kosten der Zwillinge gesunken und zwar prozentual betrachtet stärker. Die ambulanten Leistungsausgaben der Zwillinge im ersten Jahr sind im Alternativ-Modell 5,3 % geringer als im Basis-Modell, wohingegen die Kosten der Teilnehmer nur 3,4 % geringer sind.

Aufgrund der höheren Kosten der Teilnehmer sowie der geringeren Kosten der Zwillinge sind die Mittelwertdifferenzen im Alternativ-Modell im ersten Jahr verglichen mit dem Basis-Modell erheblich angestiegen:

<u>Gesamt:</u>	Basis-Modell	26 Euro	vs.	1.179 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	- 196 Euro	vs.	821 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	311 Euro	vs.	403 Euro	Alternativ-Modell

Dementsprechend kann festgehalten werden, dass trotz Verschiebung der Auswertung um ein halbes Jahr durch die Teilnahme am DMP im stationären Bereich keine unmittelbare Senkung der Leistungsausgaben im ersten Jahr erreicht werden konnte. Auch im ambulanten Bereich kann trotz niedrigerer ambulanter Leistungsausgaben der Teilnehmer im ersten Jahr nach Teilnahme am DMP verglichen mit dem Basis-Modell die Senkung nicht eindeutig auf das DMP zurückgeführt werden. Die Steigerung der Mittelwertdifferenzen stärkt die Aussage, dass das DMP zumindest kurzfristig und unmittelbar nach Programmteilnahme zu keiner Reduktion der Leistungsausgaben führt.

Jahr 2:

Im zweiten Jahr nach Anschreiben sinken die Leistungsausgaben der Teilnehmer und Zwillinge verglichen mit dem ersten Jahr.

Im stationären Bereich sinken die Kosten der Teilnehmer jedoch nicht so stark wie die der Zwillinge, sodass sich die Mittelwertdifferenz weiter zugunsten der Zwillinge verändert (+ 1.040 Euro). Diese hohe Differenz kann mithilfe der QQ-Plots erläutert werden:

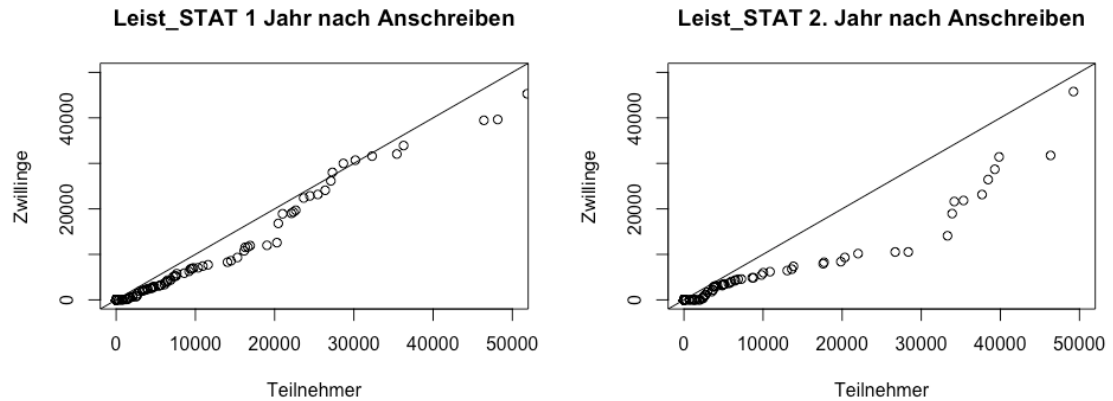


Abb. 21: QQ-Plots Programm-Qualität: Leistungen Stationär - Alternative 1

Bezogen auf das erste Jahr hat das QQ-Plot des Alternativ-Modells kaum eine Ähnlichkeit mit dem QQ-Plot des Basis-Modells. Im Basis-Modell ist das QQ-Plot gezeichnet durch eine Vielzahl an Punkten, die sich ab einem Bereich von 20.000 Euro in der Hälfte der Nicht-Teilnehmer, weit entfernt von der Winkelhalbierenden befinden.⁶³ Im Alternativ-Modell hingegen liegen die Quantile überwiegend in der Hälfte der Teilnehmer, nahe der Winkelhalbierenden. Insbesondere die Quantile der oberen Leistungsbereiche liegen in der Teilnehmerhälfte, weshalb es zum Anstieg der Mittelwertdifferenz im Alternativ-Modell kommt.

Bezogen auf das zweite Jahr hingegen ähneln sich die QQ-Plots des Basis- und des Alternativ-Modells wieder, jedoch mit einer Verschiebung der abgetragenen Quantile zulasten der Teilnehmer im QQ-Plot des Alternativ-Modells. Ab einer Leistungshöhe von 10.000 Euro entfernen sich die Quantile weit von der Winkelhalbierenden, was die Mittelwertdifferenz von 1.040 Euro begründet. Das Maximum der Zwillinge in Höhe von 120.460 Euro im Vergleich zum Maximum der Teilnehmer in Höhe von 57.917 Euro kompensiert eine noch höhere Mittelwertdifferenz geringfügig.

Dementsprechend ergibt auch die um ein halbes Jahr versetzte Analyse der stationären Kosten im zweiten Jahr keine Senkung der Leistungsausgaben durch die Teilnahme am DMP.

Im Unterschied zu den stationären Kosten ist die Senkung der Teilnehmer (- 1.011 Euro) vom ersten zum zweiten Jahr im ambulanten Bereich deutlich größer als die Senkung der Zwillinge (- 532 Euro), wodurch der Mittelwert der Teilnehmer kleiner ist als der Mittelwert der Zwillinge, was eine Mittelwertdifferenz von - 76 Euro ergibt. Dennoch ist die Mittelwertdifferenz verglichen mit der Differenz im Basis-Modell geringer (Basis-Modell: - 260 Euro).⁶⁴ Die QQ-Plots sind in beiden Modellen weitestgehend ähnlich:

⁶³ Vgl. Abb. 17: QQ-Plots Programm-Qualität: Leistungen Stationär

⁶⁴ Vgl. Tab. 7: Leistungsausgaben nach Anschreiben (Mittelwerte)

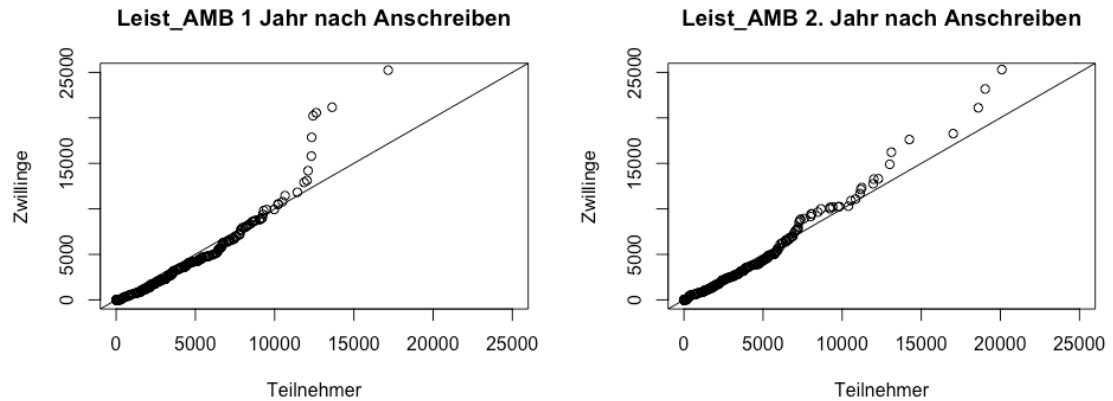


Abb. 22: QQ-Plots Programm-Qualität: Leistungen Ambulant - Alternative 1

Analog zum Basis-Modell liegen die Quantile in beiden Jahren in einem Bereich bis 10.000 Euro weitestgehend auf der Winkelhalbierenden. Oberhalb der 10.000 Euro hingegen verursacht die Gruppe der Zwillinge höhere Leistungsausgaben als die Gruppe der Teilnehmer. Auch im Alternativ-Modell ist der durchschnittliche Mittelwert der Teilnehmer im ersten Jahr höher als der durchschnittliche Mittelwert der Zwillinge, was ebenfalls hauptsächlich am Maximum liegt. Während das Maximum der Teilnehmer bei 126.739 Euro liegt, beläuft sich das Maximum der Zwillinge auf lediglich 68.037 Euro.

Im zweiten Jahr hingegen liegt das Maximum der Teilnehmer bei 41.566 Euro, wohingegen das Maximum der Zwillinge deutlich höher bei 70.257 Euro liegt, was zum höheren Mittelwert der Zwillinge im zweiten Jahr beiträgt.

Die Analyse der ambulanten Leistungsausgaben im Basis-Modell ergab, dass durch die Teilnahme am DMP die ambulanten Leistungen in einem hohen Bereich reduziert werden können. Beim Vergleich der QQ-Plots des Basis- und des Alternativ-Modells fällt jedoch auf, dass die Quantile im Alternativ-Modell im zweiten Jahr in der Hälfte der Zwillinge näher an der Winkelhalbierenden liegen als im Basis-Modell.⁶⁵ Dies führt zu einer geringeren Mittelwertdifferenz im Alternativ-Modell verglichen mit dem Basis-Modell.

Die um ein halbes Jahr erweiterte Analyse des Alternativ-Modells unterstützt die Schlussfolgerung aus der Auswertung der ambulanten Leistungsausgaben im Basis-Modell, dass durch das DMP die ambulanten Leistungen in einem hohen Bereich reduziert werden können. Jedoch ist zu erkennen, dass die Quantile im Alternativ-Modell bereits näher an der Winkelhalbierenden liegen als noch im Basis-Modell, was bedeutet, dass sich die Kosten der Zwillinge und die Kosten der Teilnehmer nicht mehr so stark unterscheiden. Folglich sollte zu Validierungszwecken ein erweiterter Zeitraum betrachtet werden.

⁶⁵ Vgl. Abb. 18: QQ-Plots Programm-Qualität: Leistungen Ambulant

Im Folgenden wird geprüft, ob sich die geringeren ambulanten Leistungen der Teilnehmer im Bereich ab 15.000 Euro auch in den ambulanten Depressionskosten widerspiegeln, oder ob analog zum Basis-Modell diese Beobachtung in den Leistungsdaten für Depressionsdiagnosen nicht nachzuweisen ist.

