

# Matching mit den vorhandenen Anwendungsroutinen des SAS Enterprise Guide

Tobias Vogelmann  
 mhplus Krankenkasse  
 Franckstraße 8  
 71636 Ludwigsburg  
 tobias.vogelmann@mhplus.de

Tino Schubert  
 mhplus Krankenkasse  
 Franckstraße 8  
 71636 Ludwigsburg  
 tino.schubert@mhplus.de

## Zusammenfassung

In den letzten Jahren sind verschiedene Programme und Makros entstanden, mit denen unterschiedliche Arten des Matchings in SAS realisiert werden können. Prinzipiell können diese Programme auch in die grafische Benutzeroberfläche des SAS Enterprise Guide eingebunden werden. Allerdings ergeben sich hier, je nach Programm bzw. Makro, Restriktionen was die Auswahl und die Benennung von Variablen angeht. Die Anpassung an die eigenen Erfordernisse setzt zudem tiefere Kenntnisse in Base SAS voraus, die bei Endanwendern des SAS Enterprise Guide nicht immer vorhanden sind.

Dieser Beitrag stellt einen Weg vor, wie mit Hilfe der Anwendungsroutinen, die im SAS Enterprise Guide ab Installation vorhanden sind, verschiedene Matchingmethoden durchgeführt werden können. Das Matching wird dabei mehrstufig über die Anwendungsroutine `Abfrage erstellen` realisiert. Es wird dargestellt, wie mit Hilfe dieser Abfragen ein direktes 1:1-Matching mit einem Ziehen ohne Zurücklegen, als auch ein Propensity Score Matching (unter Zuhilfenahme der Anwendungsroutine `Logistische Regression`) umgesetzt werden kann und wie verschiedene Matchingparameter (bspw. Caliper) eingesetzt werden können. Zudem wird beschrieben, wie der Erfolg des Matchings überprüft werden kann. Programmierkenntnisse in Base SAS sind nicht notwendig, sämtliche Schritte sind in SAS Enterprise Guide vollständig grafisch umsetzbar.

**Schlüsselwörter:** Matching, SAS Enterprise Guide, Propensity Score, Anwendungsroutine, GKV, Erfolgsmessung

## 1 Einleitung

Matchingverfahren sind in vielen Wissenschafts- und Wirtschaftszweigen zu einem wichtigen Instrument, zur Messung von Interventionseffekten, geworden. Dabei sind Matchingverfahren als Teilgruppe der quasiexperimentellen Kontrollgruppenstudien besonders dann von Interesse, wenn die Bildung einer zufälligen Kontrollgruppe vor der Intervention nicht oder nur unter nicht zu vertretendem Aufwand möglich ist. Matchingverfahren werden dann genutzt, um die Interventions- und Kontrollgruppe anzugleichen und so systematische Verzerrungen zu verringern.

Da im Umfeld der Gesetzlichen Krankenversicherung (GKV) randomisierte Kontrollgruppenstudien oft aus rechtlichen oder ethischen Gründen nicht möglich sind, wurden hier in den vergangenen Jahren quasiexperimentelle Kontrollgruppenstudien mithilfe von Matchingansätzen vielfältig eingesetzt. Derartige Studien können bei gesund-

heitsökonomischen Fragestellungen mit Abrechnungsdaten verwendet werden, beispielsweise um die Kostenwirksamkeit verschiedener Medikamente miteinander vergleichen zu können. Durch ein Matching wird versucht, verschiedene Gruppen mit unterschiedlichen Medikamenten so weit anzugleichen, dass sie sich nur noch hinsichtlich der unterschiedlichen Medikation unterscheiden, in anderen Eigenschaften jedoch keine Unterschiede aufweisen [1].

Ziel dieser Arbeit ist es, einen Ansatz zu entwickeln, mit dem unterschiedliche Varianten des Matchings innerhalb des SAS Enterprise Guide (SEG) implementiert werden können. Dabei wird auf selbstgeschriebene Programme oder Makros verzichtet, und die Matching-Ansätze werden vollständig mit den vorhandenen Anwendungsroutinen des SAS Enterprise Guide abgebildet. Hierfür wird zunächst kurz auf die Bedeutung und Verbreitung von Matchingansätzen im Umfeld der Gesetzlichen Krankenversicherung eingegangen. Im Hauptteil der Arbeit wird dargestellt, wie verschiedene Matching-Ansätze im SAS Enterprise Guide, implementiert werden können.

Sämtliche Auswertungen wurden mit dem SAS Enterprise Guide 4.3, basierend auf einem System mit Base SAS 9.3 mit SAS/GRAPH und SAS/STAT, durchgeführt.

## **2 Matching mit GKV-Daten**

Im Umfeld der gesetzlichen Krankenkassen hat in den vergangenen Jahren, aufgrund der bestehenden Rahmenbedingungen, der Einsatz von quasiexperimentellen Kontrollgruppendesigns an Bedeutung gewonnen. Sie stellen eine Alternative zu den im GKV-Umfeld häufig verwendeten Prä-Post-Vergleichen dar, die, wenn sie ohne Vergleichsgruppe durchgeführt werden, anfällig für starke Verzerrungen sind. Daraus können falsche Managemententscheidungen resultieren. Einige Fragestellungen, die mithilfe von Matching mit Routinedaten der GKV bearbeitet wurden sind: Haben die Disease-Management-Programme zu einer verbesserten Versorgung der Versicherten mit Typ-2-Diabetes beigetragen [2]? Ist die Versorgung von Versicherten in Ärztenetzen kostengünstiger als die so genannte Regelversorgung [3]? Können Wahltarife mit einem Selbstbehalt für die Versicherten dazu beitragen, Kosten zu vermeiden [4]? Wie hoch sind die Arzneimittelausgaben und wie häufig sind stationäre Aufenthalte bei verschiedenen Therapien des Typ-2-Diabetes [5], [6]?

