ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Schneider, Hilmar

Working Paper

Sind ARGEn wirklich besser? Anmerkungen zur Evaluation der Betreuungsorganisation im Bereich der Grundsicherung

IZA Standpunkte, No. 16

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Schneider, Hilmar (2009) : Sind ARGEn wirklich besser? Anmerkungen zur Evaluation der Betreuungsorganisation im Bereich der Grundsicherung, IZA Standpunkte, No. 16, Forschungsinstitut zur Zukunft der Arbeit (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/91818

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

STANDPUNKTE

IZA Standpunkte Nr. 16

Sind ARGEn wirklich besser? Anmerkungen zur Evaluation der Betreuungsorganisation im Bereich der Grundsicherung

Hilmar Schneider

Juli 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Sind ARGEn wirklich besser? Anmerkungen zur Evaluation der Betreuungsorganisation im Bereich der Grundsicherung

Hilmar Schneider
IZA

IZA Standpunkte Nr. 16
Juli 2009

IZA

Postfach 7240
53072 Bonn

Tel.: (0228) 3894-0
Fax: (0228) 3894-180
E-Mail: iza@iza.org

Die Schriftenreihe “IZA Standpunkte” veröffentlicht politikrelevante Forschungsarbeiten und Diskussionsbeiträge von IZA-Wissenschaftlern, IZA Research Fellows und IZA Research Affiliates in deutscher Sprache. Die Autoren sind für den Inhalt der publizierten Arbeiten verantwortlich. Im Interesse einer einheitlichen Textzirkulation werden Aktualisierungen einmal publizierter Arbeiten nicht an dieser Stelle vorgenommen, sondern sind gegebenenfalls nur über die Autoren selbst erhältlich.

IZA Standpunkte Nr. 16
Juli 2009

ZUSAMMENFASSUNG

Sind ARGEn wirklich besser? Anmerkungen zur Evaluation der Betreuungsorganisation im Bereich der Grundsicherung*

Mit der Einführung von Arbeitslosengeld II wurde auch die behördliche Zuständigkeit für die Grundsicherung einer grundlegenden Reorganisation unterzogen. Während in der überwiegenden Mehrheit der Kommunen in Form der sogenannten Arbeitsgemeinschaften eine Mischverwaltung aus Bundesagentur für Arbeit und örtlichen Kommunen geschaffen wurde, wurde den sogenannten Optionskommunen eine eigenständige Verwaltung zugestanden. Die Verwaltungspraxis sollte zeigen, welche von beiden Verwaltungsformen die effektivere Betreuung von erwerbsfähigen Bedürftigen gewährleistet. Deshalb legte der Gesetzgeber fest, dass beide Verwaltungsformen bis Ende 2008 evaluiert werden sollten. In dem entsprechenden Evaluationsbericht des Bundesministeriums für Arbeit und Soziales wird festgestellt, dass die Arbeitsgemeinschaften im Großen und Ganzen besser abschneiden als die Optionskommunen. Das vorliegende Papier setzt sich kritisch mit diesem Bericht auseinander. Es zeigt, dass das verwendete Untersuchungsdesign keine eindeutigen Schlussfolgerungen zulässt. Paradoxerweise könnten die vordergründig für die Arbeitsgemeinschaften sprechenden Ergebnisse sogar ein Beleg dafür sein, dass in Wahrheit die Optionskommunen effektiver arbeiten.

JEL-Codes: C31, C81, D73, H43, H75, I38, J48

Schlagworte: Soziale Sicherung, Grundsicherung, Sozialhilfe, Evaluation, Hartz-Reform

Kontaktadresse:

Hilmar Schneider
IZA
Postfach 7240
D-53072 Bonn
E-Mail: schneider@iza.org

* Ich danke meinen Kollegen Ulf Rinne und Arne Uhlendorff vom IZA für Ihre wertvollen Hinweise zur Erstellung dieses Textes, der sich der schwierigen Aufgabe widmet, ein keineswegs unwichtiges methodisches Argument in eine weitgehend inhaltlich geführte Debatte einzubringen.

Ein eindeutiges Evaluationsergebnis?

Mit der zum 01. Januar in Kraft getretenen Reform der Grundsicherung (Hartz IV) wurde auch die institutionelle Form der Betreuung von Hilfebedürftigen einer grundlegenden Änderung unterzogen. Bis dahin oblag die Betreuung von Sozialhilfebeziehern den Kommunen, während die Betreuung von Arbeitslosenhilfebeziehern in der Verantwortung der Bundesagentur für Arbeit lag. Im Zuge der faktischen Abschaffung der Arbeitslosenhilfe war seitens der Bundesregierung ursprünglich geplant gewesen, den Verantwortungsbereich der Bundesagentur für Arbeit auf sämtliche erwerbsfähigen Hilfebedürftige auszuweiten. Die kommunale Zuständigkeit sollte auf nicht erwerbsfähige Hilfebedürftige begrenzt werden. Nach massiven Protesten der Kommunen setzten die Länder im Vermittlungsausschuss jedoch schließlich eine Mischverwaltung von Kommunen und Bundesagentur für Arbeit in Form der sogenannten Arbeitsgemeinschaften (ARGEn) durch und für eine begrenzte Zahl von sogenannten Optionskommunen (69) wurde darüber hinaus im Rahmen der sogenannten Experimentierklausel (§ 6 SGB II) eine autonome kommunale Zuständigkeit für erwerbsfähige Hilfebedürftige vereinbart. Der Gesetzgeber behielt sich allerdings vor, die Effektivität beider Verwaltungsformen nach einer Anlaufphase auf den Prüfstand zu stellen. Mit Hilfe wissenschaftlich abgesicherter Erkenntnisse sollte bis Ende 2008 ermittelt werden, welche der beiden neu zu schaffenden Organisationsformen effektiver arbeitet, um anschließend auf dieser Grundlage ggfs. eine einheitliche Verwaltungsform beschließen zu können.