Leistungsausgaben für F-Diagnosen:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	1.946	1.036	910	0,988
	2. Jahr	1.332	913	419	0,866

Stationär	1 Jahr	1.259	551	708	0,965
	2. Jahr	817	542	275	0,777

Ambulant	1 Jahr	687	485	202	0,993
	2. Jahr	515	371	144	0,98

Tab. 11: Leistungsausgaben F-Diagnosen nach Anschreiben (Mittelwerte) - Alternative 1

Wie der Tabelle zu entnehmen ist, sind sämtliche Mittelwerte der Teilnehmer höher als die Mittelwerte ihrer Zwillinge. Aus der Analyse der Mittelwerte kann folglich auf keine Reduktion der Leistungsausgaben für Depressionserkrankungen durch die Teilnahme am DMP in den 2,5 Jahren nach Programmteilnahme geschlossen werden.

Die dargestellten Kosten im ersten Jahr sind signifikant höher. Im zweiten Jahr hingegen sind die Gesamtkosten sowie die Kosten im stationären Bereich nicht signifikant höher. Lediglich die ambulanten Kosten für F-Diagnosen sind mit einem Unterschied in Höhe von 144 Euro im zweiten Jahr signifikant höher. Der geringere Mittelwert der Teilnehmer im ambulanten Bereich aus der Analyse der Leistungsausgaben für alle ambulanten Krankheiten lässt sich folglich nicht auf eine Reduktion der Ausgaben für ambulante F-Diagnosen zurückführen. Dies konnte bereits im Basis-Modell gesehen werden. Die QQ-Plots der ambulanten Leistungsausgaben für F-Diagnosen sind daher im Basis- und im Alternativ-Modell ähnlich.⁶⁶

Trotz eindeutig höherer Mittelwerte der Teilnehmer gegenüber der Zwillinge im Alternativ-Modell sollte ebenfalls die Veränderung zum Basis-Modell betrachtet werden:

Jahr 1:

Angenommen wurde, dass durch die Verschiebung der Auswertung um ein halbes Jahr die Leistungsausgaben der Teilnehmer insbesondere im ersten Jahr nach Anschreiben geringer

⁶⁶ Vgl. Abb. 20: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant sowie Abb. 37: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant - Alternative 1

sind als im Basis-Modell, da die Kosten für die intensive ärztliche Betreuung der Teilnehmer nicht mehr in die Mittelwerte einfließen. Anders als bei Betrachtung der gesamten Leistungsausgaben führt die Verschiebung tatsächlich dazu, dass die Gesamtausgaben für F-Diagnosen und die ambulanten Depressionsausgaben der Teilnehmer im ersten Jahr geringfügig niedriger (- 40 und - 82 Euro), die stationären Kosten jedoch minimal höher sind als im Basis-Modell (+ 42 Euro).⁶⁷

Teilnehmer Jahr 1:

<u>Gesamt:</u>	Basis-Modell	1.986 Euro	vs.	1.946 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	1.217 Euro	vs.	1.259 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	769 Euro	vs.	687 Euro	Alternativ-Modell

Ähnlich wie bei den Teilnehmern führt die zeitliche Verschiebung im Alternativ-Modell dazu, dass die Kosten der Zwillinge im ersten Jahr in allen Bereichen niedriger sind als im Basis-Modell:

Zwillinge Jahr 1:

<u>Gesamt:</u>	Basis-Modell	1.367 Euro	vs.	1.036 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	790 Euro	vs.	551 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	577 Euro	vs.	485 Euro	Alternativ-Modell

Die Leistungsausgaben der Zwillinge sinken jedoch stärker als die der Teilnehmer, was insbesondere an den Gesamtkosten ersichtlich wird. Während die Gesamtkosten der Teilnehmer nur um 2 % sinken, reduzieren sich die Gesamtkosten der Zwillinge im ersten Jahr um 24,2 %. Infolgedessen vergrößert sich die Mittelwertdifferenz noch weiter im Vergleich zum Basis-Modell.

Jahr 2:

Im zweiten Jahr sind die Kosten der Teilnehmer im Alternativ-Modell deutlich niedriger verglichen mit dem Basis-Modell:

Teilnehmer Jahr 2:

<u>Gesamt:</u>	Basis-Modell	1.655 Euro	vs.	1.332 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	1.078 Euro	vs.	817 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	576 Euro	vs.	515 Euro	Alternativ-Modell

⁶⁷ Vgl. Tab. 8: Leistungsausgaben F-Diagnosen nach Anschreiben (Mittelwerte)

Bei Betrachtung der Leistungsausgaben für F-Diagnosen der Zwillinge ist im zweiten Jahr das Gegenteil zu beobachten. Mit Ausnahme der ambulanten Leistungsausgaben sind die Kosten der anderen Bereiche höher als im Basis-Modell:

Zwillinge Jahr 2:

<u>Gesamt:</u>	Basis-Modell	818 Euro	vs.	913 Euro	Alternativ-Modell
<u>Stationär:</u>	Basis-Modell	392 Euro	vs.	542 Euro	Alternativ-Modell
<u>Ambulant:</u>	Basis-Modell	427 Euro	vs.	371 Euro	Alternativ-Modell

Die ambulanten Kosten der Zwillinge sind die einzigen Kosten, die im zweiten Jahr im Alternativ-Modell geringer sind als im Basis-Modell. Da sich die ambulanten Kosten von Teilnehmern und Zwillingen vom ersten zum zweiten Jahr ungefähr im gleichen Maß verringern, ändert sich die Mittelwertdifferenz im Vergleich zum Basis-Modell kaum (Basis-Modell: 149 Euro; Alternativ-Modell: 144 Euro).

Besonders auffällig hingegen ist die Steigerung der stationären Leistungsausgaben der Zwillinge vom Basis- zum Alternativ-Modell um 38,3 %. In Folge der Senkung der Teilnehmerkosten und der Steigerung der Zwillingskosten im zweiten Jahr ist die Mittelwertdifferenz im Alternativ-Modell sichtbar geringer als im Basis-Modell (Basis-Modell: 686 Euro; Alternativ-Modell: 275 Euro). Dennoch liegt der Mittelwert der Teilnehmer über dem Mittelwert der Zwillinge.

Möglich wäre, dass die Mittelwertsteigerung der stationären Kosten der Zwillinge im zweiten Jahr mit dem typischen Verlauf der Depression zusammenhängt. Personen, die zum Gesundheitsprogramm eingeladen wurden, litten kurz zuvor an einer Depression, sodass die Kosten in dieser Zeit enorm anstiegen.⁶⁸ Auf die depressive Phase folgt anschließend eine „gute“ Phase in der die Leistungsausgaben dementsprechend sinken, was insbesondere daran zu erkennen ist, dass in jeder Auswertung die Kosten im zweiten Jahr geringer sind als die Kosten im ersten Jahr. Denkbar wäre folglich, dass die durchschnittlichen Kosten im Alternativ-Modell im ersten Jahr geringer sind, da die Zwillinge in dieser Zeit gesund sind. Im zweiten Jahr sind sie zwar noch immer geringer als im ersten Jahr, jedoch bereits höher als im Basis-Modell, da im halben Jahr, um das die Auswertung verschoben wurde, einige Personen erneut an einer Depression erkrankten. Die Senkung der Leistungsausgaben in der Gruppe der Teilnehmer wäre hingegen ein Zeichen für den Erfolg des DMPs. Die Mittelwerte steigen im Alternativ-Modell nicht an, sondern sinken in der um ein halbes Jahr versetzten Analyse sogar, was bedeuten kann, dass die Teilnehmer nicht erneut erkranken. Es ist jedoch zu bedenken, dass es sich um Mittelwerte handelt. Im ersten Jahr nach Anschreiben wurden von lediglich 30 Teilnehmern und 16 Zwillingen stationäre Leistungsabrechnungen

⁶⁸ Vgl. Abb. 9: Absolute Leistungsausgaben für F-Diagnosen: Angeschriebene Personen

eingereicht. Noch weniger Personen wurden im zweiten Jahr aufgrund von Depressionen stationär behandelt. Im zweiten Jahr liegen stationäre Rechnungsbeträge für F-Diagnosen von lediglich 21 Teilnehmer sowie 10 Zwillinge vor. Die Mittelwerte sind demzufolge anhand weniger Werte berechnet, was die Aussagekraft reduziert.

Die geringe Anzahl wird ebenfalls anhand der QQ-Plots ersichtlich:

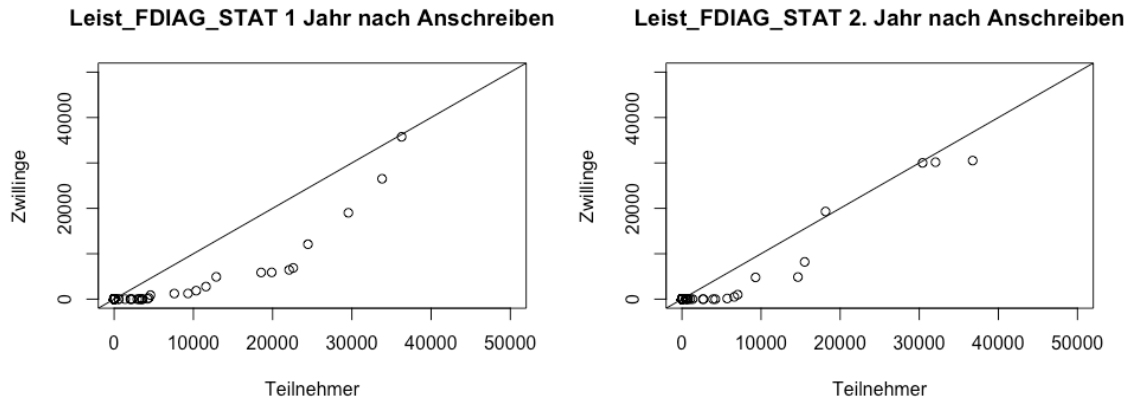


Abb. 23: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär - Alternative 1

Insbesondere im zweiten Jahr kann aus den wenigen Punkten keine genaue Struktur erkannt werden. Aufgrund dessen sollte zur Validierung der Aussage einerseits ein längerer Zeitraum sowie andererseits eine höhere Anzahl an Leistungsabrechnungen bzw. an Teilnehmern betrachtet werden.

Analog zum Basis-Modell kann auch durch die Analyse der Depressionskosten im Alternativ-Modell nicht erkannt werden, dass die Kosten der Teilnehmer im ambulanten oder im stationären Bereich durch das DMP in den 2,5 Jahren nach Einladung geringer werden als die Kosten ihrer Zwillinge. Ebenfalls konnte auch im Alternativ-Modell nicht bestätigt werden, dass sich die ambulanten Leistungen in einem hohen Bereich durch die Teilnahme am DMP reduzieren. Jedoch sollte im Rahmen einer längerfristigen Analyse ausgewertet werden, ob sich der Trend der stationären Kostensteigerung in der Gruppe der Zwillinge und der Kostenreduktion in der Gruppe der Teilnehmer fortsetzt. Eine solche Entwicklung wäre ein Zeichen für den Erfolg des DMP, auch wenn die Kosten der Teilnehmer im Alternativ-Modell über denen der Zwillinge liegen.