Alle Fragestellungen beinhalten dieselbe Problematik, denn es sollen zwei oder mehr Gruppen miteinander verglichen werden, die sich in einer Merkmalsausprägung unterscheiden (Intervention), sich aber in den anderen beobachtbaren Merkmalen möglichst ähnlich sind.

Krankenkassen haben ein großes Interesse daran, mittels Matchingverfahren eine Vergleichbarkeit zwischen Patienten, die beispielsweise an einem besonderen Vertrag teilnehmen und den Nichtteilnehmern herzustellen. Insbesondere auch dann, wenn man aus ethischen aber auch aus rechtlichen Gründen Versicherte nicht von der Inanspruch-

nahme ausschließen kann und so vor der Intervention keine Kontrollgruppe gebildet werden kann.

Limitationen des Matchings in der GKV sind ganz wesentlich darin zu sehen, dass beispielsweise Vitalparameter und andere Parameter zum Krankheitsverlauf und zur Lebensqualität in der Regel nicht in gewünschter Form in den Routinedaten abgebildet sind. Dies ist der Tatsache geschuldet, dass der eigentliche Zweck der GKV-Routinedaten-Erhebung primär in der Leistungsabrechnung und nicht in der Vergleichbarkeit von Krankheitsverläufen zu sehen ist. An einigen Stellen kann jedoch auch über die Schwere der Erkrankung eine Aussage getroffen werden, beispielsweise wenn eine Unterscheidung zwischen Krankheitsstufen getroffen wird (als Beispiel kann die NYHA-Klassifikation bei Herzinsuffizienz genannt werden, die von den Ärzten bei der Abrechnung im ICD-10 Code, grundsätzlich abgebildet werden kann). Dennoch ist auch hier anzumerken, dass die Codierqualität und -spezifität deutlich, beispielsweise zwischen Vertragsteilnehmern und -nichtteilnehmern differiert. Weiterhin ist die zeitliche Verfügbarkeit der GKV-Daten zu beachten: Den Krankenkassen stehen Abrechnungsdaten, je nach Leistungsbereich, erst Wochen oder Monate nach der Leistung zur Verfügung. Die Erfolgsmessung einer Intervention ist also nur mit einem größeren Zeitverzug möglich.

Folgende Limitationen sollten daher bei der Wahl des Verfahrens und bei der Ergebnisdarstellung mit GKV-Daten diskutiert werden:

- Anzahl und Auswahl der Parameter im Matching, insbesondere was die Berücksichtigung relevanter Erkrankungen in beiden Gruppen betrifft,
- Bedeutung der nicht berücksichtigten Parameter für das betrachtete Krankheitsbild,
- Darstellung der Kodierqualität, auch was Unterschiede in der Interventions- und der Kontrollgruppe betrifft,
- Parameter, an denen der Outcome gemessen wird.

Zusammenfassend lässt sich festhalten, dass die Vorüberlegungen und die Interpretation, und nicht das eigentliche Rechnen bzw. die konkrete Umsetzung in einem Statistikprogramm, die Herausforderungen beim Matching-Verfahren darstellen.

### **3 Möglichkeiten des Matching in SAS Enterprise Guide**

#### **3.1 Grundlagen**

Der SAS Enterprise Guide ist eine grafische Benutzeroberfläche (GUI) für die SAS-Umgebung. Im SEG stehen dem Anwender vorgefertigte Datenauswertungstools (so genannte Anwendungsroutinen) zur Verfügung, beispielsweise für deskriptive Statistiken, Regressionsanalysen oder die grafische Darstellung von Ergebnissen. Darüber hinaus kann der Benutzer Base SAS-Programme in die GUI einbinden und diese Programme als eigene Anwendungsroutinen speichern. Anwender können daher existie-

rende Matchingalgorithmen, die meistens auf den SAS-Prozeduren SQL und LOGISTIC basieren, (bspw. [7], [8], [9]) auch in SEG-Projekten verwenden.

Die Möglichkeiten sind für SEG-Anwender aus unserer Sicht aber nicht ideal, da

- für eine Veränderung und Individualisierung der bestehenden Programme Kenntnisse in Base SAS vorhanden sein müssen und diese nicht mit SEG-Kenntnissen durchgeführt werden können,
- das Look-and-Feel (Benutzeroberfläche) des SEG verlassen werden muss,
- die Einbindung von Programmcode, der vom Endanwender nicht nachvollzogen werden kann, immer auch mit Black-Box-Risiken einhergeht,
- je nach SAS-Code Restriktionen bezüglich der Auswahl, Anzahl und Namensgebung der verwendeten Variablen besteht.

Aus diesen Gründen ist eine Einbindung von Fremdcode nicht in jedem Fall die optimale Strategie, um ein Matching im SEG zu implementieren. Aus diesen Überlegungen heraus wurde nach Wegen gesucht, Matchingverfahren innerhalb der vorinstallierten Anwendungsroutinen des SAS Enterprise Guide zu bewerkstelligen, und gänzlich auf Base SAS-Programme zu verzichten. Bisher sind uns hierzu keine Arbeiten bekannt.

Im Folgenden werden für zwei verschiedene Matchingverfahren Möglichkeiten der Umsetzung im SEG aufgezeigt. Beiden Beispielen ist gemein:

- 1:1-Matching: Jeder Beobachtung der Interventionsgruppe wird maximal ein Matchingpartner der Kontrollgruppe zugewiesen.
- Greedy Matching: Eine Optimierung des Matching-Ergebnisses wird für jede Beobachtung einzeln vorgenommen, eine Gesamtbetrachtung und Optimierung des Matching-Ergebnisses über alle Beobachtungen hinweg erfolgt nicht.
- Ziehen ohne Zurücklegen: Jede Beobachtung der Kontrollgruppe kommt für maximal eine Beobachtung in der Interventionsgruppe als Matchingpartner infrage.

