

# Benchmarking von Propensity Score basierten Matchingalgorithmen

Philipp Wewering | Roland Linder

mainz // 2011

56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS)

6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie e.V. (DGEpi)

Mainz | 29. September 2011

# Inhalt

- Motivation und Fragestellungen
- Methodik
- Datenbasis & Benchmarking
- Diskussion & Ausblick

# Motivation und Fragestellungen

- Das WINEG hat unter anderem die Aufgabe, die Wirksamkeit von präventiven, diagnostischen und therapeutischen Maßnahmen zu bewerten, insbesondere unter Berücksichtigung von Kosten-Nutzen-Aspekten.



- Häufig ist eine Anwendung von randomisierten, kontrollierten Studien (RCT) nicht möglich.
- Eine alternative Methodik ist das Propensity Score Matching (PSM).

# Methodik

## Ablauf des Propensity Score Matchings



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

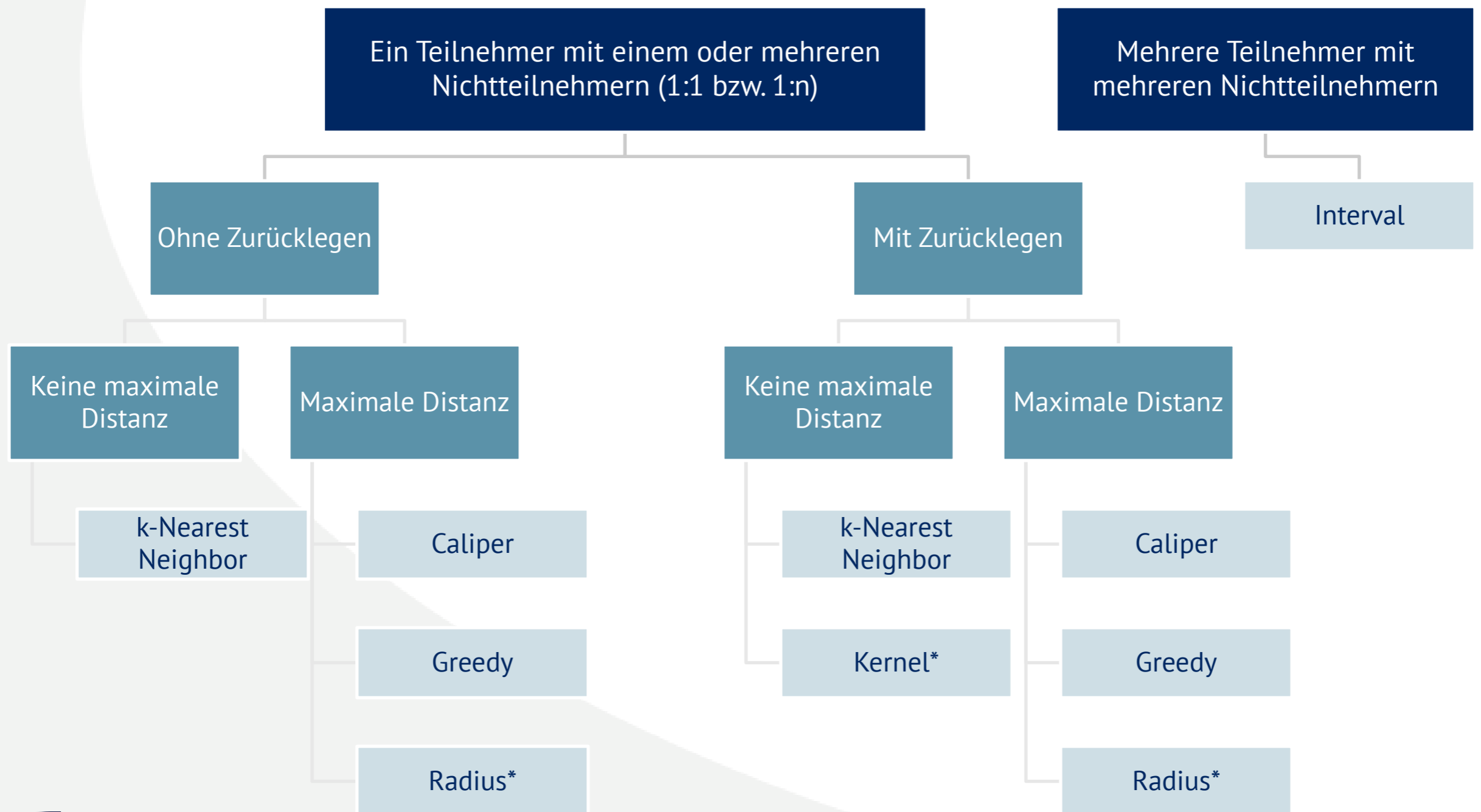
- Abhängige Variable: Teilnehmer: ja/nein
- Unabhängige Variablen: Alter, Geschlecht, Vorerkrankungen, ...