Anzahl F-Diagnosen:

		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	6	4	2	1
	2. Jahr	4	3	1	0,999

Stationär	1 Jahr	1	0	1	0,945
	2. Jahr	0	0	0	0,707

Ambulant	1 Jahr	4	3	1	0,988
	2. Jahr	3	2	1	0,994

Tab. 12: Anzahl F-Diagnosen nach Anschreiben (Mittelwerte) - Alternative 1

Die durchschnittliche Anzahl an F-Diagnosen ist im Alternativ-Modell ähnlich wie im Basis-Modell verteilt.⁶⁹ Teilnehmer haben in allen Bereichen und in allen Jahren durchschnittlich mehr F-Diagnosen als die zugeordneten Zwillinge.

Die Analyse der Ausgaben für F-Diagnosen im zweiten Jahr im Alternativ-Modell ergab, dass im Vergleich zum Basis-Modell in der Gruppe der Zwillinge stationäre Kostensteigerungen und in der Gruppe der Teilnehmer Kostensenkungen zu beobachten waren. Daher wird im Folgenden geprüft, ob dies auch in der Anzahl der stationären F-Diagnosen zu erkennen ist. Im stationären Bereich liegen gerundet keine F-Diagnosen bei Teilnehmern und Zwillingen vor, weshalb zur Analyse der Unterschiede in beiden Modellen die ungerundeten Zahlen hinzugezogen werden sollten. Wird die Anzahl der stationären F-Diagnosen im Basis-Modell mit der Anzahl im Alternativ-Modell im zweiten Jahr verglichen, so kann in beiden Gruppen ein geringerer Mittelwert erkannt werden. Teilnehmer und Zwillinge haben bei einer um ein halbes Jahr versetzten Analyse in beiden Jahren eine durchschnittlich geringe Anzahl an stationären F-Diagnosen. Im zweiten Jahr sinkt der Mittelwert der Teilnehmer deutlich stärker als der Mittelwert der Zwillinge:

Stationäre F-Diagnosen Jahr 2:

<u>Teilnehmer:</u>	Basis-Modell	0,57	vs.	0,38	Alternativ-Modell
<u>Zwillinge:</u>	Basis-Modell	0,33	vs.	0,29	Alternativ-Modell

Die Anzahl der stationären F-Diagnosen der Teilnehmer sinkt zwar im Alternativ-Modell, jedoch steigt die durchschnittliche Anzahl der Zwillinge nicht, sodass die Beobachtung aus der Analyse der Kosten für F-Diagnosen nicht gekräftigt werden kann. Jedoch ist an der Anzahl

⁶⁹ Vgl. Tab. 9: Anzahl F-Diagnosen nach Anschreiben (Mittelwerte)

der Personen ohne stationäre F-Diagnose im zweiten Jahr nach Anschreiben erkennbar, dass die Mittelwerte anhand weniger Personen berechnet wurden:

Keine stationäre F-Diagnose:

Basis-Modell:	Teilnehmer:	215	Zwillinge:	230
Alternativ-Modell:	Teilnehmer:	218	Zwillinge:	229

Bei den meisten der 239 Teilnehmer sowie der 239 Zwillinge wurde im zweiten Jahr nach Anschreiben keine stationäre F-Diagnose in den Rechnungsbelegen dokumentiert. Im Alternativ-Modell haben verglichen mit dem Basis-Modell drei weitere Teilnehmer keine F-Diagnose. Im Vergleich dazu verringert sich die Anzahl der Zwillinge ohne F-Diagnose um eine Person.

Fazit Programm-Qualität Alternativ-Modell 1:

Die um ein halbes Jahr versetzte numerische und grafische Analyse ergab analog zum Basis-Modell, dass keine eindeutige Reduktion der Teilnehmerkosten durch das DMP in den zwei Jahren nach Programmteilnahme erreicht werden konnte. Jedoch konnte im Rahmen der Analyse der Depressionskosten in der Gruppe der Zwillinge im Alternativ-Modell verglichen mit dem Basis-Modell ein Anstieg der stationären Kosten im zweiten Jahr beobachtet werden. In der Gruppe der Teilnehmer hingegen sinken die Leistungsausgaben für stationäre F-Diagnosen im zweiten Jahr. Sofern die Daten für einen längerfristigen Zeitraum vorliegen, sollte geprüft werden, ob sich dieser Trend weiter fortsetzt.

6 Propensity-Score: Alternativ-Modell 2

Im Alternativ-Modell 2 erfolgt das Matching anhand der gleichen Parameter wie bereits im Basis-Modell. Jedoch werden nun nicht nur Personen des Jahres X3 betrachtet, sondern alle Personen, die in den Jahren X1, X2 und X3 zur Teilnahme am DMP eingeladen wurden, ausgenommen der Programmabbrecher. Personen, die im Jahr X4 angeschrieben wurden, werden in der Auswertung nicht berücksichtigt, da die Leistungsausgaben bis lediglich März X6 vorliegen. Folglich sind bei den Personen, die zwischen April X4 und Dezember X4 angeschrieben wurden, keine Leistungsdaten für die zwei Jahre nach Anschreiben vorhanden.

Den 625 Teilnehmern stehen im Alternativ-Modell 2 nun 11.226 Nicht-Teilnehmer gegenüber. Durch die Erweiterung des Personenkreises muss ein erneutes Matching durchgeführt werden. Analog zum Basis-Modell muss nach Durchführung des Matchings die Güte des PS-Modells beurteilt werden. Im Anschluss kann die Programmqualität auf Basis aller Programmteilnehmer analysiert werden.

6.1 Analyse der Matching-Qualität

Das Matching erfolgt anhand der identischen Parameter wie bereits im Basis-Modell. Im Folgenden wird die Güte des Matchings anhand der Mittelwertdifferenzen und der prozentualen Veränderung der Mittelwerte sowie bei Bedarf anhand der QQ-Plots beurteilt.

Merkmal	Differenz der Mittelwerte		proz. Änderung
	vor Matching	nach Matching	
ALTER	0,83	0,02	97,45
GESCHLECHT	0,02	0,00	100,00
V_LEIST_ANTIDEP_COUNT	0,53	0,09	83,47
COMORBIDITY	- 0,05	- 0,02	65,56
LAST_FDIAG_AMB	23,23	- 3,57	85,14
LAST_FDIAG_STAT	- 18,22	- 0,57	97,14
LEIST_AMB_360	229,92	90,60	47,27
LEIST_STAT_360	1.172,77	- 288,17	71,74
LEIST_FDIAG_AMB_360	44,30	- 3,95	93,42
LEIST_FDIAG_STAT_360	438,86	- 32,10	91,66
ANZ_FDIAG_AMB_360	0,65	- 0,07	90,36
ANZ_FDIAG_STAT_360	0,29	- 0,01	94,77

Tab. 13: Differenz der Mittelwert vor- und nach Matching – Alternative 2

Differenz der Mittelwerte vor Matching:

Analog zum Basis-Modell unterscheiden sich Teilnehmer und Nicht-Teilnehmer vor Durchführung des Matchings in allen Variablen. Die Gruppe der Teilnehmer ist durchschnittlich fast ein Jahr älter, enthält geringfügig mehr Frauen als Männer und erhält in der Zeit vor Anschreiben mehr Antidepressiva verschrieben (+ 0,53), hat jedoch einen kleineren Komorbiditätsindex als die Gruppe der Nicht-Teilnehmer (- 0,05).

Der Zeitpunkt der letzten ambulanten F-Diagnose liegt bei den Teilnehmern im Vergleich zu den Nicht-Teilnehmern näher am Anschreibedatum (+ 23,23 Tage), der Zeitpunkt der letzten stationären F-Diagnose hingegen weiter weg (- 18,22 Tage).

Die Kosten der Teilnehmer sind in allen Bereichen höher als die der Nicht-Teilnehmer. Im Vergleich zum Basis-Modell ist jedoch zu beobachten, dass die Mittelwerte im Alternativmodell 2 in den beiden Gruppen nicht so weit voneinander entfernt liegen, wie bei Betrachtung der im Jahr X3 angeschriebenen Personen im Basis-Modell.⁷⁰ Die Mittelwertdifferenz der ambulanten Kosten für F-Diagnosen in Höhe von 44,30 Euro ist sehr gering, wohingegen die Mittelwertdifferenz der stationären Gesamtkosten mit 1.172,77 Euro sehr hoch ist, aber dennoch niedriger verglichen mit dem Basis-Modell.

Die höheren Kosten der Teilnehmer spiegeln sich ebenfalls in der Anzahl der F-Diagnosen im Jahr vor Anschreiben wider. Ambulant haben Teilnehmer durchschnittlich 0,65 und stationär durchschnittlich 0,29 F-Diagnosen mehr als Nicht-Teilnehmer.

Differenz der Mittelwerte nach Matching:

Die Mittelwertdifferenzen haben sich in Folge des Matchings überwiegend stark reduziert. Die Teilnehmer und die zugeordneten Zwillinge sind nach Matching fast gleich alt (+ 0,02). Bezogen auf das Geschlecht sind beide Gruppen gleich. Jede Gruppe enthält 356 Männer und 269 Frauen. Die Anzahl der verschriebenen Antidepressiva (+0,09) sowie der Komorbiditätsindex (- 0,02) sind nach Matching in beiden Gruppen so gut wie gleichverteilt.

Im Hinblick auf die Zeitpunkte der letzten F-Diagnosen sind beide Gruppen vergleichbar geworden, mit prozentualen Änderungen über 85 %. Der Zeitpunkt der letzten ambulanten F-Diagnose liegt in der Gruppe der Zwillinge durchschnittlich 3,57 Tage und der Zeitpunkt der letzten stationären F-Diagnose durchschnittlich 0,57 Tage weiter in der Vergangenheit als in der Gruppe der Teilnehmer.

Als einziger Parameter liegt die Änderung der Mittelwertdifferenz der ambulanten Gesamtkosten unter 50 %. Trotz der geringen prozentualen Änderung beläuft sich die Mittelwertdifferenz nach Matching auf nur 90,60 Euro. Der Mittelwert der Teilnehmer nach Matching beträgt 4.903 Euro und der Mittelwert der Zwillinge 4.812 Euro. Verglichen mit den hohen Mittelwerten führt die Mittelwertdifferenz trotz der prozentualen Änderung von unter 50 % nicht zu einer schlechten Matching-Qualität.

