

Propensity Score Matching zur Evaluation des DMP "Diabetes mellitus Typ 2"

Dr. Susanne Ahrens, Philipp Wewering, Prof. Dr. Roland Linder
WINEG | Wissenschaftliches Institut der TK, Hamburg

Symposium **Routinedaten im Gesundheitswesen 2012**
München | 7. März 2012

Ein Ziel des WINEG:

Wirksamkeit präventiver, diagnostischer oder therapeutischer Maßnahmen **bewerten** unter Berücksichtigung von **Kosten-Nutzen-Aspekten**

Beispiel:

Evaluation des DMP Diabetes mellitus Typ 2 basierend auf Routinedaten der Techniker Krankenkasse

Fragen:

- Wie kann man die Auswirkung der Teilnahme am DMP (Behandlungseffekt, Therapieerfolg) beurteilen?
- Können Routinedaten helfen, Programmeffekte hinsichtlich medizinischem Outcome und Kosten zu identifizieren?

Inhalt

- **Motivation**
- **Methodik (Propensity Score Matching)**
- **Ergebnisse**
- **Diskussion**
- **Fazit & Ausblick**

Motivation

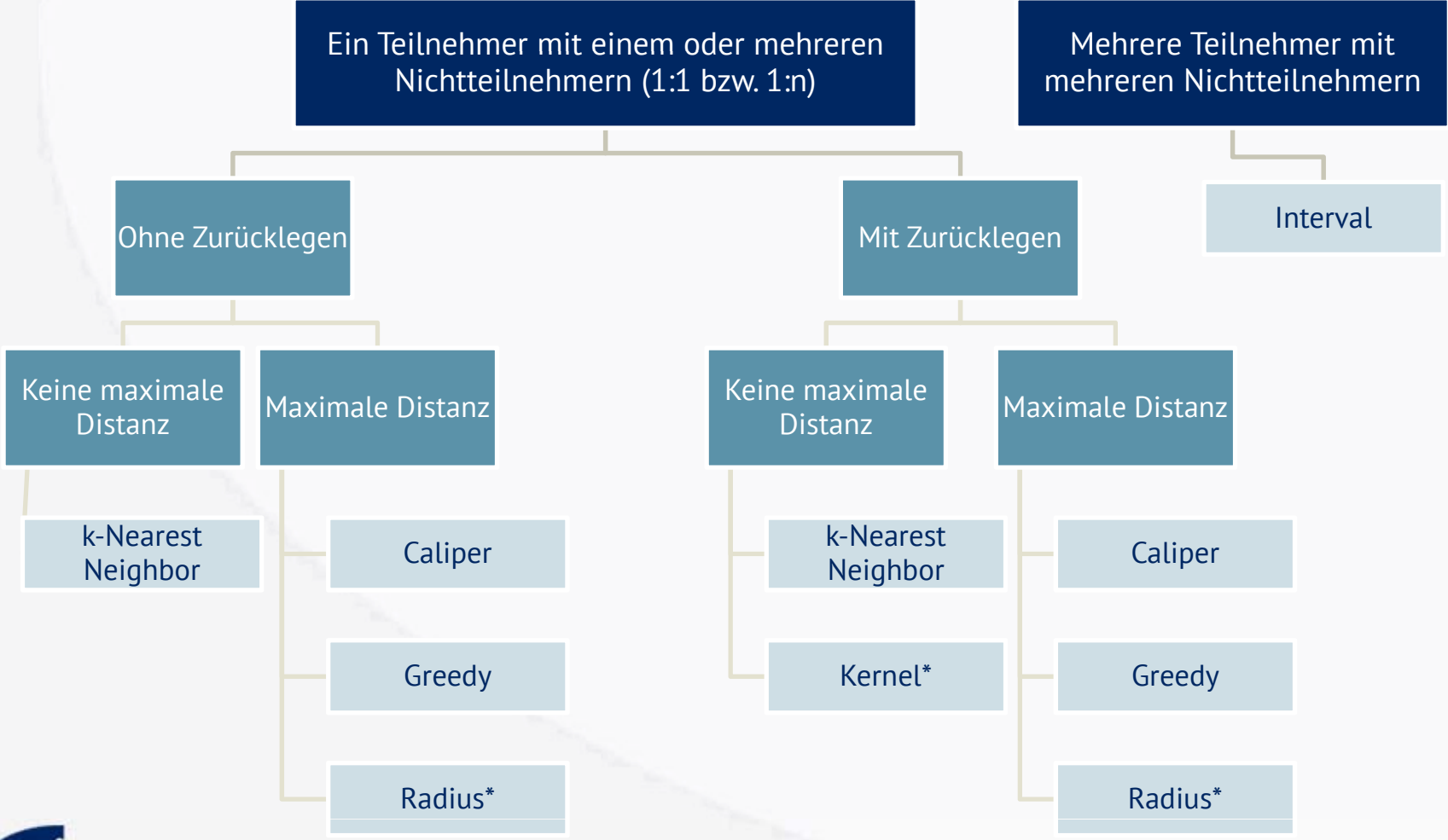
- **Aufwendung von 1,1 Milliarden Euro** für DM-Programme in 2009
- **konzeptionelle Inhalte der DMP's müssen gesetzlich vorgeschriebene Anforderungen erfüllen***:
 - **Bewertung der Wirksamkeit und der Kosten (Evaluation)**
 - **Möglichkeiten des Vergleiches zu einer Kontrollgruppe (keine DMP-Versicherte oder keine DMP-Ärzte) sind zu prüfen**