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

$$P_i(X) = \frac{\exp(\beta \cdot X_i)}{1 + \exp(\beta \cdot X_i)}$$

# Methodik

## Matchingverfahren



\* nur 1:n Matching

# Methodik

## Matchingverfahren

### k-Nearest Neighbor Matching

- Jedem Teilnehmer  $i$  der behandelten Gruppe werden ein oder mehrere Nichtteilnehmer  $j$  als Matchingpartner zugeordnet.
- Hierbei werden jeweils die Nichtteilnehmer berücksichtigt, die die geringste Distanz zum beobachteten Teilnehmer aufweisen.

### Caliper Matching

- Jedem Teilnehmer  $i$  der behandelten Gruppe werden ein oder mehrere Nichtteilnehmer  $j$  als Matchinpartner zugeordnet.
- Hierbei werden jeweils die Nichtteilnehmer berücksichtigt, die die geringste Distanz zum beobachteten Teilnehmer aufweisen. Zusätzlich muss sich jeder Partner innerhalb eines definierten Radius (= Caliper) um den Teilnehmer befinden.

# Methodik

## Matchingverfahren

### Greedy Matching

- Die Idee liegt darin, zunächst Kontrollpersonen als Matchingpartner zu finden, die eine hohe Übereinstimmung in den Scores haben.
- Anschließend werden sukzessive für die noch "partnerlosen" Teilnehmer Matchingpartner gefunden, deren Übereinstimmung nicht mehr so groß ist.
- Ein übliches Maß ist hierbei die Genauigkeit der Übereinstimmung in der Anzahl der Nachkommastellen (6 => 3 Matching).

### Interval Matching

- Die Teilnehmer werden anhand ihrer Scores sortiert und in gleich große Strata eingeteilt (z.B. Decentile).
- Die sich hieraus ergebenden Grenzen werden auf die Gruppe der Nichtteilnehmer übertragen und anschließend das Stratum mit der geringsten Anzahl an Personen innerhalb der Nichtteilnehmergruppe bestimmt.
- Anschließend werden die überzähligen Nichtteilnehmer der anderen Strata zufällig gelöscht.

# Methodik

## Messung der Matchinggüte

### Statistische Verfahren

- Die in der Literatur am häufigsten genutzten Verfahren sind:
  - Chi-Quadrat-Test bzw. Wilcoxon-Test.
- Weitere genutzte Tests:
  - Standardized Bias

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

- Sample Percent Reduction in Bias

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

# Datenbasis

## DMP T2DM

- Daten zum DMP Diabetes Mellitus Typ 2 der TK
- 84 410 Teilnehmer bzw. 144 910 Nichtteilnehmer
- Daten zu Alter, Geschlecht, Pflegestufen, Komorbiditäten, etc. (19 Variablen)

## NSW PSID

- Daten zum National Supported Work-Programm
- 185 Teilnehmer bzw. 2 490 Nichtteilnehmer
- Daten zu Alter, Ausbildung, Familienstand, Einkommen, etc. (8 Variablen)

## NSW CPS

- Daten zum National Supported Work-Programm
- 185 Teilnehmer bzw. 15 992 Nichtteilnehmer
- Daten zu Alter, Ausbildung, Familienstand, Einkommen, etc. (8 Variablen)