⁷⁰ Vgl. Tab. 6: Differenz der Mittelwert vor- und nach Matching – Basis-Modell (2)

Besonders reduziert hat sich die Mittelwertdifferenzen der stationären Gesamtkosten. Vor Durchführung des Matchings unterschieden sich die Mittelwerte der beiden Gruppen deutlich. Diese hohe Mittelwertdifferenz von 1.172,77 Euro konnte auf - 288,17 Euro verringert werden. Ein Unterschied ist dennoch gegeben, allerdings nicht mehr so ausgeprägt, wie vor Matching. Nach Matching verursachen die den Teilnehmern zugeordneten Zwillinge durchschnittlich stationäre Mehrkosten in Höhe von 288 Euro im Jahr vor Anschreiben. Zur Verdeutlichung werden im Folgenden die QQ-Plots dargestellt:

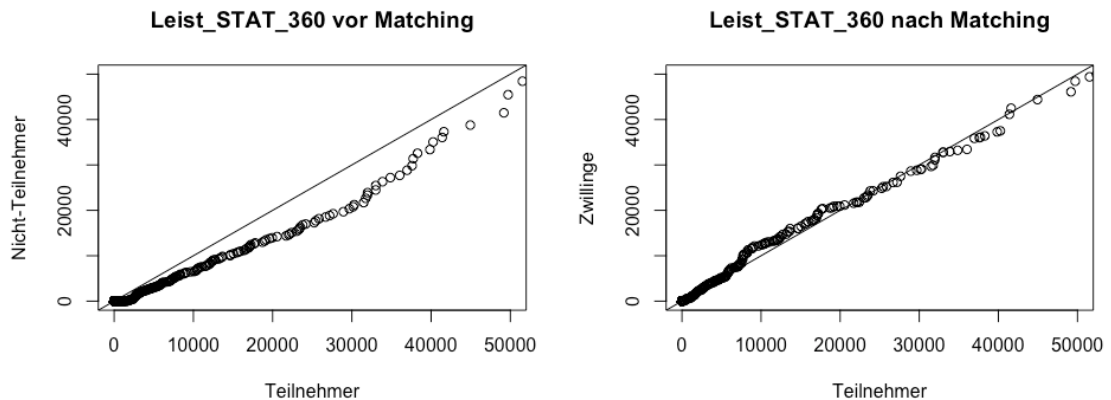


Abb. 24: QQ-Plots Matching-Qualität: Leistungen Stationär - Alternative 2

Die Verringerung der Mittelwertdifferenz ist ebenfalls bei Betrachtung der QQ-Plots ersichtlich. Vor Matching liegen alle Quantile in der Teilnehmerhälfte. Ab einem Bereich von 20.000 Euro entfernen sich die Quantile zudem deutlich von der Winkelhalbierenden. Nach Matching liegen weitestgehend alle Punkte auf bzw. nahe der Winkelhalbierenden. Einige Quantile befinden sich rund um eine Leistungshöhe von 10.000 Euro in der Hälfte der Zwillinge. Folglich ist der Mittelwert der Zwillinge geringfügig höher als der der Teilnehmer.

Die Mittelwertdifferenzen der Leistungsausgaben für F-Diagnosen und der Anzahl der F-Diagnosen haben sich sowohl im ambulanten als auch im stationären Bereich mit prozentualen Änderungen von jeweils über 90 % deutlich verringert. Teilnehmer und Zwillinge sind in diesen Parametern vergleichbar geworden.

Fazit Matching-Qualität Alternativ-Modell 2:

Durch das Matching konnte der Gruppe der Teilnehmer eine Gruppe an Zwillingen gegenübergestellt werden, die in ihren Merkmalen weitestgehend ähnlich ist. Das Zeichen auf ein schlechtes Matching aufgrund der geringen prozentualen Änderungen des Merkmals „LEIST_AMB_360“ bzw. der trotz Matching hohen Mittelwertdifferenz des Merkmals „Leist_STAT_360“ konnte durch Analyse der Mittelwerte sowie des QQ-Plots entkräftet werden. Deshalb kann die Güte des PS-Modells als gut eingestuft werden, sodass im Folgenden mit der Beurteilung der Programm-Qualität fortgefahren werden kann.

6.2 Analyse der Programm-Qualität

Die Analyse der Programm-Qualität anhand einer höheren Anzahl von Teilnehmern und Zwillingen ergibt ebenfalls keine Kostenreduktion infolge der Teilnahme am DMP.

Leistungsausgaben					
		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	9.006	8.100	906	0,889
	2. Jahr	7.956	7.489	467	0,763

Stationär	1 Jahr	3.114	2.786	328	0,72
	2. Jahr	2.609	2.595	14	0,511

Ambulant	1 Jahr	5.003	4.613	390	0,875
	2. Jahr	4.498	4.147	351	0,865

Leistungsausgaben für F-Diagnosen					
		Teilnehmer	Zwillinge	Differenz	p-Wert
Gesamt	1 Jahr	1.393	971	422	0,985
	2. Jahr	1.092	1.094	- 2	0,498

Stationär	1 Jahr	697	415	282	0,94
	2. Jahr	585	713	- 128	0,302

Ambulant	1 Jahr	697	557	140	0,994
	2. Jahr	506	372	134	0,999

Tab. 14: Beurteilung der Programm-Qualität - Alternativ-Modell 2

Bei Betrachtung der Leistungsausgaben für alle Krankheiten haben die Teilnehmer immer höhere Ausgaben als die Zwillinge. Im zweiten Jahr sinken die Kosten sowohl bei Teilnehmern als auch bei Zwillingen im Vergleich zum ersten Jahr. Dabei sinken die Kosten der Teilnehmer stärker als die Kosten der Zwillinge, wodurch sich die Mittelwertdifferenzen reduzieren.

Als wichtiges Zeichen für den Erfolg des DMP gelten die Ausgaben für Depressionen. Durch das DMP sollen diese Kosten langfristig gesenkt werden. Im ersten Jahr liegen die Kosten der Teilnehmer immer oberhalb der Kosten der Zwillinge. Wie bereits erläutert, ist dieser Anstieg aufgrund der Teilnahmekosten zu erwarten, sodass die Kosten der Teilnehmer frühestens im zweiten Jahr nach Teilnahme unterhalb der Kosten der Zwillinge liegen. Entgegen des Ergebnisses aus dem Basis-Modell sinken die durchschnittlichen stationären Kosten der Teilnehmer im zweiten Jahr tatsächlich so stark, dass der Mittelwert unterhalb des Mit-

telwerts der Zwillinge liegt. Der Unterschied ist zwar nicht signifikant, aber dennoch hoch (-128 Euro). Der Mittelwert der Zwillinge ist im zweiten Jahr höher als der Mittelwert der Teilnehmer, da die Gruppe der Zwillinge höhere Kosten in einem Bereich ab 20.000 Euro verursacht als die Gruppe der Teilnehmer.

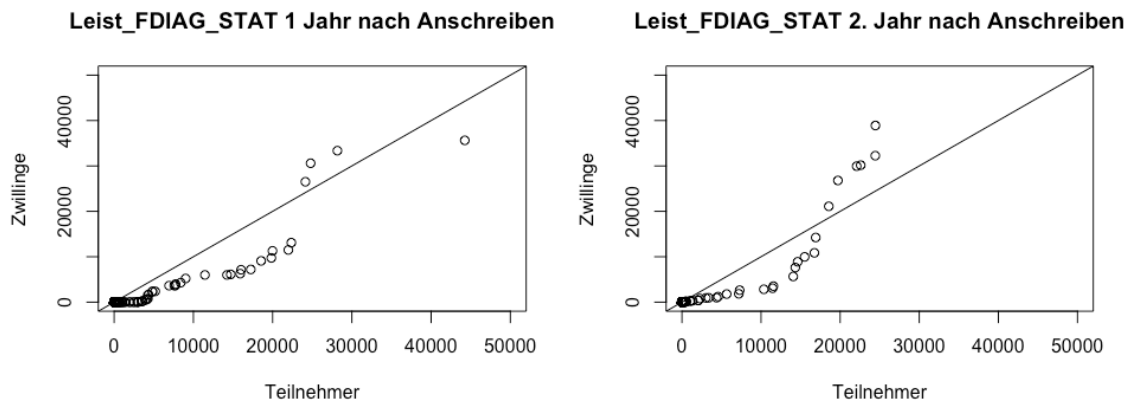


Abb. 25: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Stationär - Alternative 2

Im ersten Jahr nach Anschreiben liegen die meisten Quantile in der Teilnehmerhälfte, woraus der höhere Mittelwert der Teilnehmer folgt. Wenige Punkte sind zwar in der Zwillingshälfte zu finden, jedoch liegen diese nahe der Winkelhalbierenden.

In Jahr 2 ist ein Großteil der Quantile weiterhin in der Teilnehmerhälfte vorzufinden. Die Quantile im Bereich ab 20.000 Euro liegen jedoch deutlich in der Zwillingshälfte. Zudem ist der Maximalwert der Zwillinge mit 67.264 Euro mehr als doppelt so hoch wie der Maximalwert der Teilnehmer in Höhe von 32.055 Euro. Dies führt dazu, dass der Mittelwert der Teilnehmer geringer ist als der Mittelwert der Zwillinge. Folglich liegen die Kosten der Teilnehmer nicht immer unterhalb der Kosten der Zwillinge, sodass daraus nicht geschlossen werden kann, dass das DMP erfolgreich war. Es wäre jedoch möglich, dass durch die Teilnahme am DMP stationäre Kosten im hohen Bereich reduziert werden können. Diese Erkenntnis sollte durch die Betrachtung aller Teilnehmer und durch die Analyse eines längerfristigen Zeitraums validiert werden.