## **3.2 Ausgangslage**

Der verwendete Beispieldatensatz besteht aus zwei Tabellen. Die erste Tabelle beinhaltet pseudonymisierte demographische und Abrechnungsdaten zu 100 Versicherten, die hier als Interventionsgruppe bezeichnet werden. Die zweite Tabelle beinhaltet die gleichen Daten zu 5960 Personen, die hier als Kontrollgruppe bezeichnet werden. Als Attribute sind in beiden Tabellen vorhanden: Eindeutiges Pseudonym der Beobachtung, Geschlecht, Alter, Bundesland, Gesamtkosten 2010, Gesamtkosten 2011, unterschiedliche Antidepressivagruppen in 2010 (AD-Gruppen), Anzahl Verordnungen über Antidepressiva im Jahr 2010 (AD-Verordnungen), Gruppenzugehörigkeit (1=Interventionsgruppe, 2=Kontrollgruppe). Betrachtet wird eine Situation, in der die 100 Interventionsgruppenpatienten, die aus der gleichen Grundgesamtheit stammen wie die 5960 Kontrollgruppenpatienten, 2011 selektiv von Ärzten in einen Vertrag eingeschrieben wurden, ohne dass die genauen Auswahlkriterien der Krankenkasse bekannt sind. Untersucht werden

soll, aus Sicht und Datenlage der Krankenkasse, ob die Intervention (Einschreibung in einen Vertrag) im Jahr 2011 zu einer kostengünstigeren Behandlung geführt hat, also ob die Gesamtkosten 2011 der Interventionsgruppe unter den Gesamtkosten 2011 der Kontrollgruppe lagen. Anhand dieser Fragestellung werden die Methoden des direkten Matchings und des Propensity-Score-Matchings im SEG vorgestellt.

Tabelle 1 zeigt die durchschnittlichen Werte der Interventions- und der Kontrollgruppe vor dem Matching. Zu sehen sind statistisch signifikante Unterschiede in allen betrachteten Variablen vor dem Matching. Die Interventionsgruppe ist offensichtlich keine Zufallsstichprobe aus der Grundgesamtheit, die sich aus der Interventionsgruppe und der Kontrollgruppe ergibt. Ein einfacher Mittelwertvergleich der beiden Gruppen im Jahr 2011 ist also nicht geeignet, den Erfolg der Intervention zu messen.

**Tabelle 1:** Parameter vor dem Matching

Parameter	vor Matching		
	Interventionsgruppe	Kontrollgruppe	Differenz
N	100	5960	
Alter	57,1	43,9	13,1***
Kosten 2010	11930	2991	8939***
AD-Gruppen	0,8700	0,623	0,248***
AD-Verordnungen	2,20	1,48	0,74***
<b>Kosten 2011</b>	8472	3425	5046***

### 3.3 Direktes Matching

Beim direkten Matching (auch: simple matching) werden zwei Gruppen dann als vergleichbar betrachtet, wenn in allen für das Matching relevanten Variablen keine signifikanten Unterschiede zwischen den beiden Gruppen erkennbar sind. Im folgenden Beispiel werden die demographische Variable Jahresalter sowie die Variablen Gesamtkosten 2010, unterschiedliche verordnete Antidepressiva-Gruppen und die Anzahl der Antidepressiva-Verordnungen 2010 als Matching-Variablen verwendet.

Das Matching im SEG geschieht dabei in vier Schritten, die im Folgenden vorgestellt werden:

- 1 Zuordnung der Beobachtungen der Kontrollgruppe auf passende Beobachtungen der Interventionsgruppe unter Berücksichtigung der zulässigen Caliper.
- 2 Sicherstellung 1:1-Matching, indem jeder Beobachtung aus der Interventionsgruppe, für die ein Matchingpartner in der Kontrollgruppe verfügbar ist, genau ein Matchingpartner der Kontrollgruppe zugeordnet wird.
- 3 Sicherstellung eines Ziehens ohne Zurücklegen in der Kontrollgruppe, indem jedem Matchingpartner der Kontrollgruppe genau eine Beobachtung in der Interventionsgruppe zugeordnet wird.

- 4 Suche nach einem Matchingpartner für diejenigen Beobachtungen in der Interventionsgruppe, die durch den dritten Schritt jetzt keinem Matchingpartner mehr zugeordnet sind.

### 3.3.1 Erster Schritt: Direktes Matching der Interventions- mit der Kontrollgruppe

Um für jede Beobachtung in der Interventionsgruppe einen „Zwilling“ in der Kontrollgruppe zu finden, wird zunächst jede Beobachtung der Interventionsgruppe mit jeder Beobachtung der Kontrollgruppe verglichen. Dazu wird das kartesische Produkt beider Tabellen gebildet und diejenigen Beobachtungen der Interventions- und Kontrollgruppe ausgegeben, die in den gewählten Matching-Parametern übereinstimmen bzw. nicht mehr als vorgegeben voneinander abweichen. Im SEG werden die beiden Ausgangstabellen *Interventionsgruppe* und *Kontrollgruppe* hierzu mit der Anwendungsroutinevorlage `Abfrage erstellen` zusammengeführt. Die Tabelle Interventionsgruppe ist im Folgenden mit *t1*, die Tabelle Kontrollgruppe mit *t2* bezeichnet.

Die Zusammenführung geschieht über einen `cross join` über das eindeutige Pseudonym, dies kann im Fenster `Tabellen joinen` ausgewählt werden. Über den Reiter *Daten filtern* werden anschließend die Matchingparameter eingestellt: Für jede Variable, die für das Matching verwendet wird, wird ein Filter eingestellt. Ein exaktes Matching kann über den Operator `Gleich` erreicht werden. Die Ausgabetabelle enthält anschließend alle Beobachtungen der Interventionsgruppe, für die eine Beobachtung in der Kontrollgruppe gefunden wurde, die bezüglich der Matchingvariablen identisch sind. Für eine bessere Übersichtlichkeit wurden die Variablen in Abbildung 1 umbenannt: Variable, die aus der Tabelle *Interventionsgruppe* stammen, wurde der Suffix *I\_* vorangestellt, Variable aus der Tabelle *Kontrollgruppe* erhalten das Suffix *K\_*.