# Benchmarking

## Verglichene Matchingverfahren:

- k-Nearest Neighbor Matching mit 1 Partner und ohne Mehrfachverwendung
- k-Nearest Neighbor Matching mit 2 Partnern und ohne Mehrfachverwendung
- Caliper Matching mit einem Caliper von 0,002 und 1 Partner, ohne Mehrfachverwendung
- automatisiertes Caliper Matching mit Matchingquote von 95%
- Interval Matching mit 10 Intervallen
- Interval Matching mit 100 Intervallen
- Interval Matching mit 1000 Intervallen
- Greedy Matching,  $6 \Rightarrow 3$

# Benchmarking Testergebnisse

Sample Percent Reduction in Bias - DMP

Variable	kNN1	kNN2	Gr	C1	Co1	I10	I100	I1000
GESCHL	0,4453198	0,0567842	0,9323709	0,9541379	0,9460212	0,8009227	0,8916243	0,6173826
AGE	0,9334508	0,0444287	0,7542776	0,7730660	0,7788642	0,9900906	0,8024936	0,7781328
A1	-3,7270688	-0,9932554	0,2150385	0,2776182	0,3335873	0,6469196	-0,5132458	0,3022054
A2	0,9254453	0,0669973	0,7918151	0,8861008	0,8883215	0,9715951	0,7568170	0,9177520
A3	0,9170889	0,0163886	0,8318693	0,8529928	0,8565194	0,9110673	0,8420477	0,8867280
B1	-15,5451822	0,0552792	0,8029641	0,7930549	0,7969375	0,9354514	0,8004925	0,8456951
B2	4,7895412	-0,0535948	0,7735577	0,7216242	0,7049802	0,4637348	0,7512315	0,6447013
B3	4,4316157	0,2250068	0,8968536	0,9400380	0,9400068	0,7859303	0,9539635	0,9660936
B4	-0,8082041	0,0315271	0,8517943	0,8638776	0,8745960	0,8432018	0,8532046	0,9194791
B5	-5,2615342	0,1334654	0,9195051	0,9058679	0,9168931	0,9504859	0,5673020	0,9162699
B6	9,6674949	0,0180361	0,8499690	0,9569752	0,9549038	0,7410112	0,9223949	0,8329788
P1	-3,4152063	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
P2	-5,9373413	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649	0,8798649
P3	-6,4526769	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
HERZINFARKT	-0,8663810	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
SCHLAGANFALL	1,4178180	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
PAVK	10,0184788	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
ERBLINDUNG	0,4216310	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
NIERENINSUFFIZIENZ	-1,8439354	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
POLYNEUROPATHIE	23,8622172	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
FUßAMPUTATION	2,3543958	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
ZAHL_BTR	0,0705379	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
SUM_DDD	47,0080728	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649
SUM_NETTO_M	14,9213315	0,0149291	0,8712628	0,8566342	0,8527849	0,6542133	0,9149048	0,8798649

Standardised Bias (I) - DMP

Variable	kNN 1		kNN 2		GR	
	SB before	SB after	SB before	SB after	SB before	SB after
GESCHL	2,6855052	1,4921399	2,6855052	2,5324274	2,6855052	0,1817347
AGE	8,0961642	0,5501902	8,0961642	7,7081496	8,0961642	-2,0270271
A1	0,3990021	-1,8709935	0,3990021	0,7941857	0,3990021	0,3120117
A2	-3,9284467	0,3156046	-3,9284467	-3,6493982	-3,9284467	0,8696309
A3	-7,5477180	0,6657520	-7,5477180	-7,4173035	-7,5477180	1,3345151
B1	-15,5451822					
B2	4,7895412					
B3	4,4316157					
B4	-0,8082041					
B5	-5,2615342					
B6	9,6674949					
P1	-3,4152063					
P2	-5,9373413					
P3	-6,4526769					
HERZINFARKT	-0,8663810					
SCHLAGANFALL	1,4178180					
PAVK	10,0184788					
ERBLINDUNG	0,4216310					
NIERENINSUFFIZIENZ	-1,8439354					
POLYNEUROPATHIE	23,8622172					
FUßAMPUTATION	2,3543958					
ZAHL_BTR	0,0705379					
SUM_DDD	47,0080728					
SUM_NETTO_M	14,9213315					