Im ambulanten Bereich sind die Teilnehmerkosten in beiden Jahren signifikant höher als die Zwillingskosten (Jahr 1: + 140 Euro; Jahr 2: + 134 Euro). Trotz Teilnahme am DMP konnten die durchschnittlichen ambulanten Kosten der Teilnehmer nicht unterhalb der Kosten der Zwillinge sinken. Besonders gut erkennbar ist dies auch im QQ-Plot:

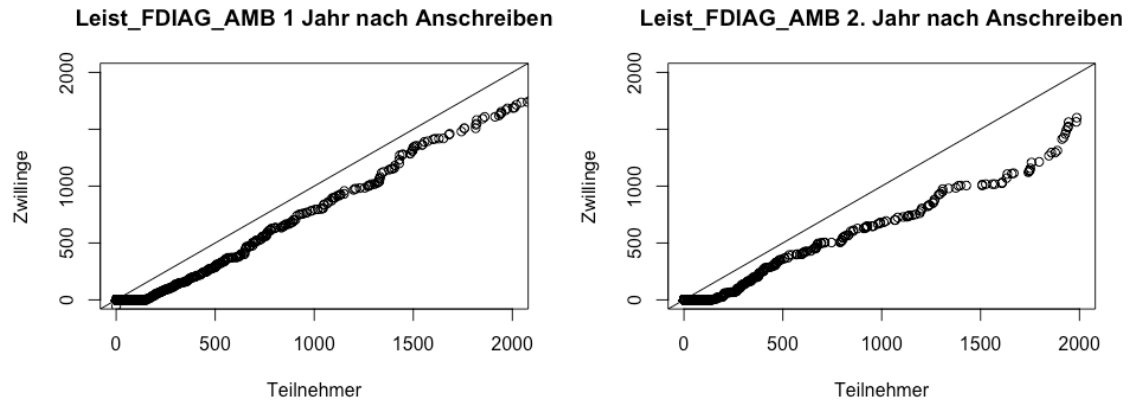


Abb. 26: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant - Alternative 2

Sowohl in Jahr 1 als auch in Jahr 2 nach Anschreiben liegen alle Quantile in der Hälfte der Teilnehmer. Insbesondere im zweiten Jahr entfernen sich diese weit von der Winkelhalbierenden.

Fazit Programm-Qualität Alternativ-Modell 2:

Im Gegensatz zum Basis-Modell erfolgt die Auswertung im Alternativ-Modell 2 anhand einer deutlich höheren Anzahl an Personen. Das Ergebnis zeigt, dass die Leistungskosten der Teilnehmer bezogen auf alle Krankheiten immer oberhalb der Kosten der Zwillinge liegen. Werden lediglich die Kosten F-Diagnosen betrachtet, sieht das Ergebnis leicht anders aus. Die stationären Kosten für F-Diagnosen im zweiten Jahr sind in der Gruppe der Teilnehmer durchschnittlich geringer als in der Gruppe der Zwillinge. Die Analyse der QQ-Plots ergibt, dass die Kosten der Teilnehmer in einem hohen Bereich reduziert werden konnten. Diese Erkenntnis sollte jedoch durch weitere Analysen validiert werden.

Die ambulanten Depressionskosten hingegen sind in der Teilnehmergruppe in beiden Jahren signifikant höher als in der Zwillingungsgruppe.

7. Fazit

Zur Messung des Erfolgs eines Gesundheitsprogramms sollte dieses regelmäßig evaluiert werden. Üblicherweise wird der Erfolg von Therapien in der medizinischen Forschung mittels randomisierter kontrollierter Studien geprüft. Jedoch kann diese Methode nicht im Kontext von DMPs zur Anwendung kommen. Eine Alternative für die statistische Auswertung ist die Propensity-Score-Methode, bei der es sich um eine nichtrandomisierte Studie handelt. Im Rahmen dieser wird zunächst eine Vergleichbarkeit zwischen Teilnehmern am DMP (Interventionsgruppe) und Nicht-Teilnehmern (Kontrollgruppe) hergestellt. Nur wenn die Gruppen miteinander vergleichbar sind, kann im Anschluss die Qualität des Programms beurteilt werden.

Ziel dieser Arbeit war es, zum einen geeignete Merkmale zu finden, anhand derer die beiden Gruppen gematcht werden, sodass die Matching-Qualität gut ist. Zum anderen sollte anschließend nach erfolgreichem Matching die Programm-Qualität beurteilt werden.

Für das Matching wurden zwölf Merkmale gefunden anhand derer das Matching durchgeführt wurde und die zu einer guten Matching-Qualität führten:

- Alter
- Geschlecht
- Anzahl an verschriebenen Antidepressiva
- Komorbiditätsindex
- Letzte ambulante sowie letzte stationäre F-Diagnose
- Gesamtkosten im ambulanten sowie im stationären Bereich
- Kosten für F-Diagnosen im ambulanten sowie im stationären Bereich
- Anzahl an ambulanten sowie stationären F-Diagnosen

Für das Kriterium „Geschlecht“ wurde ein exaktes Matching gewählt, da eine Depression zwischen den Geschlechtern unterschiedlich ausgeprägt ist. Alle weiteren Kriterien wurden mittels der „nearest neighbor“- Methode gematcht. Die numerische Analyse anhand von Mittelwerten sowie die grafische Analyse mittels QQ-Plots ergab eine gute Matching-Qualität. Durch das Matching wurde der Teilnehmergruppe eine Gruppe an Zwillingen gegenübergestellt, die in ihren Merkmalen ähnlich ist.

Auf Basis dieses Matchings konnte anschließend die Programm-Qualität beurteilt werden. Analog zur Beurteilung der Matching-Qualität erfolgte die Analyse der Programm-Qualität numerisch sowie grafisch. Es wurden drei verschiedene Modelle betrachtet, wobei zur Beurteilung der Programm-Qualität jeweils die Gesamtkosten im ambulanten sowie im stationären Bereich, die Kosten für F-Diagnosen im ambulanten sowie im stationären Bereich und die Anzahl an ambulanten sowie stationären F-Diagnosen analysiert wurden. Im Basis-Modell erfolgte die Auswertung anhand der Personen, die im Jahr X3 zum Programm eingeladen wurden. Dabei wurden die beiden Jahre nach Versanddatum der Einladung separat betrachtet, wobei hauptsächlich das zweite Jahr relevant ist, da im ersten Jahr mit einem

Kostenanstieg in der Teilnehmergruppe aufgrund der Kosten für die Teilnahme am DMP zu rechnen ist. Im Alternativ-Modell 1 erfolgte die identische Auswertung wie im Basis-Modell nur um ein halbes Jahr versetzt, um so die Kosten der Programmteilnahme zu eliminieren. Zuletzt wurde die Auswertung analog des Basis-Modells, jedoch unter Einbeziehung der im Jahr X1, X2 und X3 zum Programm eingeladenen Personen, ausgenommen der Programmabbrecher, durchgeführt.

Grundsätzlich führten alle drei Analysen zu einem weitestgehend einheitlichen Ergebnis. Durch die Teilnahme am DMP und die damit verbundene Steuerung der Teilnehmer in eine leitliniengerechte Therapie sinken die Kosten der Teilnehmer zumindest kurzfristig – wie erwartet – nicht unter die Kosten der Zwillinge.

Im Basis-Modell liegen lediglich die stationären Leistungsausgaben der Teilnehmer für alle Krankheiten im ersten Jahr sowie die ambulanten Kosten der Teilnehmer für alle Krankheiten im zweiten Jahr unterhalb der Ausgaben der Zwillinge, wobei diese Reduktion nicht auf einen Rückgang der Leistungsausgaben für Depressionen zurückzuführen ist. Alle weiteren Leistungsausgaben sowie die Anzahl der F-Diagnosen sind in der Gruppe der Teilnehmer höher als in der Zwillingsgruppe.

Auch die um ein halbes Jahr versetzten Analyse ergibt, mit Ausnahme der ambulanten Leistungsausgaben für alle Krankheiten, höhere Kosten der Teilnehmer im Vergleich zu den Zwillingen. Allerdings ergibt die Analyse der Leistungsausgaben für F-Diagnosen im Alternativ-Modell in der Gruppe der Zwillinge verglichen mit dem Basis-Modell einen Anstieg der stationären Kosten für F-Diagnosen im zweiten Jahr. In der Gruppe der Teilnehmer hingegen sinken die Leistungsausgaben für stationäre F-Diagnosen im gleichen Zeitraum. Dieser Trend kann weiter analysiert werden, wenn die Daten für einen längerfristigen Zeitraum vollständig vorliegen.

Die Erhöhung der Anzahl an Personen in der Interventions- und Kontrollgruppe im Alternativ-Modell 2 ergibt eine Senkung der stationären Kosten für F-Diagnosen in der Gruppe der Teilnehmer im zweiten Jahr, sodass die stationären Kosten für F-Diagnosen im zweiten Jahr unterhalb der stationären Kosten für F-Diagnosen in der Gruppe der Zwillinge fallen. Grund dafür ist die Reduktion der Teilnehmerkosten ab einer Leistungshöhe von 20.000 Euro. Auch diese Erkenntnis kann zukünftig durch weitere Analysen validiert werden.

Anhand der vorhandenen Leistungsabrechnungen konnte in den zwei Jahren nach Programmteilnahme – wie erwartet – keine signifikante Kosteneinsparung durch die Teilnahme am DMP identifiziert werden. Sobald die Leistungsdaten für einen längeren Zeitraum vorliegen, kann die Analyse des Alternativ-Modells 2 erneut durchgeführt werden. In diesem Modell sollten dann alle DMP-Teilnehmer betrachtet werden, wodurch die Ergebnisse an Validität gewinnen. Zu betrachten sind dann die Ergebnisse ab dem zweiten Jahr, da es in-

folge der DMP-Teilnahme im ersten Jahr zu erwarteten Kostensteigerungen in der Gruppe der Teilnehmer kommt.

Zudem könnten weitere wirtschaftliche Kennzahlen in die Auswertung einbezogen werden. Möglich wäre u. a. die Einbeziehung der stationären Aufenthalte aufgrund von F-Diagnosen sowie die Dauer dieser Aufenthalte. Sofern weitere Kriterien zur Analyse der Programm-Qualität einbezogen werden sollen, wäre eine gleichzeitige Berücksichtigung im Matching sinnvoll, wenn sich dadurch die Matching-Qualität nicht verschlechtert.