	I_Pseudo	I_Alter	I_Kosten_10	I_AD_Groupen	K_Pseudo	K_Alter	K_Kosten_10	K_AD_Groupen	kostendifferenz_10
8	3XT70cph85T5	53	8121.89	0	GNoiGZbWlcQZ	56	8448.33	0	326.44
9	3XT70cph85T5	53	8121.89	0	ITsYdduUuXTG	48	8112.02	0	9.87
10	3XT70cph85T5	53	8121.89	0	LM9tYMEWw9vJ	58	7836.03	0	285.86
11	3XT70cph85T5	53	8121.89	0	PPzrn0eSHQt2	54	8224.35	0	102.46
12	5muzvfzvBDw	59	5052.36	1	YrZ33j5l0j8	55	5258.83	1	206.47
13	72a7B1X2CZly	52	21426.32	2	Yq8D5UTHoCCg	54	21753.65	2	327.33
14	7RSfUBxxJ2Jh	50	16110.36	1	DsArLGGwIXa9	55	16223.58	1	113.22
15	7RSfUBxxJ2Jh	50	16110.36	1	Ujyye9ca56L0	48	16680.24	1	569.88
16	7RSfUBxxJ2Jh	50	16110.36	1	evHILGfxg7GZ	46	16505.17	1	394.81
17	9SjKZFIimgcT5	58	5013.2	1	8TOLIdSfrigo	55	5145.34	1	132.14

Abbildung 1: Tabelle nach cross join (Auszug)

Problematisch am exakten Matching ist, dass bei einer größeren Anzahl an Variablen und relativ wenigen Beobachtungen in der Kontrollgruppe für zahlreiche Beobachtungen keine Matching-Partner gefunden werden können. Als eine Lösung des Problems bietet sich der Einsatz so genannter Caliper (deutsch etwa Greifzirkel) an. Ein Caliper bestimmt, wie weit die Matching-Partner in den einzelnen Variablen voneinander abweichen dürfen. Die definierte Abweichung kann dabei absolut (beispielsweise: „Das Alter zweier Beobachtungen darf um maximal fünf Jahre nach oben oder unten abweichen“) oder relativ („Die Gesamtkosten zweier Beobachtungen dürfen um maximal fünf Prozent nach oben oder unten abweichen“) bestimmt werden. Dabei können Caliper für

die einzelnen Variablen unterschiedlich gewählt werden. Im folgenden Beispiel wird – um die angesprochenen Möglichkeiten alle einmal darzustellen – die Anzahl der Antidepressiva-Gruppen exakt ohne Abweichung gematcht (exaktes Matching), für das Jahresalter wird eine Abweichung von +/- 5 Jahren akzeptiert, für die Anzahl der Antidepressiva-Verordnungen eine Abweichung von +/- 1 (absoluter Caliper) und für die Gesamtkosten im Jahr 2010 wird eine Abweichung von 10% nach oben oder unten akzeptiert (relativer Caliper). Der Caliper sollte nicht zu groß gewählt werden, da ansonsten die gematchten Partner nicht mehr vergleichbar sind und die common support Annahme gefährdet wird. Sämtliche Caliper können ebenfalls über den Reiter `Daten filtern`, über den Operator `Zwischen` umgesetzt werden. Für die Variable `Alter` lautet der Filter beispielsweise: `t2.jahresalter BETWEEN t1.jahresalter-5 AND t1.jahresalter+5`. Der Filter für den relativen Caliper der Gesamtkosten 2010 lautet dementsprechend `t2.Kosten_10 BETWEEN 1.05*t1.Kosten10 AND 0.95*t1.Kosten_10`. Für den nächsten Schritt wird zudem der absolute Abstand der Gesamtkosten 2010 für die gematchten Beobachtungen errechnet. Dies geschieht im Reiter `Daten auswählen`, durch die Erstellung einer neuen Variablen mit dem Namen `kostendifferenz_10` und der SAS-Funktion `abs()`.

Wie in Abbildung 1 zu sehen ist, sind nun jeder Beobachtung in der Interventionsgruppe diejenigen Beobachtungen in der Kontrollgruppe gegenüber gestellt, die den Matching-Kriterien entsprechen. Es ist ebenfalls zu erkennen, dass einer Beobachtung in der Interventionsgruppe in diesem Schritt mehrere Beobachtungen in der Kontrollgruppe zugewiesen werden können – dies wird zu einem späteren Zeitpunkt korrigiert.

### 3.3.2 Zweiter Schritt: 1:1-Zuordnung der Kontroll- zur Interventionsgruppe

Beim ersten Schritt, insbesondere beim Einsatz von Calipern, kann es vorkommen, dass für eine Beobachtung in der Interventionsgruppe mehrere Matching-Partner in der Kontrollgruppe gefunden werden. Da das erklärte Ziel ein 1:1-Matching darstellt, muss aus den mehrfachen Matching-Partnern genau ein Partner für jede Beobachtung in der Interventionsgruppe ausgewählt werden. Für die Zuordnung genau eines Matchingpartners ist es erforderlich, das Kriterium der Auswahl dieses einen Partners näher zu spezifizieren. Denkbar ist eine Zufallsauswahl unter allen infrage kommenden Partnern (über eine zuvor generierte Zufallszahl oder durch Auswahl der ersten Beobachtung), die Minimierung des Abstandes eines ausgewählten Parameters oder die Minimierung eines Globalmaßes, das die Abweichungen über mehrere oder alle Parameter misst. Die folgende Darstellung geht von der Minimierung der Unterschiede in einem einzelnen, normativ festgelegten Parameter aus. Als zu minimierender Parameter wird dabei in unserem Beispiel die Differenz der Gesamtkosten 2010 zwischen den möglichen Matchingpartnern, wie im vorherigen Schritt berechnet, verwendet.