*Verordnung über das Verfahren zum Risikostrukturausgleich in der gesetzlichen Krankenversicherung (Risikostruktur-Ausgleichsverordnung - RSAV Anlage 1 zu § 28b bis § 28g): Anforderungen an strukturierte Behandlungsprogramme für Diabetes mellitus Typ 2

Kontrollgruppenbildung via Matching

- **Berücksichtigung von Selektionseffekten:**
 - Matching auf Basis des Propensity Score kann Selektionsbias reduzieren
- **Übereinstimmung:**
 - Matching-Methoden gleichen Gruppen hinsichtlich gegebener Kriterien an
- **Intention:**
 - Beseitigung systematischer Unterschiede in Störvariablen in beiden Gruppen
 - Vergleichbarkeit der Gruppen erhöhen
 - Bias minimieren
 - Validität der Ergebnisse maximieren

Matching-Verfahren im Überblick



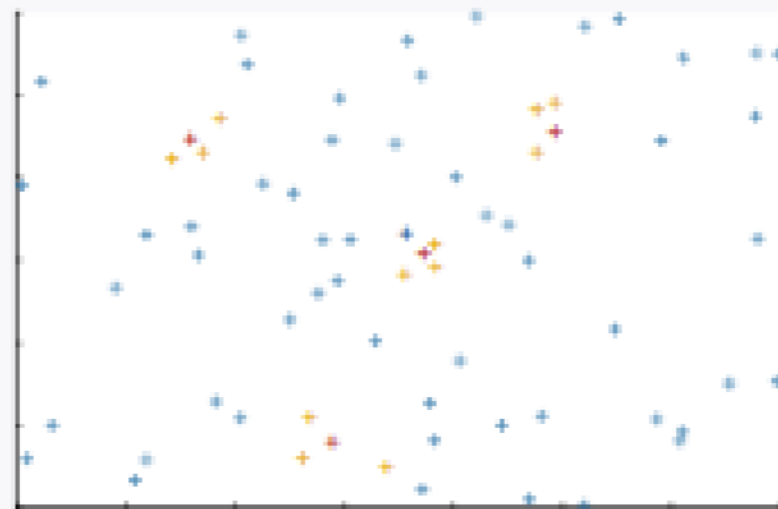
* nur 1:n Matching

Matching-Verfahren im Detail

1. k-Nearest Neighbor Matching:

- Jedem DMP-Teilnehmer wird ein oder werden mehrere Nichtteilnehmer zugeordnet
- Nichtteilnehmer mit der geringsten Distanz zum Teilnehmer werden berücksichtigt

Illustration des k-Nearest Neighbor Matching (k=3)