Anhang

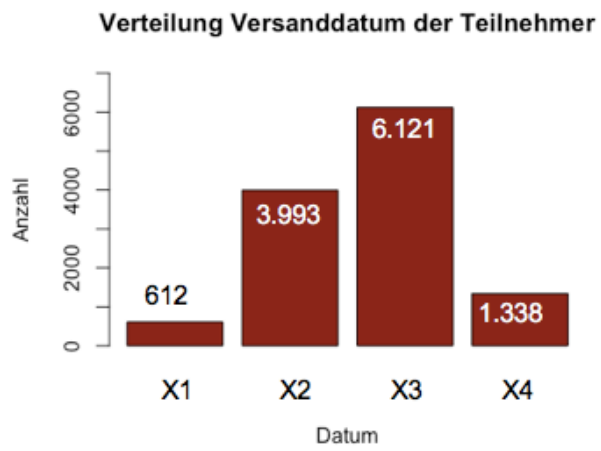


Abb. 27: Verteilung Versanddatum der Teilnehmer

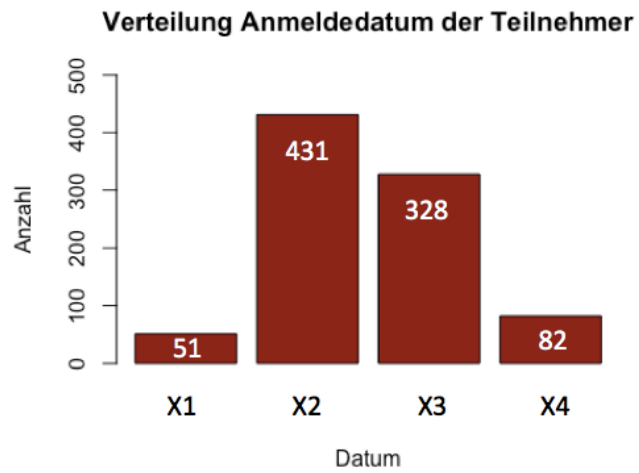


Abb. 28: Verteilung Anmeldedatum der Teilnehmer

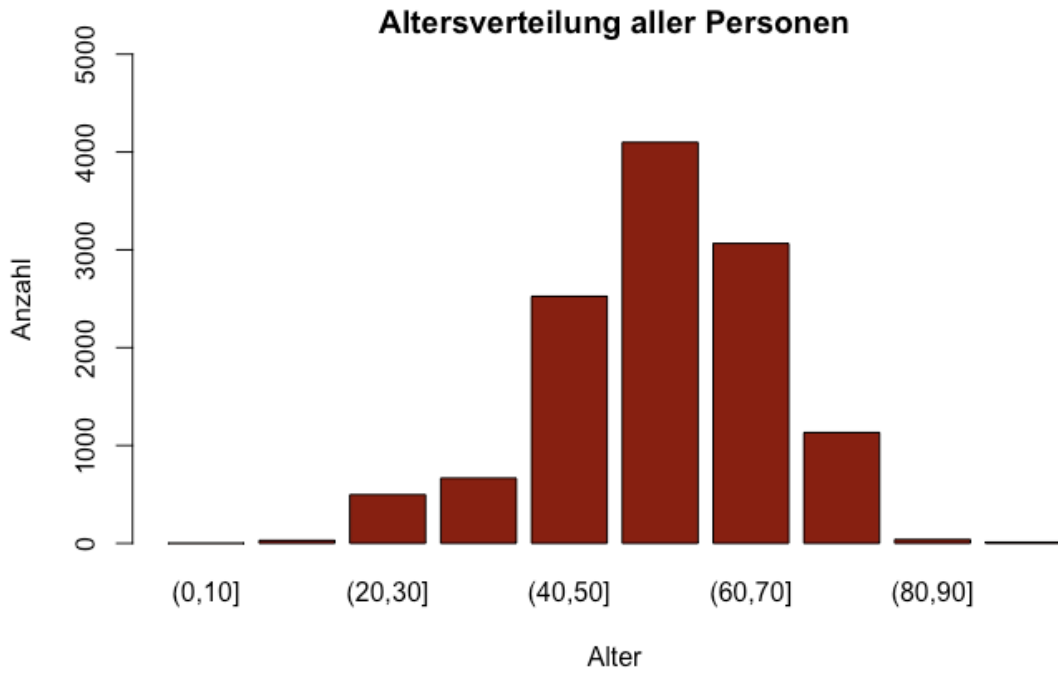


Abb. 29: Altersverteilung aller Personen

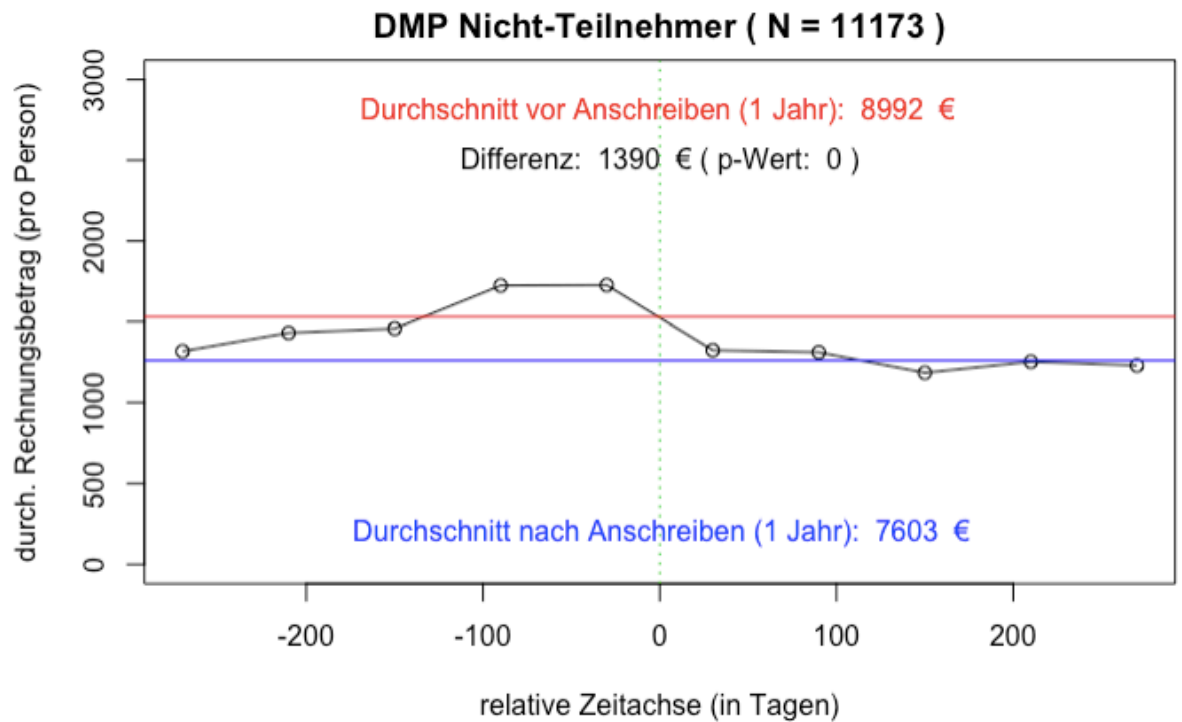


Abb. 30: Durchschnittliche Leistungsausgaben der DMP Nicht-Teilnehmer

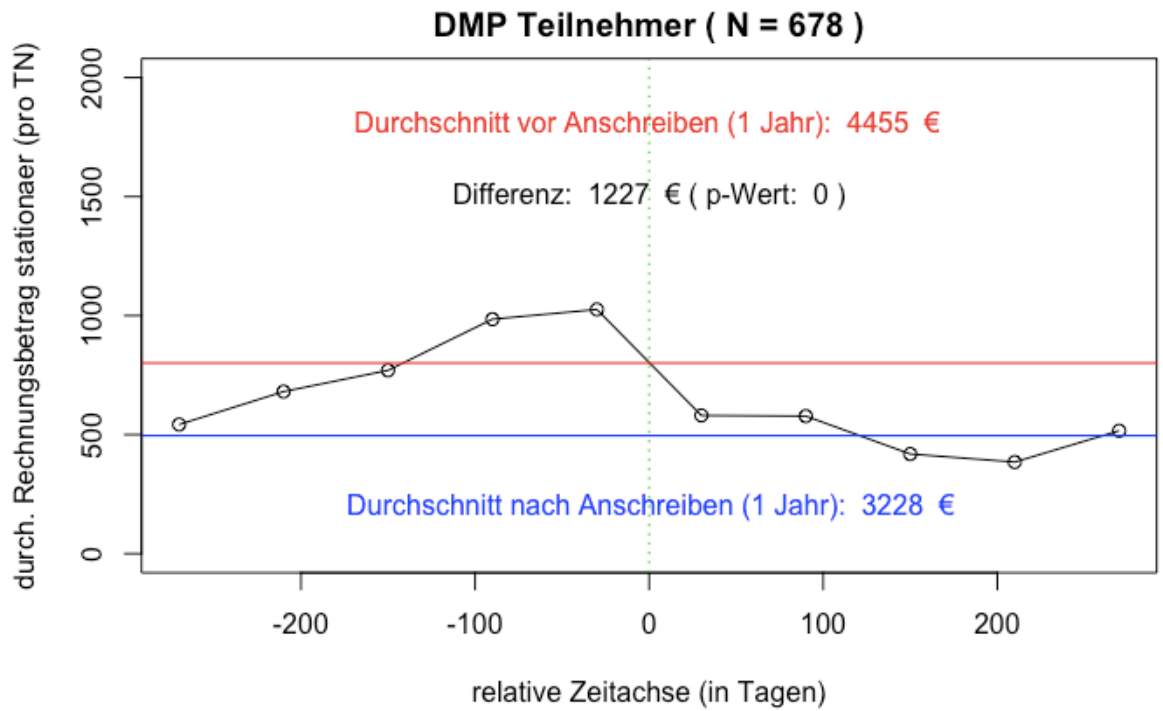


Abb. 31: Durchschnittliche stationäre Leistungsausgaben der DMP Teilnehmer

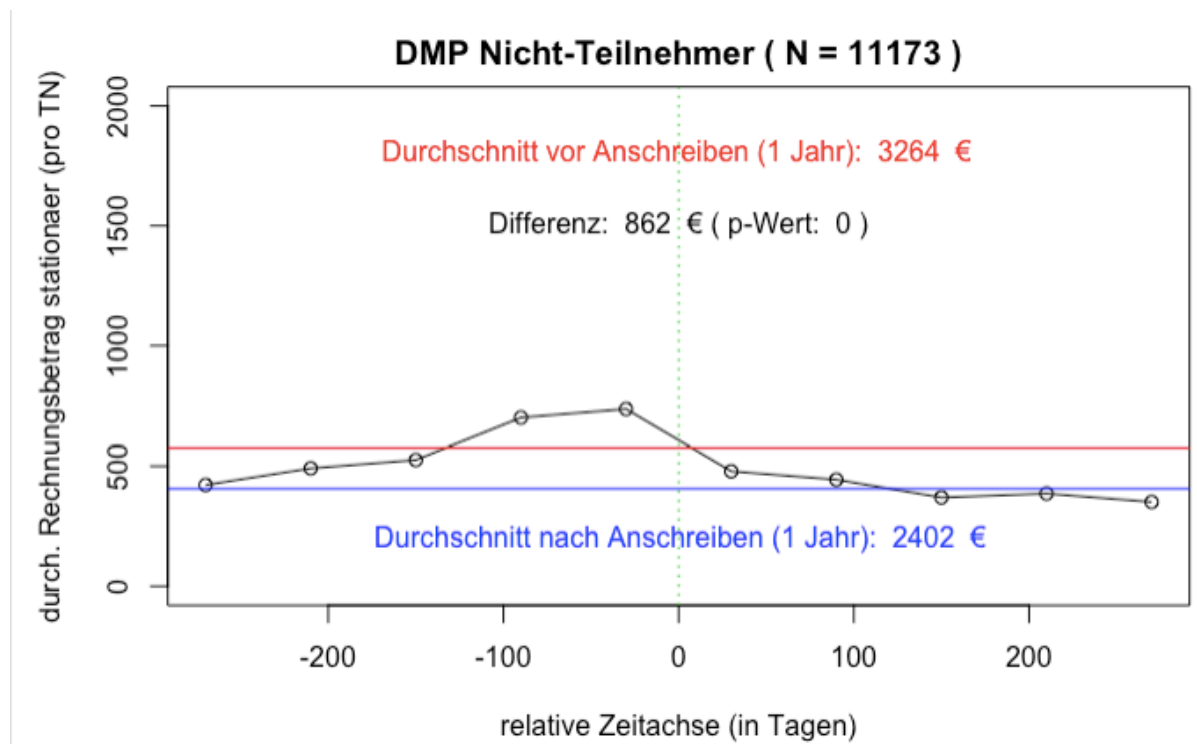


Abb. 32: Durchschnittliche stationäre Leistungsausgaben der DMP Nicht-Teilnehmer

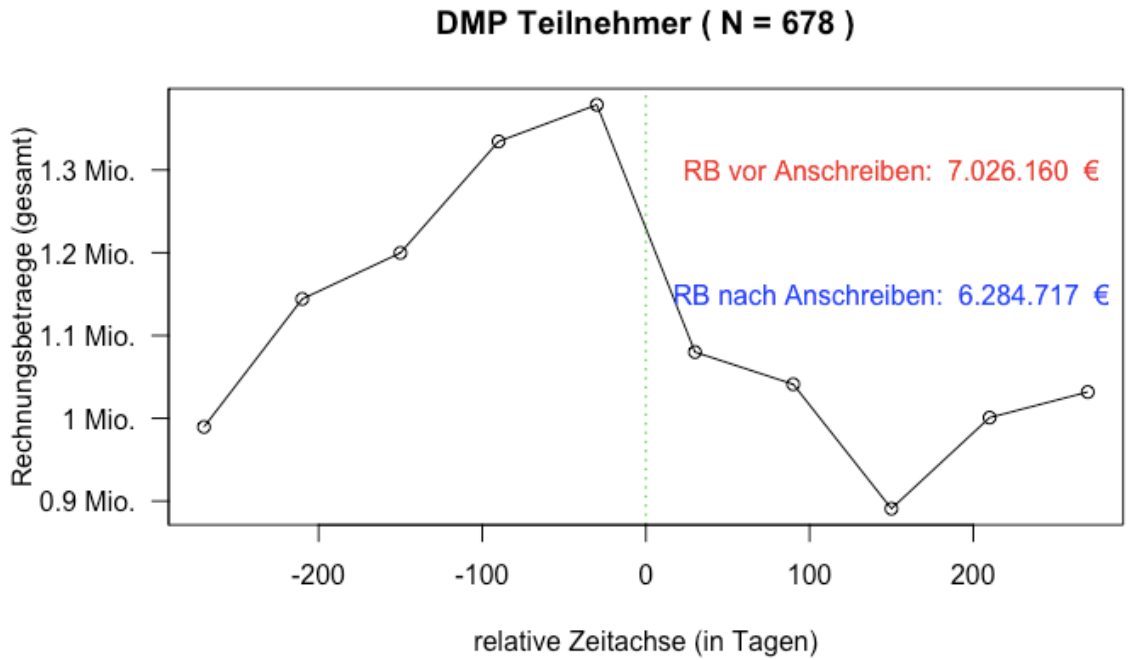


Abb. 33: Absolute Leistungsausgaben der DMP-Teilnehmer

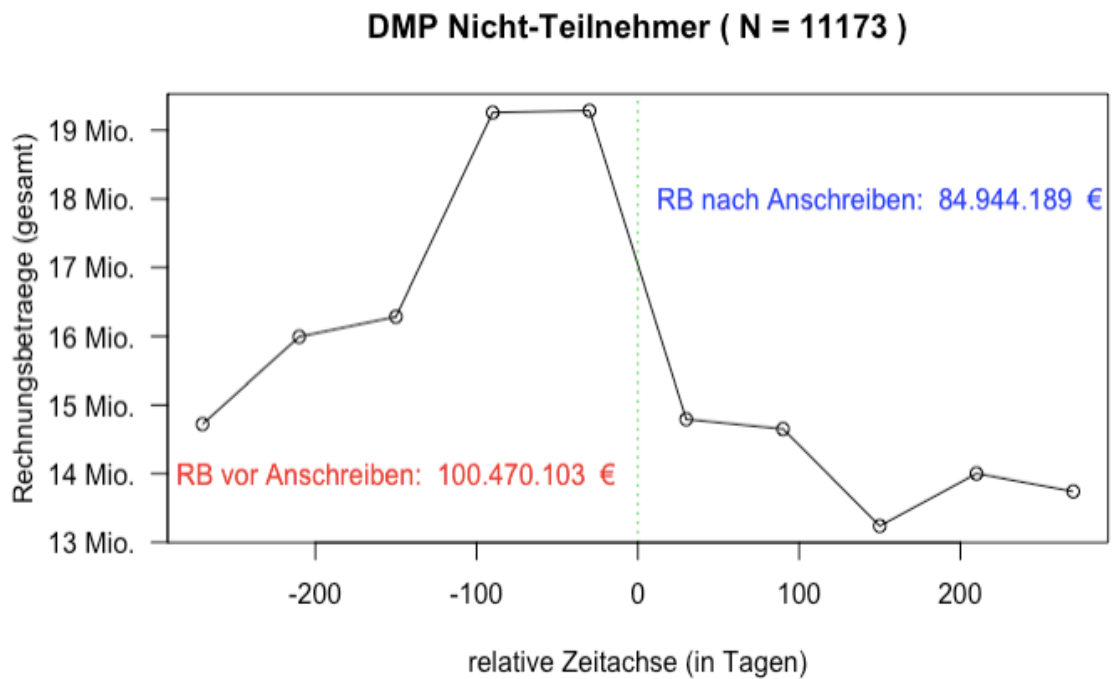


Abb. 34: Absolute Leistungsausgaben der DMP Nicht-Teilnehmer

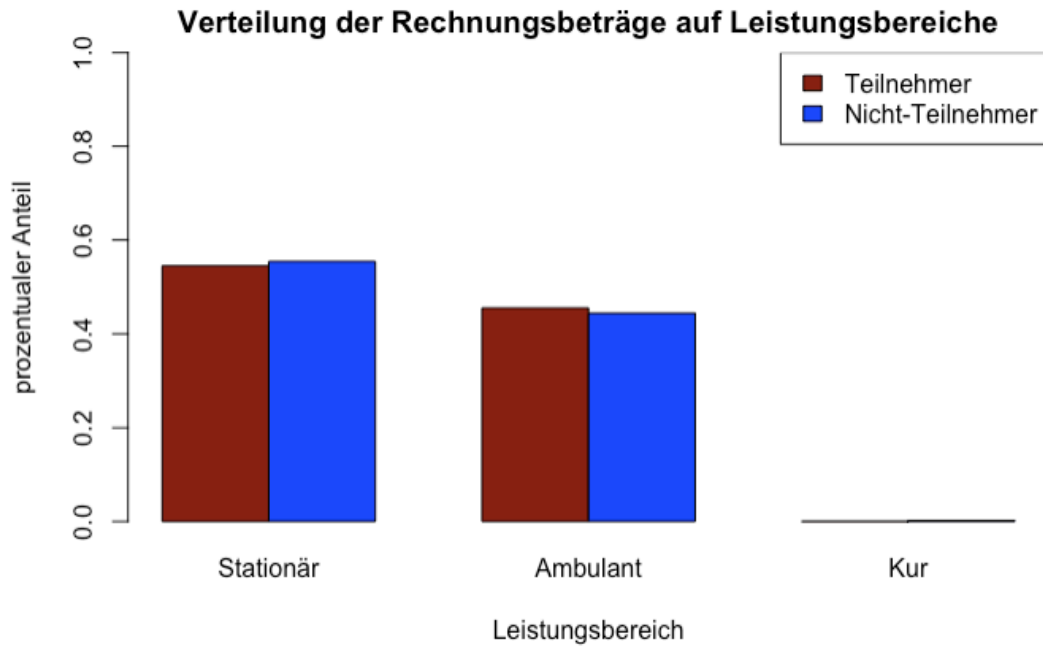


Abb. 35: Verteilung der Rechnungsbeträge für F-Diagnosen auf Leistungsbereiche

	Means Treated	Means Control	Mean-Diff