In einem neuen Knoten wird mit der Anwendungsroutine `Abfrage erstellen`, das Ergebnis des vorherigen Schrittes modifiziert. Als auszugebende Variable werden die



Pseudonyme der Interventionsgruppe (Suffix  $I\_$  beachten!) und die Kostendifferenz 2010 ausgewählt, wobei letztere über Auswahl der Funktion `MIN` in der Spalte *Zusammenfassung* über die Pseudonyme der Interventionsgruppe minimiert wird. Der Output zeigt nun die Pseudonyme der einzelnen Beobachtungen der Interventionsgruppe und den Abstand der Gesamtkosten zum am nächsten liegenden Matchingpartner der Kontrollgruppe<sup>1</sup>. Über eine weitere Anwendungsroutine `Abfrage erstellen` kann nun die Wiederzuordnung der Kontrollgruppenbeobachtungen erfolgen. Das Ergebnis der eben erstellten Abfrage und das Ergebnis aus Schritt 1 werden per `inner join` über die Pseudonyme der Interventionsgruppe zusammengeführt, mit einem Filter, der die Gleichheit der minimalen Kostendifferenz mit der berechneten Kostendifferenz des ersten Schrittes sicherstellt

(Filter: `t1.MIN_of_kostendifferenz10 = t2.kostendifferenz_10`). Im Ergebnis ist nun jeder Beobachtung in der Interventionsgruppe entweder keine Beobachtung (kein Matchingpartner gefunden) oder genau eine Beobachtung in der Kontrollgruppe zugeordnet. Die Situation nach dem zweiten Schritt ist in **Abbildung 2 Fehler! Verweisquelle konnte nicht gefunden werden.** nochmals dargestellt.

	I_Pseudo	I_alter	I_Kosten_10	I_AD_Groupen	K_Pseudo	K_alter	K_Kosten_10	K_AD_Groupen	kostendifferenz_10
1	0kFMjJtyVocL	63	3247.55	0	Ww3lhJKzsXaz	59	3221.95	0	25.6
2	0wKtspysGzSt	52	4715.63	1	99NSfqAYfttk	57	4719.49	1	3.86
3	3XT70cph85T5	53	8121.89	0	ITsYdduUuXTG	48	8112.02	0	9.87
4	5muzvfzvBDw	59	5052.36	1	YlrZ33j5l0j8	55	5258.83	1	206.47
5	72a7B1X2CZly	52	21426.32	2	Yq8D5UTHoCCg	54	21753.65	2	327.33
6	7RSfUExxJ2Jh	50	16110.36	1	DsArLGgWfXa9	55	16223.58	1	113.22
7	9SjKZFImgcT5	58	5013.2	1	tOBffFOK0Rw	56	4914.16	1	99.04
8	9pg0ZVGcuffNN	60	18238.13	1	YvJXQ1MV25op	57	17656.09	1	582.04
9	DKCnRDihHtku	61	13252.63	1	g50EnG5h3hUr	60	12830.19	1	422.44

**Abbildung 2:** Tabelle nach dem zweiten Schritt: Jede Beobachtung der Interventionsgruppe ist jetzt maximal einer Beobachtung in der Kontrollgruppe zugeordnet

Die 1:1-Zuordnung kann wie angedeutet auch auf andere Weisen sichergestellt werden, beispielsweise kann statt der Abstandsminimierung über eine Variable auch ein Globalmaß des Abstandes zwischen den Beobachtungen der Interventions- und Kontrollgruppe berechnet und dieses Globalmaß anschließend minimiert werden. Die Berechnung eines derartigen Globalmaßes kann – wie die Berechnung des absoluten Abstandes einer Variablen – in der Anwendungsroutine `Abfrage erstellen` in einer neuen Variablen geschehen. Als berechnetes Globalmaß wird in der Literatur beispielsweise die Mahalanobis-Distanz empfohlen ([10], S. 162). Die weiteren Ausführungen dieser Arbeit sind grundsätzlich auch auf derartige Globalmaß-Optimierungen anwendbar.

<sup>1</sup> Die Anzahl der Zeilen in der ausgegebenen Tabelle ist gleich der Anzahl der Beobachtungen in der Interventionsgruppe, für die mind. ein Matchingpartner gefunden wurde. Im verwendeten Beispiel sind dies 76 von ursprünglich 100, d. h. für 24 Beobachtungen wurde kein Partner in der Kontrollgruppe gefunden, eine Erfolgsmessung für diese Beobachtungen scheidet aus.



### 3.3.3 Dritter Schritt: Ziehen ohne Zurücklegen in der Kontrollgruppe

Im vorherigen Schritt wurde sichergestellt, dass eine 1:1-Zuordnung der Beobachtungen in der Interventionsgruppe zu den Beobachtungen der Kontrollgruppe erfolgt. Bis zum jetzigen Punkt ist aber umgekehrt nicht ausgeschlossen, dass eine Beobachtung in der Kontrollgruppe mehreren Beobachtungen in der Interventionsgruppe zugeordnet wird. Um eine derartige Zuordnung, die im Sinne der Kombinatorik einem Ziehen ohne Zurücklegen (ZoZ) entspricht, zu erreichen, ist ein dritter Schritt im Matching erforderlich.

Das Vorgehen entspricht dabei im Wesentlichen dem Vorgehen im zweiten Schritt. Über einen neuen Knoten `Abfrage erstellen`, werden jetzt als auszugebende Variablen die Pseudonyme der Kontrollgruppe (Suffix `K_` beachten!) und die Kostendifferenz 2010 ausgewählt, wobei letztere über Auswahl der Funktion `MIN` in der Spalte *Zusammenfassung* über die Pseudonyme der Kontrollgruppe minimiert wird. Anschließend werden wieder die Beobachtungen der Interventionsgruppe zugeordnet, über einen `inner join` über das Kontrollgruppenpseudonym und einen Filter, der den richtigen Abstand der Gesamtkosten kontrolliert (s. vorheriger Schritt).