Chi Square /Wilcoxon Mann Whitney (II) - DMP

Variable	C1		Co1		Cr1	
	p-Value before	p-Value after	p-Value before	p-Value after	p-Value before	p-Value after
GESCHL	<0,0001	0,8092922	<0,0001	0,7764157	<0,0001	0,4917918
AGE	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
A1	0,3566004	0,5735252	0,3566004	0,6036666	0,3566004	0,5180300
A2	<0,0001	0,3534287	<0,0001	0,3630608	<0,0001	0,2901551
A3	<0,0001	0,0224746	<0,0001	0,0259338	<0,0001	0,0263497
B1	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
B2	<0,0001	0,0077727	<0,0001	0,0077551	<0,0001	0,0092674
B3	<0,0001	0,0207491	<0,0001	0,0142452	<0,0001	0,0179467
B4	0,0637993	0,5629148	0,0637993	0,4738462	0,0637993	0,5855842
B5	<0,0001	0,4332998	<0,0001	0,4332995	<0,0001	0,4005076
B6	<0,0001	0,0092395	<0,0001	0,0164981	<0,0001	0,0076455
P1	<0,0001	0,5081590	<0,0001	0,5595724	<0,0001	0,7120354
P2	<0,0001	0,5689935	<0,0001	0,5503127	<0,0001	0,4297607
P3	<0,0001	0,0117829	<0,0001	0,0095827	<0,0001	0,0092706
HERZINFARKT	0,0465952	0,3293804	0,0465952	0,3163458	0,0465952	0,4874203
SCHLAGANFALL	0,0009728	0,4324161	0,0009728	0,3431824	0,0009728	0,4551534
PAVK	<0,0001	0,7133194	<0,0001	0,6813051	<0,0001	0,6345024
ERBLINDUNG	0,3282715	0,9137786	0,3282715	0,9137786	0,3282715	0,7725602
NIERENINSUFFIZIENZ	<0,0001	0,2386722	<0,0001	0,2734331	<0,0001	0,2013986
POLYNEUROPATHIE	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
FUßAMPUTATION	<0,0001	0,7927724	<0,0001	0,8950823	<0,0001	0,9468080
ZAHL_BTR	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
SUM_DDD	<0,0001	0,00190697	<0,0001	0,00218291	<0,0001	0,0020682
SUM_NETTO_M	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