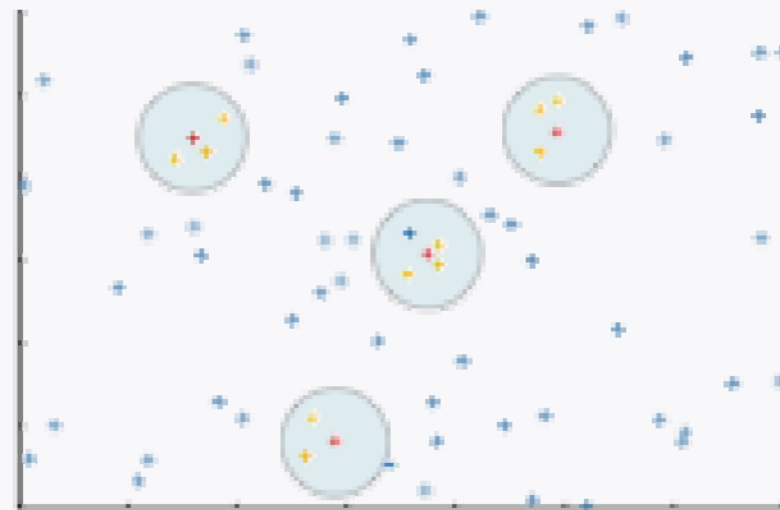


- ★ = Teilnehmer
- ★ = Matchingpartner
- ★ = weitere Versicherte

2. Caliper Matching:

- Funktionalität:
 - ähnlich wie k-Nearest Neighbor Matching
- Toleranzlevel:
 - gibt eine maximale Distanz zwischen den Matching-Partnern vor (Radius um den Fall = "caliper")
 - Risiko für „schlechte“ Matches wird minimiert

Illustration des Caliper Matching (k=3)



- ★ = Teilnehmer
- ★ = Matchingpartner
- ★ = weitere Versicherte

3. Interval Matching:

- Aufteilung der Propensity Score-Verteilung in Intervalle (Strata)
- Phasen:
 - Sortierung der Propensity Scores für die Teilnehmer und Einteilung in Strata
 - Übertragung der Stratagrenzen auf die Scores der Nichtteilnehmer
 - Auszählen der Nichtteilnehmer in den Strata und zufällige Eliminierung überzähliger Stratamitglieder
 - Zusammenfassen der Nichtteilnehmer zur Kontrollgruppe

Schematischer Ablauf des Interval Matching mit 5 Strata

Teilnehmer	Nichtteilnehmer	Teilnehmer	Nichtteilnehmer
0,150204540241104	0,1105584033644	0,150204540241104	0,1105584033644
0,10856436714066	0,2659054287700	0,10856436714066	0,2659054287700
0,203371179940241	0,70597040506468	0,203371179940241	0,70597040506468
0,27240073589358	0,268945403382388	0,27240073589358	0,268945403382388
0,28128160902207	0,26539897930989	0,28128160902207	0,26539897930989
0,34056371292457	0,340340603822868	0,34056371292457	0,340340603822868
0,46838932004258	0,38885738737188	0,46838932004258	0,38885738737188
0,44918645780149	0,40747744382387	0,44918645780149	0,40747744382387
0,328283848148	0,45827918293888	0,328283848148	0,45827918293888
0,54458038988874	0,407321188382178	0,54458038988874	0,407321188382178
0,81103748911287	0,88487868848408	0,81103748911287	0,88487868848408
0,75726378488827	0,78870078879774	0,75726378488827	0,78870078879774
0,8890988418218	0,78870078879774	0,8890988418218	0,78870078879774
0,310211384578228	0,85896778878734	0,310211384578228	0,85896778878734
0,871384904827088	0,88824888879707	0,871384904827088	0,88824888879707

(a) Aufteilung der Teilnehmer in Strata

(b) Übertragung der Stratagrenzen auf die Nichtteilnehmer

Teilnehmer	Nichtteilnehmer	Teilnehmer	Nichtteilnehmer
0,150204540241104	0,1105584033644	0,150204540241104	0,1105584033644
0,10856436714066	0,2659054287700	0,10856436714066	0,2659054287700
0,203371179940241	0,70597040506468	0,203371179940241	0,70597040506468
0,27240073589358	0,268945403382388	0,27240073589358	0,268945403382388
0,28128160902207	0,26539897930989	0,28128160902207	0,26539897930989
0,34056371292457	0,340340603822868	0,34056371292457	0,340340603822868
0,46838932004258	0,38885738737188	0,46838932004258	0,38885738737188
0,44918645780149	0,40747744382387	0,44918645780149	0,40747744382387
0,328283848148	0,45827918293888	0,328283848148	0,45827918293888
0,54458038988874	0,407321188382178	0,54458038988874	0,407321188382178
0,81103748911287	0,88487868848408	0,81103748911287	0,88487868848408
0,75726378488827	0,78870078879774	0,75726378488827	0,78870078879774
0,8890988418218	0,78870078879774	0,8890988418218	0,78870078879774
0,310211384578228	0,85896778878734	0,310211384578228	0,85896778878734
0,871384904827088	0,88824888879707	0,871384904827088	0,88824888879707

(c) Zufällige Eliminierung der überzähligen Stratamitglieder

(d) Zusammenfassen der Nichtteilnehmer zur neuen Kontrollgruppe

Propensity Score Matching

- Ein direktes Matching auf Basis der (vorhandenen) Variablen ist aufgrund der Multidimensionalität wegen begrenzter Performance nicht sinnvoll.
- Deswegen wird das Propensity Score Matching (PSM) benutzt.