Das Ergebnis des dritten Schritts ist eine Tabelle, die jeder Beobachtung der Interventionsgruppe genau eine Beobachtung der Kontrollgruppe zuordnet die hinsichtlich der gewählten Matching-Parameter vergleichbar ist, und jede Beobachtung der Kontrollgruppe dient maximal einer Beobachtung der Interventionsgruppe als Matchingpartner.

### 3.3.4 Vierter Schritt: „Second best“ für die Interventionsgruppe

Werden im dritten Schritt doppelte Zuordnungen der Kontrollgruppe zur Interventionsgruppe entfernt, dann sind Beobachtungen in der Interventionsgruppe, die im ersten Schritt noch einen Matching-Partner erhalten haben, jetzt ohne Matching-Partner. Für diese Beobachtungen werden nun im vierten Schritt die zweitbesten Partner aus der Kontrollgruppe zugeordnet.

Hierfür werden zunächst über einen neuen Knoten `Abfrage erstellen` die eindeutigen Pseudonyme der Interventionsgruppe aus dem ersten Schritt mit denen des dritten Schritts verglichen, um diejenigen Beobachtungen zu finden, die nach dem ersten Schritt noch einen Matchingpartner hatten, nach dem dritten aber nicht mehr. Dies geschieht durch einen `left join` der beiden Tabellen über die Pseudonyme der Interventionsgruppe und einen Filter, der ein Fehlen der Pseudonyme in der Ausgabe des dritten Schritts voraussetzt (`t2.Pseudo IS MISSING`). Anschließend erfolgt die Auswahl des zweitbesten Partners analog dem zweiten Schritt: Es wird diejenige Beobachtung in der Kontrollgruppe ausgewählt, die den minimalen Abstand in den Gesamtkosten 2010 hat und noch nicht im dritten Schritt einem Matching-Partner zugewiesen wurde. Die Ausgabe dieses Schritts ist in Abbildung 3 dargestellt. Anschließend erfolgt, wie im zweiten Schritt, eine Zuordnung zu den Beobachtungen der Kontrollgruppe über die minimierte Kostendifferenz.

	I_Pseudo	MIN_of_kostendifferenz_10
1	KX32ImpTsPyP	48.45
2	QzS26MsPmMJp	58.41
3	IHeafIkYO3Zf	191.97
4	nSxUzYmXRcR6	413.08
5	qnfq62HENTTd	288.35
6	rNmttU7sGFY3	88.49

**Abbildung 3:** Kostenminimierung Beobachtungen, die nach dem dritten Schritt jetzt keinen Matchingpartner mehr haben

Je nach Datensatz kann es vorkommen, dass einzelne Beobachtungen keinen zweitbesten Matchingpartner besitzen und somit ohne Matchingpartner bleiben.

Über einen neuen Knoten `Tabelle` anhängen können nun die Ergebnisse des vierten Schritts mit denen des dritten Schritts zusammengeführt werden. Ergebnis ist eine Tabelle, die dem erklärten Ziel, ein direktes 1:1-Matching mit einem Ziehen ohne Zurücklegen in der Kontrollgruppe, entspricht.

### 3.4 Propensity Score Matching

Eine Alternative zum Matching auf einzelnen Variablen stellt ein Matching über die Propensity Scores dar. Der Propensity Score ist die Wahrscheinlichkeit, mit der eine bestimmte Beobachtung zur Interventionsgruppe gehört. Die Theorie hinter dem Propensity Score besagt, dass beim Vergleich zweier Gruppen nur die Variablen kontrolliert werden müssen, die die Wahrscheinlichkeit der Teilnahme beeinflussen. Es reicht darüber hinaus aus, nur die Teilnahmewahrscheinlichkeit selbst zu kontrollieren, um Verzerrungen auszuschließen. Anders als beim direkten Matching müssen die Beobachtungen also nicht mehr in einer Vielzahl von Variablen übereinstimmen ([12], S. 81). Das Propensity-Score-Matching macht sich diese Theorie zu Nutze und verwendet als Matchingparameter lediglich die Propensity Scores, die üblicherweise vor dem Matching mit einer logistischen Regression geschätzt werden.

#### 3.4.1 Erster Schritt: Schätzung der Propensity Scores

Die Propensity Scores (PS) werden im Rahmen dieser Arbeit mithilfe einer logistischen Regression geschätzt. Hierfür wird die SEG-Anwendungsroutine `Logistische Regression` verwendet. Abhängige Variable ist die Variable der Gruppenzugehörigkeit (zur Erinnerung: 1=Interventionsgruppe, 2=Kontrollgruppe). Da die PS die Wahrscheinlichkeit der Zugehörigkeit zur Interventionsgruppe angeben, muss das Modell an das Niveau der Interventionsgruppe (hier: 1) angepasst werden. Dies kann im Reiter *Modell – Ausprägung* eingestellt werden. Im folgenden Beispiel werden die Propensity Scores mithilfe der gleichen unabhängigen Variablen geschätzt, die auch in das direkte Matching eingegangen sind: Jahresalter, Gesamtkosten 2010, Anzahl unterschiedlicher Antidepressivagruppen und Anzahl der Antidepressiva-Verordnungen 2010.