ALTER	55,87	55,93	-0,06
GESCHLECHTM	0,63	0,54	0,09
GESCHLECHTW	0,37	0,46	-0,09
V_LEIST_ANTIDEP_COUNT	2,27	1,87	0,40
COMORBIDITY	0,62	0,65	-0,04
LAST_FDIAG_AMB	-101,97	-136,69	34,72
LAST_FDIAG_STAT	-145,96	-116,10	-29,86
LEIST_AMB_360	5.428,55	5.087,39	341,16
LEIST_STAT_360	6.025,88	3.490,24	2.535,63
LEIST_FDIAG_AMB_360	723,07	680,43	42,65
LEIST_FDIAG_STAT_360	1.524,74	606,42	918,31
ANZ_FDIAG_AMB_360	3,78	3,15	0,63
ANZ_FDIAG_STAT_360	1,06	0,39	0,67

Tab. 15: Mittelwerte vor Matching – Basis-Modell

	Means Treated	Means Control	Mean-Diff
ALTER	55,87	55,87	100,00
GESCHLECHTM	0,63	0,63	100,00
GESCHLECHTW	0,37	0,37	84,17
V_LEIST_ANTIDEP_COUNT	2,27	2,33	-88,78
COMORBIDITY	0,62	0,54	98,18
LAST_FDIAG_AMB	-101,97	-102,60	42,20
LAST_FDIAG_STAT	-145,96	-128,70	65,35
LEIST_AMB_360	5.428,55	5.310,33	44,19
LEIST_STAT_360	6.025,88	4.610,77	55,41
LEIST_FDIAG_AMB_360	723,07	742,09	82,32
LEIST_FDIAG_STAT_360	1.524,74	1.362,38	88,09
ANZ_FDIAG_AMB_360	3,78	3,86	49,93
ANZ_FDIAG_STAT_360	1,06	0,72	100,00