- Ein-/ Ausschlüsse aufgrund bestimmter Kriterien
- Anreicherung mit zusätzlichen Daten (Alter, Geschlecht, Vorerkrankungen, ...)

- Abhängige Variable: Teilnehmer: ja/nein
- Unabhängige Variablen: Alter, Geschlecht, Vorerkrankungen, ...

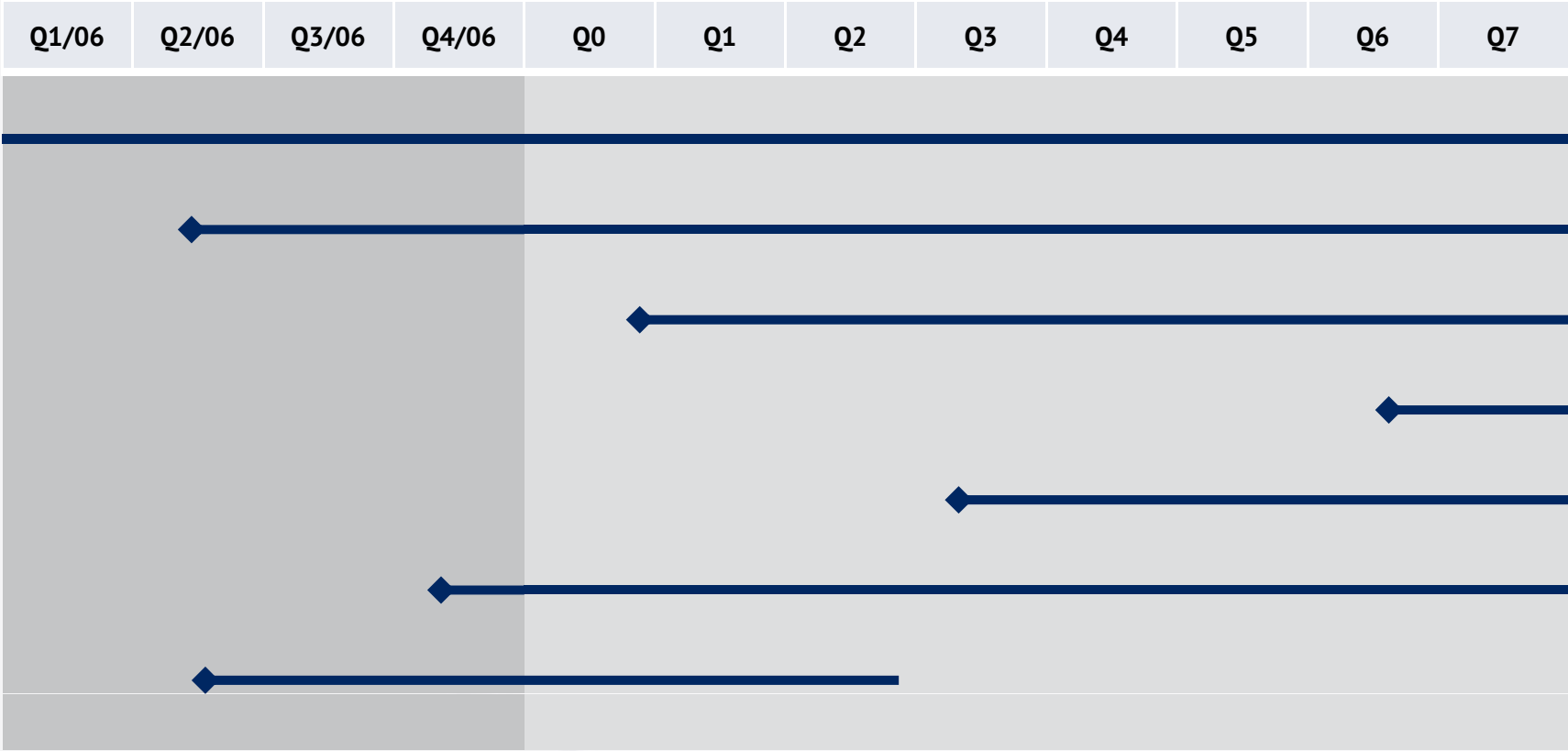
- individ. Berechnungen für jeden Teilnehmer und Nichtteilnehmer auf Grundlage der bestimmten Regressionskoeffizienten und den Variablenausprägungen gemäß:

$$P(X_i) = P(Y = 1 | X_i = x_i) = \frac{\exp(\beta \cdot X_i)}{1 + \exp(\beta \cdot X_i)}$$