### 3.4.2 Zweiter Schritt: Matching der Propensity Scores mit Caliper

Nachdem die Propensity Scores geschätzt sind, erfolgt im zweiten Schritt ein Matching der geschätzten Scores. Auch bei diesem Matching kommt ein Caliper zum Einsatz und auch hier gilt grundsätzlich: Je kleiner der Caliper, desto genauer ist das Matching, je größer der Caliper, desto mehr Beobachtungen können gematcht werden. Die richtige Wahl des Calipers hängt also entscheidend von der Größe der Interventions- und der Kontrollgruppe ab. Außerdem gilt: Je mehr Variablen für die Schätzung der PS verwendet wurden, desto mehr einzigartige geschätzte PS werden erzeugt und desto eher muss ein angebrachter Caliper verwendet werden, um einen „statistischen Zwilling“ zu finden. Umgekehrt gilt: Je weniger Variablen in die Schätzung der PS eingehen, desto weniger eindeutige Werte liefert die logistische Regression und umso ungleicher sind sich die gematchten Paare. Auch an dieser Stelle existiert also wieder ein trade-off zwischen der Vergleichbarkeit der gematchten Paare und der Anzahl der gefundenen „Zwillinge“. Im Beispiel wird ein Caliper von +/- 0.0001 verwendet.

### 3.4.3 Weitere Schritte

Die weiteren Schritte entsprechen im Wesentlichen den Schritten 2 bis 4 beim direkten Matching: Sicherstellung einer 1:1-Zuordnung der Interventions- zur Kontrollgruppe, Ziehen ohne Zurücklegen für die Kontrollgruppe und „second best“ Lösung für die verbleibenden Beobachtungen der Interventionsgruppe. Statt über einen normativ ausgesuchten Parameter (im Beispiel die Gesamtkosten 2010) erfolgt die Optimierung nun über den Abstand der geschätzten Propensity Scores zwischen den Beobachtungen. Je kleiner der Abstand der geschätzten Propensity Scores, desto ähnlicher sind sich die Beobachtungen.

## 4 Erfolgskontrolle des Matchings

Zur Kontrolle, ob das Matching ein Erfolg in dem Sinne darstellt, dass es die beobachtbaren Unterschiede zwischen den Gruppen nivelliert hat, können vor und nach dem Matching jeweils t-Tests für unabhängige Stichproben über die Anwendungsroutine  $t$ -Test durchgeführt werden. Das Matching kann dann als erfolgreich angesehen werden, wenn vorher signifikante Unterschiede in den betrachteten Variablen reduziert wurden. Dies gilt sowohl für die Variablen, die in das Matching einbezogen wurden, als auch auf eventuell nicht berücksichtigte Variablen, die einen Einfluss auf den Outcome (hier: Kosten 2011) haben.

Unser Beispiel ist in Tabelle 2 zu sehen: Vor dem Matching sind in jeder Variable signifikante Unterschiede zwischen der Interventionsgruppe (IG) und der Kontrollgruppe (KG) zu erkennen. Durch das relativ einfache direkte Matching werden die signifikanten Unterschiede weitgehend beseitigt. Beim Propensity-Score-Matching wird deutlich, dass die verwendeten Parameter nicht in der Lage waren, sämtliche signifikanten Unterschiede zu beseitigen. Dies kann damit begründet werden, dass wir in unserem einfachen Beispiel zur Schätzung der Propensity Scores nur einige wenige Variab-

len verwendet haben (Alter, Verordnungen, Gesamtkosten 2010). Ein Matching auf die geschätzten Propensity Scores liefert aber nur dann erwartungstreue Ergebnisse, wenn alle Variablen, die gleichzeitig die Teilnahmewahrscheinlichkeit und den Outcome beeinflussen, in die Schätzung der PS eingehen [11]. Sehr wahrscheinlich existieren weitere Variablen, die die Teilnahmewahrscheinlichkeit und das Ergebnis beeinflussen. Ob diese beobachtbar und im Datensatz vorhanden sind, ob also mit anderen Worten die Voraussetzungen für das PS-Matching erfüllt sind, ist für jede Forschungsfrage separat zu prüfen. In unseren Beispielen ist nicht ausgeschlossen, dass beispielsweise die Ansprache durch den Arzt maßgeblich entscheidend ist.

Der Outcome, der hier vorliegenden Intervention, sind die Kosten 2011. Diese sind vor dem Matching in der Interventionsgruppe signifikant höher als in der Kontrollgruppe. Nach dem Matching sind die signifikanten Unterschiede beim direkten Matching verschwunden: Die beiden Gruppen unterscheiden sich hinsichtlich der Gesamtkosten im Jahr 2011 (nach der Intervention) nicht mehr signifikant. Ein Vergleich zwischen IG und KG kann durchgeführt werden. In dem hier verwendeten Beispiel führte die Betreuung im Versorgungsprogramm zu keinem finanziellen Vorteil. Hieraus wären entsprechende Managementmaßnahmen abzuleiten. Beim Propensity-Score-Matching verbleibt auch im Jahr 2010 ein signifikanter Unterschied zum 5%-Niveau, weshalb ein Vergleich zwischen IG und KG in der Form nicht zielführend ist.

**Tabelle 2:** Parameter vor und nach dem Matching

Parameter	vor Matching			direktes Matching			PS-Matching		
	IG	KG	Diff.	IG	KG	Diff.	IG	KG	Diff.
N	100	5960		76	76		78	78	
Alter	57,1	43,9	13,1***	55,7	55,1	0,6	55,4	56,8	1,4
Kosten 2010	11930	2991	8939***	9401	9393	8	9034	5973	3060**
AD-Gruppen	0,8700	0,623	0,248***	0,79	0,79	0	0,78	0,99	0,21*
AD-Verordnungen	2,20	1,48	0,74***	1,74	1,73	0,1	2,03	2,36	0,33
<b>Kosten 2011</b>	8472	3425	5046***	6073	5313	760	6115	3639	2476**

## 5 Fazit und Ausblick

Wir haben dargestellt, wie sich im SAS Enterprise Guide mit den vorhandenen Anwendungsroutinen ein Matching anhand von GKV-Routinedaten durchführen lässt. Dafür wurde auf die Anwendungsroutinen `Abfrage erstellen` und `Logistische Regression` zurückgegriffen. Sämtliche dargestellte Matching-Verfahren basieren dabei auf einem Greedy-Matching, das heißt, das Matchingergebnis wird für jede Beobachtung optimiert, nicht jedoch über alle Beobachtungen hinweg. Die Implementierung einer derartigen Optimierungsstrategie über alle Beobachtungen hinweg, steht noch aus. Ebenso ist der verwendete Datensatz mit insgesamt etwa 6000 Beobachtungen ver-

gleichsweise klein. Für größere Datenmengen ist das beschriebene Verfahren über einen `cross join` vermutlich mit langen Rechenzeiten verbunden.