www.wineg.de



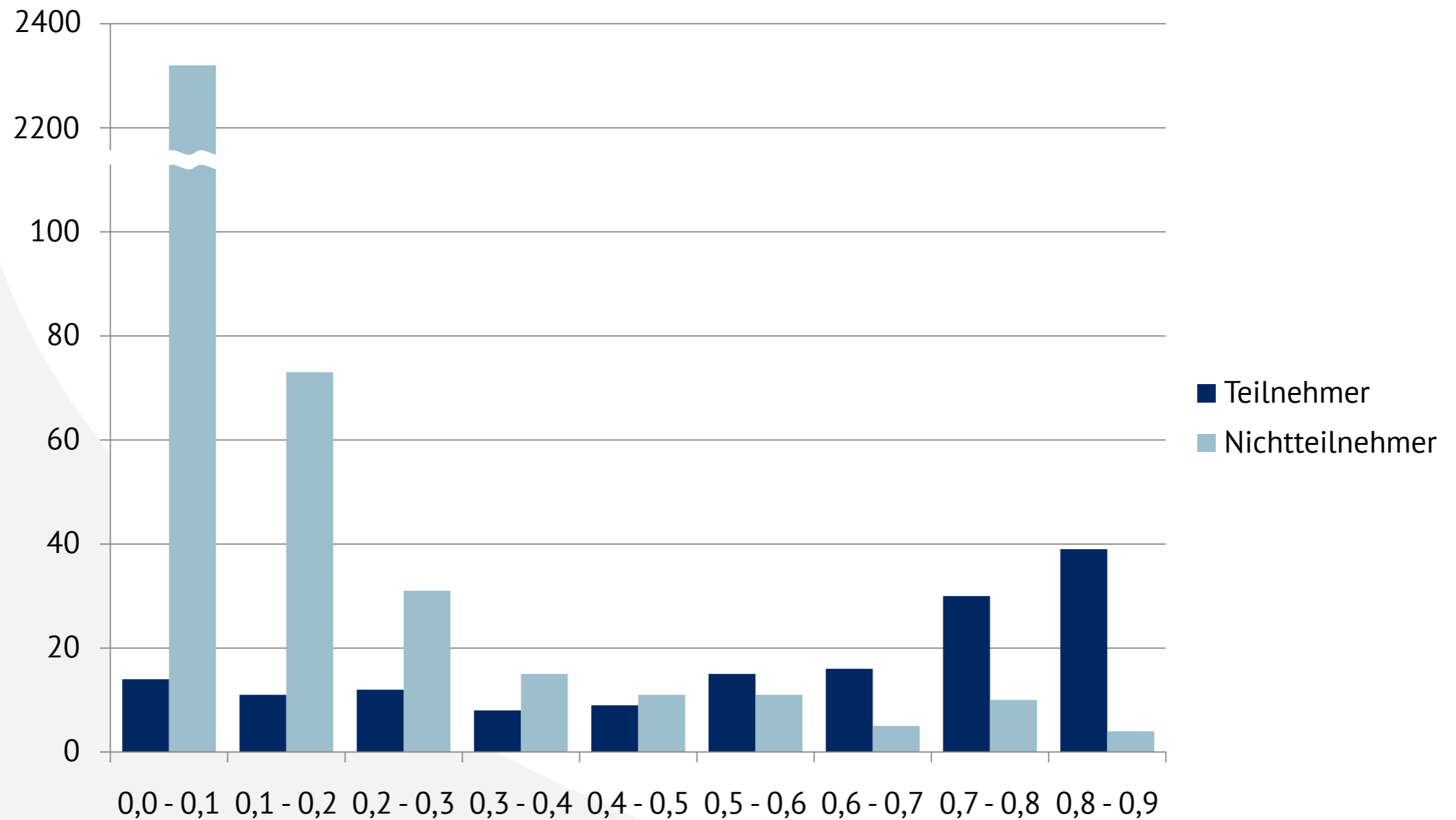
<b>DMP T2DM</b>	<b>Gruppengröße</b> 84 410 T 144 910 NT	<b>Wilcoxon-/Chi<sup>2</sup></b> 19 Variablen mit signifikanten Unterschieden	<b>SPRB</b>	<b>Standardized Bias</b>
<b>kNN-1</b>	84 410 T 84 410 NT	8 Variablen mit sign. Unterschieden	2 Variable mit bestem, 1 Variable mit niedrigstem Ergebnis	2 Variablen mit Verschlechterung
<b>kNN-2</b>	72 455 T 72 455 NT	18 Variablen mit sign. Unterschieden	1 Variable mit bestem, 20 Variablen mit niedrigstem Ergebnis	5 Variablen mit Verschlechterung
<b>Caliper (0.002   1)</b>	76 749 T 76 749 NT	5 Variablen mit sign. Unterschieden	4 Variablen mit bestem Ergebnis	1 Variable mit Verschlechterung
<b>Caliper (optimiert   1)</b>	76 709 T 76 709 NT	5 Variablen mit sign. Unterschieden	1 Variable mit bestem Ergebnis	1 Variable mit Verschlechterung
<b>Interval (10)</b>	84 410 T 55 470 NT	9 Variablen mit sign. Unterschieden	5 Variablen mit bestem, 1 Variable mit niedrigstem Ergebnis	1 Variable mit Verschlechterung
<b>Interval (100)</b>	84 410 T 31 800 NT	6 Variablen mit sign. Unterschieden	3 Variablen mit bestem Ergebnis	2 Variablen mit Verschlechterung
<b>Interval (1000)</b>	84 410 T 15 984 NT	4 Variablen mit sign. Unterschieden	5 Variablen mit bestem, 2 Variablen mit niedrigstem Ergebnis	1 Variable mit Verschlechterung
<b>Greedy (6 =&gt; 3)</b>	76 466 T 76 466 NT	5 Variablen mit sign. Unterschieden	2 Variablen mit bestem Ergebnis	1 Variable mit Verschlechterung