Wahrscheinlichkeit, DMP-Teilnehmer zu sein unter der Bedingung, bestimmte Merkmalsausprägungen zu besitzen

Versichertenstichprobe:

84.410 Teilnehmer, 144.910 Nicht-Teilnehmer



◆ Zeitpunkt einer (möglichen) DMP-Einschreibung

Selektionseffekte und Programmeffekte

Bei der Bewertung der vorliegenden Ergebnisse sind grundsätzlich drei mögliche Effekte zu beachten:

1. Selektionseffekt bei Versicherten

d.h. dass sich "gesündere", "gesundheitsbewusstere" oder insgesamt "aktiverere" Versicherte häufiger für ein DMP entscheiden bzw. dass diese Versicherten auch häufiger von ihren behandelnden Ärzten als geeignet für ein DMP eingestuft werden: dabei ist auch eine sozialpolitische Selektion nicht auszuschließen,

2. Selektionseffekt bei den behandelnden Ärzten

d.h. dass bessere Ergebnisse im Rahmen der DMPs im Vergleich zu "Nicht-DMP" den Unterschied in der Versorgung zwischen den bislang aktiv teilnehmenden Ärzten und den eher "DMP-kritischen" abbilden ("early adopters-Phänomen"),

3. "echter" Programmeffekt,

der sich im Verlauf seit der Programmeinführung im Sinne einer kontinuierlichen Verbesserung der Versorgung zeigt.

Ergebnisse

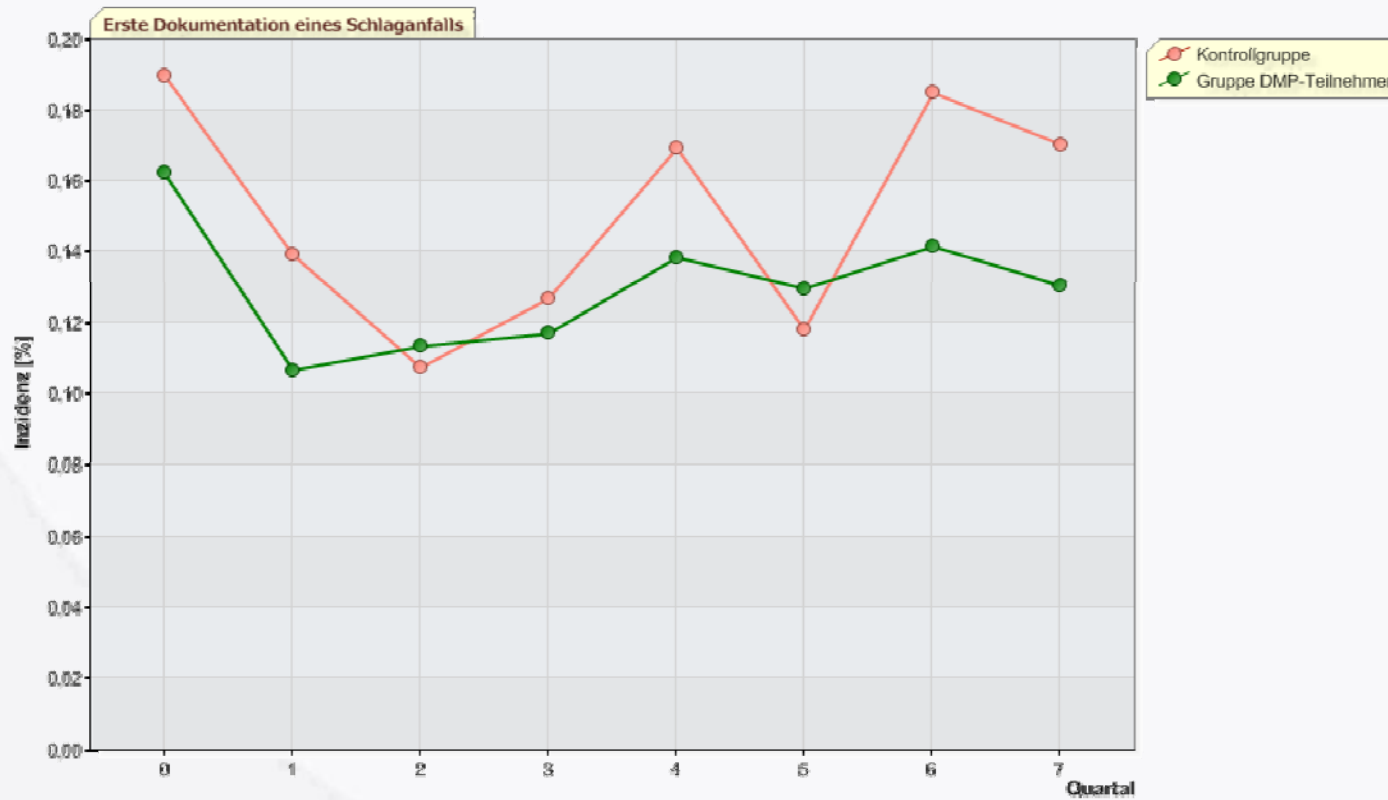
Propensity Score Interval Matching: Effekt