Tab. 16: Mittelwerte nach Matching– Basis-Modell (1)

	Means Treated	Means Control	Mean-Diff
ALTER	55,87	55,45	0,41
GESCHLECHTM	0,63	0,63	0,00
GESCHLECHTW	0,37	0,37	0,00
V_LEIST_ANTIDEP_COUNT	2,27	2,33	-0,06
COMORBIDITY	0,62	0,47	0,14
LAST_FDIAG_AMB	-101,97	-100,99	-0,98
LAST_FDIAG_STAT	-145,96	-147,27	1,31
LEIST_AMB_360	5.428,55	5.247,64	180,91
LEIST_STAT_360	6.025,88	6.043,71	-17,83
LEIST_FDIAG_AMB_360	723,07	746,01	-22,94
LEIST_FDIAG_STAT_360	1.524,74	1.446,18	78,56
ANZ_FDIAG_AMB_360	3,78	3,84	-0,05
ANZ_FDIAG_STAT_360	1,06	0,95	0,11

Tab. 17: Mittelwerte nach Matching– Basis-Modell (2)

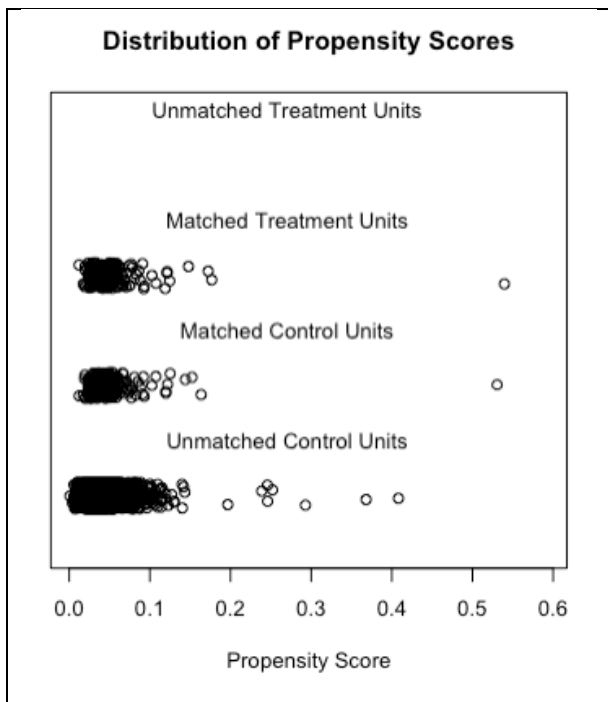
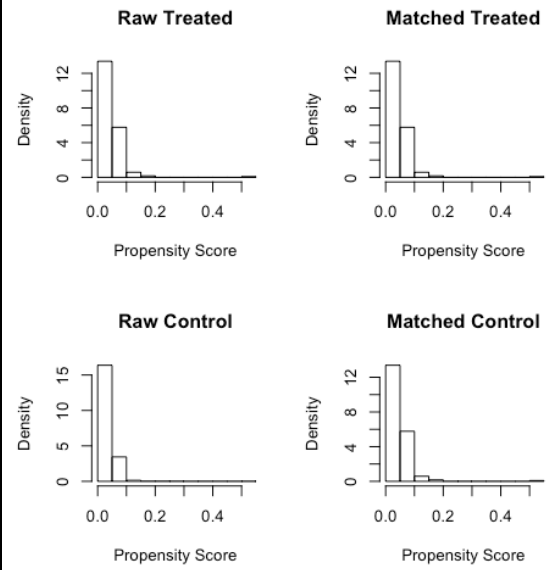
Jitter-Plot:**Histogramme:**

Abb. 36: Jitter-Plot und Histogramme des Basis-Modells

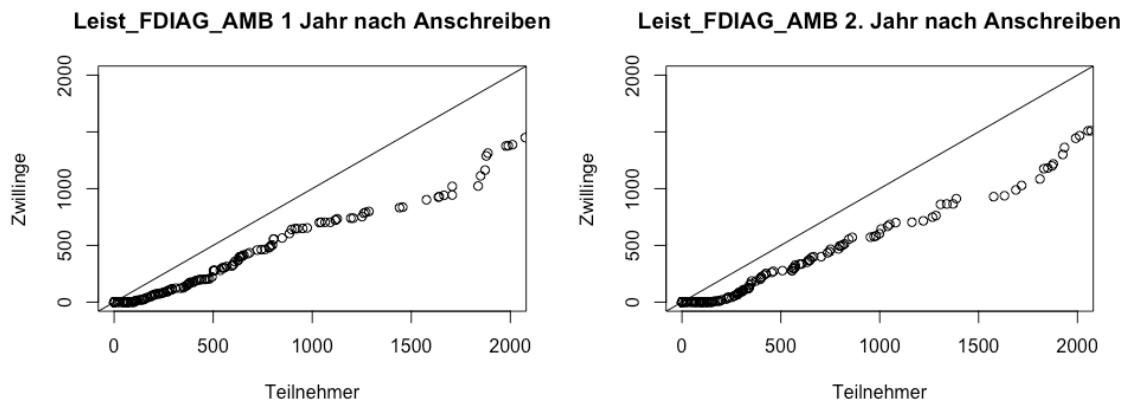


Abb. 37: QQ-Plots Programm-Qualität: Leistungen F-Diagnosen Ambulant - Alternative 1

Literaturverzeichnis

Bundesversicherungsamt (2018): Zulassung der strukturierten Behandlungsprogramme (Disease Management Programme - DMP) durch das Bundesversicherungsamt (BVA), <https://www.bundesversicherungsamt.de/weiteres/disease-management-programme/zulassung-disease-management-programme-dmp.html>, Abruf am 14.07.2018

Busch, M. A., Maske, U. E., Ryl, L., Schlack, R., & Hapke, U. (2013): Prävalenz von depressiver Symptomatik und diagnostizierter Depression bei Erwachsenen in Deutschland. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 56(5-6), 733-739

Deutsches Institut für Medizinische Dokumentation und Information (2018a): Aufbau der Systematik der ICD-10-WHO, <https://www.dimdi.de/static/de/klassi/icd-10-who/systematik/systematik.htm>, Abruf am 14.07.2018

Deutsches Institut für Medizinische Dokumentation und Information (2018b): ICD-10-WHO, <https://www.dimdi.de/static/de/klassi/icd-10-who/>, Abruf am 13.07.2018

Deutsches Institut für Medizinische Dokumentation und Information (2018c): Kapitel V - Psychische und Verhaltensstörungen (F00-F99), <https://www.dimdi.de/static/de/klassi/icd-10-who/kodesuche/onlinefassungen/htmlamtl2013/block-f30-f39.htm>, Abruf am 14.07.2018

Deutsche Krebsgesellschaft e.V. (2018): Gastrointestinale Tumoren, 1. Auflage, Elsevier, München

Gesamtverband der Deutschen Versicherungswirtschaft e. V. (2018): Konjunktur und Märkte 2018 - Versicherungswirtschaft stabil, konjunkturelle Risiken gestiegen, <https://www.gdv.de/resource/blob/32070/1a7a97ed4e8510a76a904eb95bfd50a4/pdf-konjunktur---maerkte-2018-data.pdf>

Grimmer, Arnd (2014): Statistik im Versicherungs- und Finanzwesen – Eine anwendungsorientierte Einführung, Springer Gabler, Wiesbaden

Ho, Daniel E., et al. (2011): MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, Journal of Statistical Software, 42.8, <https://gking.harvard.edu/files/gking/files/v42i08.pdf>, Abruf am 31.07.2018

Imai Kosuke, King Gary, Stuart Elizabeth A. (2008): Misunderstandings among Experimentalists and Observationalists about Causal Inference, Journal of the Royal Statistical Society A, 171, Part 2, S. 481–502, <https://imai.princeton.edu/research/files/matchse.pdf>, Abruf am 06.08.2018

Kuss, Oliver / Blettner, Maria / Börgermann, Jochen (2016): Propensity Score - eine alternative Methode zur Analyse von Therapieeffekten, Deutsches Ärzteblatt 113, Heft 35-36

Schumacher, Martin / Schulgen, Gabi (2006): Methodik klinischer Studien - Methodische Grundlagen der Planung, Durchführung und Auswertung, 2. überarbeitete Auflage, Springer-Verlag Berlin Heidelberg

Stangl, W. (2018): Stichwort: 'Komorbidität'. Online Lexikon für Psychologie und Pädagogik, <http://lexikon.stangl.eu/9654/komorbiditat/>, Abruf am 18.06.2018

Statistisches Bundesamt (Destatis) (2018): Krankheitskosten: Deutschland, Jahre, Krankheitsdiagnosen (ICD-10), Geschlecht, Altersgruppen, <https://www-genesis.destatis.de/genesis/online/logon?sequenz=tabelleErgebnis&selectionname=23631-0003&sachmerkmal=ICD10Y&sachschiessel=ICD10-F32-F34&transponieren=true>, Abruf am 07.07.2018

Stiftung Deutsche Depressionshilfe (2018a): Diagnose der Depression, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/was-ist-eine-depression/diagnose-der-depression>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018b): Häufigkeit, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/was-ist-eine-depression/haeufigkeit>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018c): Medikamentöse Behandlung, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/behandlung/medikamentoese-behandlung>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018d): Psychotherapeutische Behandlung, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/behandlung/psychotherapeutische-behandlung>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018e): Verlaufsformen, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/was-ist-eine-depression/verlaufsformen>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018f): Was ist eine Depression?, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/was-ist-eine-depression>, Abruf am 06.07.2018

Stiftung Deutsche Depressionshilfe (2018g): Wie wird eine Depression behandelt?, <https://www.deutsche-depressionshilfe.de/depression-infos-und-hilfe/behandlung>, Abruf am 06.07.2018

The R Foundation (2018): The R Project for Statistical Computing, <https://www.r-project.org>, Abruf am 21.07.2018

Wollschläger, Daniel (2012): Grundlagen der Datenanalyse mit R – Eine anwendungsorientierte Einführung, 2. Auflage, Springer-Verlag Berlin Heidelberg

World Health Organization (2017): Depression and other Common Mental Disorders – Global Health Estimates, <http://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf;jsessionid=2D7C16D399D501D2196D37DCA8003A05?sequence=1>

Ehrenwörtliche Erklärung

Ich versichere, die von mir vorgelegte Arbeit selbstständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.

Köln, den 31.08.2018

Ort, Datum

Unterschrift