Die Frage, ob ein direktes Matching oder ein Propensity Score Matching die bessere Wahl ist, muss anhand der Forschungsfrage und den zur Verfügung stehenden Daten individuell entschieden werden. Ein Matching auf die geschätzten Propensity Scores hat im Vergleich zu einem direkten Matching beispielsweise zur Folge, dass die asymptotischen Standardfehler höher sind als beim direkten Matching auf sämtliche relevanten Variablen, was schwächere statistische Tests und breitere Konfidenzintervalle zur Folge hat ([12], S. 84). Ein in der Praxis nicht zu unterschätzender Vorteil der direkten Matching-Methode ist, dass sie einfach und bildhaft zu erklären ist und auch ohne tieferegehende statistische Kenntnisse gegenüber Entscheidungsträgern im Unternehmen kommuniziert werden kann.

Auf der anderen Seite bietet das Propensity Score Matching den Vorteil, dass durch das Matching allein auf die geschätzten Propensity Scores die Beobachtungen nicht in einer ganzen Reihe von Parametern identisch sein müssen, was die Wahrscheinlichkeit erhöht, für möglichst viele Beobachtungen einen Matchingpartner zu finden. Auch führt die Schätzung der Propensity Scores zu einer Gewichtung der Parameter im Modell anhand der vorliegenden Daten, während die Gewichtung beim direkten Matching einen normativen Eingriff des Anwenders erfordert. Diese Vor- und Nachteile der verschiedenen Methoden sind vor der Durchführung einer Methodenwahl abzuwägen.

In der Praxis sind insbesondere die Variablenauswahl, die Datenverfügbarkeit und –qualität sowie die richtige Interpretation der Ergebnisse die Hauptfaktoren, die für den Erfolg und die Anwendbarkeit des Matchings entscheidend sind. Dies gilt besonders im Hinblick auf die Anfangs diskutierten Eigenarten der GKV-Routinedaten.

Die Autoren freuen sich über Austausch, Anregungen, Kritik und Vorschläge zum methodischen Vorgehen.

## Literatur

- [1] T. Reinhold, F. Andersohn, F. Hessel, B. Brüggjenjürgen, S. N. Willich: Die Nutzung von Routinedaten der gesetzlichen Krankenkassen (GKV) zur Beantwortung gesundheitsökonomischer Fragestellungen - eine Potenzialanalyse, in: Gesundheitsökonomie und Qualitätsmanagement 2011, 16 (3): 153-159.
- [2] R. Linder, S. Ahrens, D. Köppel, T. Heilmann, F. Verheyen: Nutzen und Effizienz des Disease-Management-Programms Diabetes mellitus Typ 2, in: Deutsches Ärzteblatt 2011, 108 (10): 155-162.
- [3] G. E. Braun, M. Gröbner, R. Seitz: Evaluation vernetzter Versorgungsstrukturen: Ergebnisse einer empirischen Untersuchung, in: Gesundheitsökonomie und Qualitätsmanagement 2008, 13 (6): 358-364.

- [4] N. Hemken, C. Schusterschitz, M. Thöni: Optionale Selbstbehalte—ein Instrument zur Nachfragesteuerung in der Gesetzlichen Krankenversicherung?, in: *Gesundheitsökonomie und Qualitätsmanagement* 2011, 16 (3): 171-177.
- [5] E.-G. Hagenmeyer, H. Gothe, W. Landgraf, M. Kulik, G. Schiffhorst, B. Häussler: Ressourcen-Inanspruchnahme und Kosten der Behandlung von Typ-2-Diabetikern unter Insulin-glargin- oder Insulindetemir-Therapie (LIVE-KK): Analyse von Krankenkassen-Routinedaten mittels Propensity Score Matching, in: *Gesundheitsökonomie und Qualitätsmanagement* 2010, 15 (3): 121-126.
- [6] S. Gölz, F.-W. Dippel, T. Schneider, B. Brüggenjürgen: Behandlungskosten bei Diabetes mellitus: Vergleich von drei basalinsulinbasierten Therapieansätzen, in: *Monitor Versorgungsforschung* 2011, 4 (5): 46-51.
- [7] L. S. Parsons: Using SAS Software to Perform a Case-Control Match on Propensity Score in an Observational Study, in: *SAS SUGI 2005*, 30, Paper 225-25.
- [8] W. W. Feng, Y. Jun, R. Xu: A Method/Macro Based on Propensity Score And Mahalanobis Distance to Reduce Bias in Treatment Comparison in Observational Study, in: *PharmaSUG 2006*, Paper PR05.
- [9] A. Deckert: 1:N Matching von Fällen und Kontrollen: Propensity Score vs. PROC SQL, in: C. Ortseifen, H. Ramroth, M. Weires, R. Minkenber (Hrsg.): *KSFE 2011 - Voneinander lernen. Proceedings der 15. Konferenz der SAS®-Anwender in Forschung und Entwicklung (KSFE)*, Shaker Verlag, Aachen: 105-119, 2011.
- [10] E. A. Stuart, D. B. Rubin: Best Practices in Quasi-Experimental Designs: Matching Methods for Causal Inference., in: J. W. Osborne (Hrsg.): *Best Practices in Quantitative Social Science*, Sage Publications, Thousand Oaks: 155-176, 2007.
- [11] M. Caliendo, S. Kopeinig: Some Practical Guidance for the Implementation of Propensity Score Matching, in: *Journal of Economic Surveys* 2008, 22 (1): 31-72.
- [12] J. D. Angrist, J.-S. Pischke: *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, 2009.