Parameter	Grp. D	Grp. N	p (D vs. N)	Grp. NM	p (D vs. NM)
Herzinfarkt	0.99 %	1.08 %	0.0466	1.00 %	0.9462
Schlaganfall	1.70 %	1.52 %	0.0001	1.70 %	0.9729
Periph. art. Verschlusskrankheit	6.75 %	4.45 %	< 0.0001	6.66 %	0.6211
Erblindung	0.51 %	0.48 %	0.3283	0.49 %	0.6720
Terminale Niereninsuffizienz	0.34 %	0.45 %	< 0.0001	0.33 %	0.9426
Polyneuropathie	6.80 %	1.95 %	0.0000	6.86 %	0.7386
Fußamputation	0.19 %	0.10 %	< 0.0001	0.18 %	0.5902
Pflegestufe 1	1.70 %	2.17 %	< 0.0001	1.70 %	0.9322
Pflegestufe 2	0.84 %	1.48 %	< 0.0001	0.82 %	0.7116
Pflegestufe 3	0.11 %	0.46 %	< 0.0001	0.16 %	0.0619

Zusätzlich wurden die folgenden Variablen berücksichtigt:

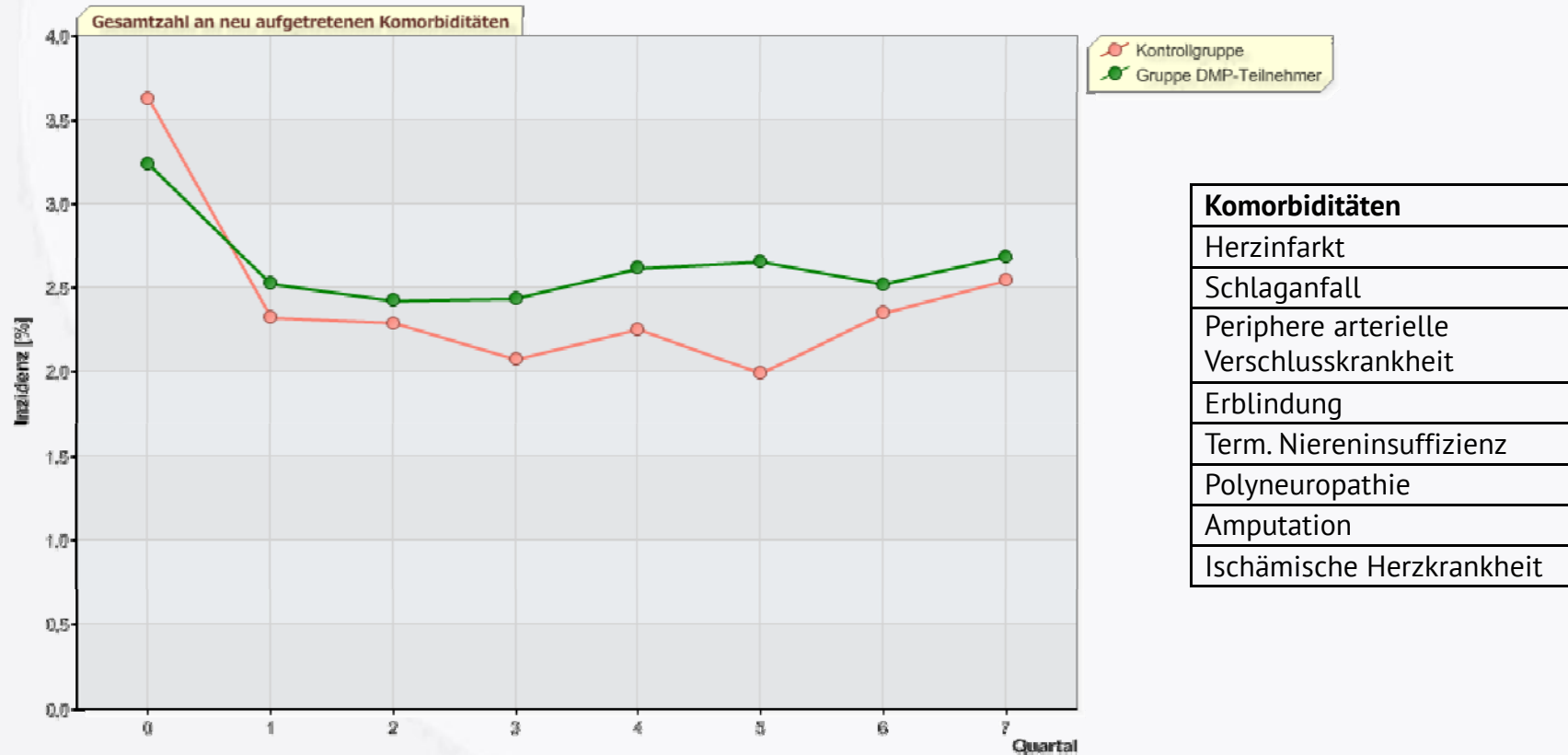
Geschlecht, Ausbildung, Stellung im Beruf, Alter, Medikamentenverbrauch in DDD, Pharma- und Krankenhauskosten