<b>NSW CPS</b>	<b>Gruppengröße</b> 185 T 15 922 NT	<b>Wilcoxon-/Chi<sup>2</sup></b> 7 (von 8) Variablen mit signifikanten Unterschieden	<b>SPRB</b>	<b>Standardized Bias</b>
<b>kNN-1</b>	185 T 185 NT	1 Variable mit sign. Unterschieden		1 Variable mit Verschlechterung
<b>kNN-2</b>		1 Variable mit sign. Unterschieden		1 Variable mit Verschlechterung
<b>Caliper (0.002   1)</b>	152 T 152 NT	1 Variable mit sign. Unterschieden		1 Variable mit Verschlechterung
<b>Caliper (optimiert   1)</b>	167 T 167 NT	1 Variable mit sign. Unterschieden	1 Variable mit bestem Ergebnis	1 Variable mit Verschlechterung
<b>Interval (10)</b>	185 T 150 NT	Keine Variablen mit sign. Unterschieden	1 Variable mit bestem Ergebnis	1 Variable mit Verschlechterung
<b>Interval (100)</b>	185 T 74 NT	2 Variablen mit sign. Unterschieden	1 Variable mit niedrigstem Ergebnis	1 Variable mit Verschlechterung
<b>Interval (1000)</b>	185 T 10 214 NT	7 Variablen mit sign. Unterschieden	7 Variablen mit niedrigstem Ergebnis	7 Variablen mit Verschlechterung
<b>Greedy (6 =&gt; 3)</b>	115 T 115 NT	Keine Variablen mit sign. Unterschieden	3 Variablen mit bestem Ergebnis	1 Variable mit Verschlechterung

<b>NSW PSID</b>	<b>Gruppengröße</b> 185 T 2 597 NT	<b>Wilcoxon-/Chi<sup>2</sup></b> 7 (von 8) Variablen mit signifikanten Unterschieden	<b>SPRB</b>	<b>Standardized Bias</b>
<b>kNN-1</b>	154 T 154 NT	3 Variable mit sign. Unterschieden		keine Variable mit Verschlechterung
<b>kNN-2</b>	154 T 308 NT	8 Variable mit sign. Unterschieden		5 Variablen mit Verschlechterung
<b>Caliper (0.002   1)</b>	41 T 41 NT	keine Variablen mit sign. Unterschieden	2 Variablen mit bestem Ergebnis	1 Variable mit Verschlechterung
<b>Caliper (optimiert   1)</b>	148 T 148 NT	3 Variablen mit sign. Unterschieden	1 Variable mit bestem Ergebnis	keine Variable mit Verschlechterung
<b>Interval (10)</b>	154 T 8 NT	8 Variable mit sign. Unterschieden	5 Variable mit niedrigstem Ergebnis	4 Variablen mit Verschlechterung
<b>Interval (100)</b>	154 T 61 NT	8 Variable mit sign. Unterschieden		4 Variablen mit Verschlechterung
<b>Interval (1000)</b>	154 T 1 472 NT	7 Variablen mit sign. Unterschieden	3 Variablen mit niedrigstem Ergebnis	7 Variablen mit Verschlechterung
<b>Greedy (6 =&gt; 3)</b>	32 T 32 NT		5 Variablen mit bestem Ergebnis	keine Variable mit Verschlechterung