Vergleich der Inzidenz von Schlaganfällen



Die Ergebnisse fallen für die DMP-Teilnehmer geringfügig besser aus.

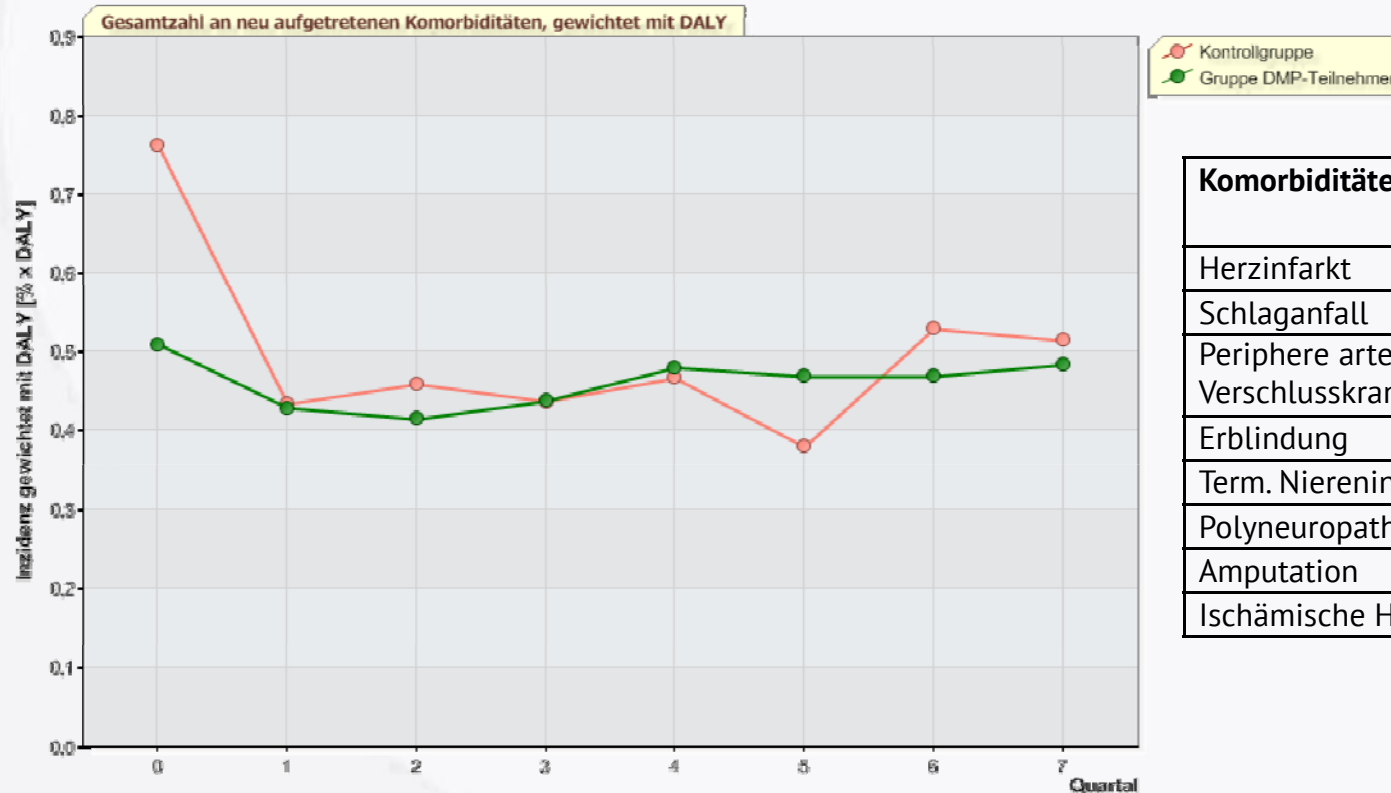
Vergleich von summierten Inzidenzen relevanter Komorbiditäten



Verglichen mit der Kontrollgruppe zeigen die DMP-Teilnehmer in nahezu allen Quartalen des Beobachtungszeitraums schlechtere Ergebnisse.