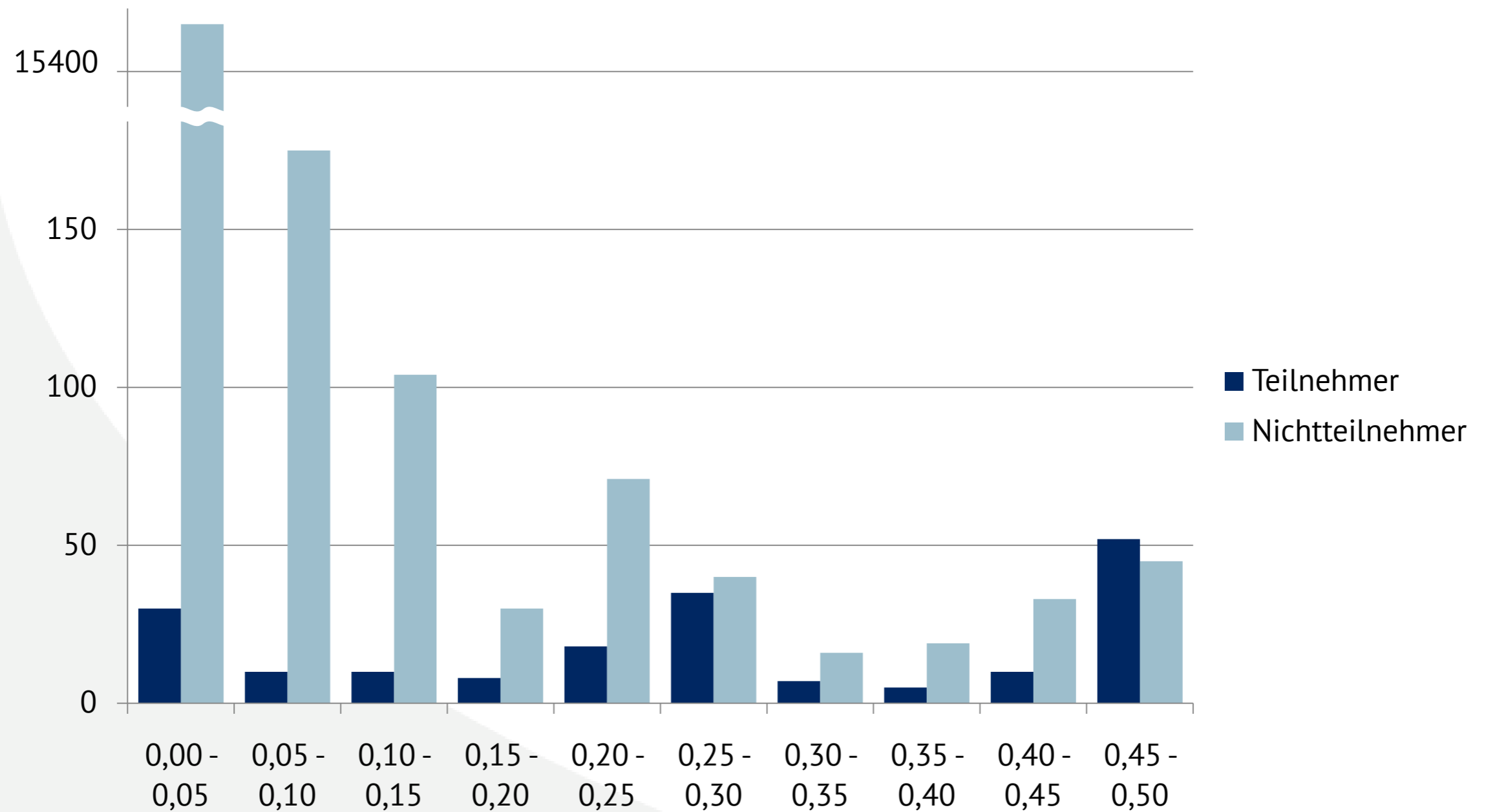
# Benchmarking

## Propensity-Score-Verteilung des Datensatzes NSW-PSID



# Benchmarking

## Propensity-Score-Verteilung des Datensatzes NSW-CPS



# Diskussion & Ausblick

## Diskussion

- In Ermangelung von RCTs ermöglicht das Propensity Score Matching die Generierung einer Kontrollgruppe.
- Hierbei kann eine große Anzahl an Kovariablen zur Minimierung von Selektionseffekten beitragen.
- Für einen direkten Vergleich verschiedener Matchingverfahren fehlt es derzeit an geeigneten Maßzahlen.
- Insgesamt betrachtet liefern Caliper und Greedy Matching die robustesten Ergebnisse hinsichtlich Variablenanzahl und Gruppengrößen.
- Der Anwender hat aber immer noch viele Stellschrauben zur Optimierung des Matching und sollte diese auch nutzen

## Ausblick

- Ausweitung der Analysen auf weitere Datensätze
- Realisierung und Benchmarking weiterer Matchingverfahren (Genetic Matching, High-Dimensional-Propensity-Score-Matching)

# Vielen Dank für Ihre Aufmerksamkeit!



Philipp Wewering | [philipp.wewering@wineg.de](mailto:philipp.wewering@wineg.de)