Ursächlich dafür ist u.a. die vergleichsweise hohe Inzidenz der Polyneuropathie, bei der die Kontrollgruppe bessere Ergebnisse zeigt (unvollständige Dokumentation?).

Vergleich von DALY-gewichteten Inzidenzen rel. Komorbiditäten



Komorbiditäten	Gewicht (DALY)
Herzinfarkt	0,439
Schlaganfall	0,920
Periphere arterielle Verschlusskrankheit	0,000
Erblindung	0,552
Term. Niereninsuffizienz	0,098
Polyneuropathie	0,072
Amputation	0,102
Ischämische Herzkrankheit	0,124

Werden die Inzidenzen mit DALYs (Disability-adjusted life years)* gewichtet, sind die Ergebnisse vergleichbar.

Das DMP zeigt hinsichtlich der Inzidenz relevanter Komorbiditäten weder Vor- noch Nachteile gegenüber der Regelversorgung.

Diskussion der Ergebnisse

- **Uneinheitliches Bild der Programmeffekte:**
 - Nutzen des DMP DM Typ 2 konnte nicht bestätigt werden
 - Effektdauer zu gering?
- **Beibehaltung einzelner Elemente sinnvoll:**
 - Ansprache der DMP-Teilnehmer durch ihre Krankenkasse (Reminding)
 - Individualisierung der DMP mit verhaltensmodifizierenden Elementen für ausgewählte Patienten
 - Unterstützungsangebote zur Förderung der Behandlerkompetenz v.a. in der Gesprächsführung
- **Studienlage verbesserungswürdig:**
 - idealerweise: durch RCT's
 - behelfsweise: durch methodisch anspruchsvolle Routinedatenanalysen

Diskussion der Methodik

- **Kontrollgruppengenerierung**
 - über Propensity Score Matching möglich, wenn RCTs nicht verfügbar sind
- **Minimierung von Selektionseffekten**
 - durch Berücksichtigung einer großen Anzahl an Kovariablen möglich
- **Methodenvergleich**
 - derzeit sind noch keine geeigneten Maßzahlen verfügbar, um verschiedene Matchingverfahren direkt vergleichen zu können

Messung der Güte des Matching

- **In der Literatur am häufigsten genannte statistische Tests:**

- Chi-Quadrat-Test bzw. Wilcoxon-Test

- **Weitere Tests:**

- Standardized Bias

$$SB = \frac{(X_T - X_C)}{\sqrt{1/2 (S_{XT} - S_{XC})}}$$

- Sample Percent Reduction in Bias

$$SPRB = 1 - \frac{|X_{AT} - X_{AC}|}{|X_{IT} - X_{IC}|}$$

Fazit & Ausblick

Fazit

- Studienlage verbessern
- DMP's in ihrer derzeitigen dokumentationslastigen Form in Deutschland nicht praktikabel
- Programme ohne falsche finanzielle Anreize

Ausblick

- Ausweitung der Analysen auf weitere Datensätze
- Erhöhung der Validität
 - Verlängerung des Untersuchungszeitraums, Nutzung weiterer Validierungsoptionen
- Umsetzung der Evaluationen mit weiteren Matchingverfahren
 - Caliper Matching
 - k-Nearest Neighbor Matching
 - Greedy Matching
 - Genetic Matching
 - High-Dimensional-Propensity-Score-Matching

Vielen Dank für Ihre Aufmerksamkeit.



Dr. Susanne Ahrens | www.wineg.de | dr.susanne.ahrens@wineg.de



ORIGINALARBEIT

Nutzen und Effizienz des Disease-Management-Programms Diabetes mellitus Typ 2

Roland Linder, Susanne Ahrens, Dagmar Köppel,
Thomas Hellmann, Frank Verheyen

MEUZIN

ORIGINAL ARTICLE

The Benefit and Efficiency of the Disease Management Program for Type 2 Diabetes

Roland Linder, Susanne Ahrens, Dagmar Köppel,
Thomas Hellmann, Frank Verheyen

MEDICINE