Unter der Regie des Bundesministeriums für Arbeit und Soziales wurde 2006 eine breit angelegte Forschungsinitiative gestartet, um dem in der Experimentierklausel festgeschriebenen Evaluationsauftrag der Bundesregierung nachzukommen. In insgesamt vier Untersuchungsfeldern (Deskriptive Analyse und Matching, Implementations- und Governanceanalyse, Wirkungs- und Effizienzanalyse, Makroanalyse und regionale Vergleiche) war eine Vielzahl von Wissenschaftlern an der Durchführung beteiligt.[1] Zugespitzt lautete die Kernfrage: ARGEn oder Optionskommunen: Wer kann es besser?

Der Ende 2008 vorgelegte Abschlussbericht kommt zu dem Schluss, dass die ARGEn im Hinblick auf die Überwindung der Hilfebedürftigkeit im Durchschnitt bis-

---

[1] Der Abschlussbericht ist als Bundestagsdrucksache Nr. 16/11488 veröffentlicht (http://www.bmas.de/coremedia/generator/31070/f390__forschungsbericht.html).
Die entsprechenden Teilberichte sind unter folgenden Web-Adressen erhältlich:
UF 1: http://www.bmas.de/coremedia/generator/30142/f385__forschungsbericht.html
UF 2: http://www.bmas.de/coremedia/generator/30142/f386__forschungsbericht.html
UF 3: http://www.bmas.de/coremedia/generator/30142/f387__forschungsbericht.html
UF 4: http://www.bmas.de/coremedia/generator/30142/f388__forschungsbericht.html

lang besser abschneiden als Optionskommunen. Zu diesem Zeitpunkt war der Vergleich zwischen ARGEn und Optionskommunen durch ein Ende 2007 ergangenes Urteil des Bundesverfassungsgericht allerdings bereits obsolet geworden. Das Gericht betrachtet die Mischverwaltung von Bundesbehörden und Kommunen als verfassungswidrigen Eingriff in das Recht der Kommunen auf eigenverantwortliche Aufgabenerledigung. Für eine Neuregelung setzte es dem Gesetzgeber eine Frist bis Ende 2010 (Az.:2 BvR 2433/04).

Die Politik steht also nach wie vor in der Pflicht, eine einheitliche Organisationsform zu schaffen. Insofern können die vorliegenden Forschungsergebnisse zumindest grundsätzlich Anhaltspunkte dafür liefern, worauf bei einer künftigen Organisationsform geachtet werden sollte. Allerdings stößt man hier schnell an Grenzen der Belastbarkeit der Ergebnisse. Insbesondere das Untersuchungsdesign im zentralen Untersuchungsfeld 3 (Wirkungs- und Effizienzanalyse) kann im Hinblick auf die Ausgangsfrage nicht für sich in Anspruch nehmen, eindeutige Schlussfolgerungen zuzulassen.

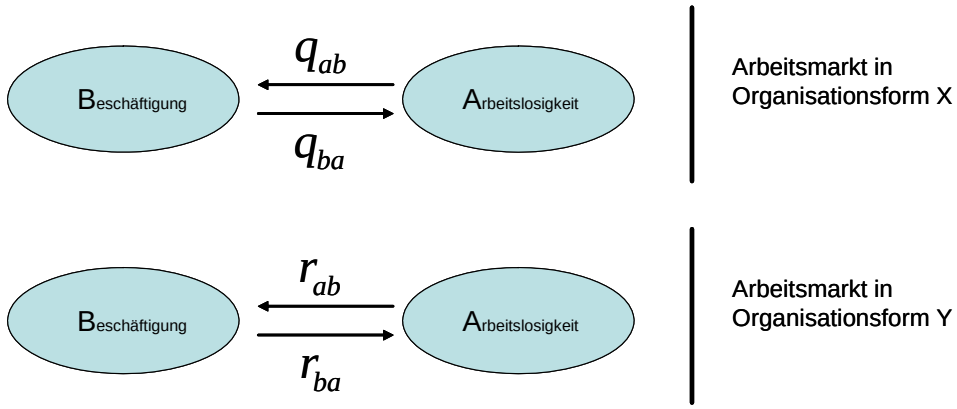
Die durchgeführte Wirkungsanalyse beruht im Kern auf einer Bestandsstichprobe von Hilfebeziehern aus ARGEn und Optionskommunen, die im Jahr 2006 im Hilfebezug standen und befragt wurden. Der ausgewählte Personenkreis wurde im Zeitraum 11/2007 bis 3/2008 erneut befragt, wobei der Verbleib im Hilfebezug als zentrales Messkriterium ermittelt wurde. Mit Hilfe verschiedener ökonometrischer Verfahren wurde auf dieser Grundlage die Wahrscheinlichkeit des Hilfebezugs modelliert. Zusammengefasst lautet das Ergebnis, dass Hilfebezieher in Optionskommunen im Durchschnitt länger im Hilfebezug verbleiben als Hilfebezieher in vergleichbaren ARGEn. Daraus leitet sich die Schlussfolgerung ab, dass Optionskommunen ihre Aufgabe weniger effektiv erfüllen als ARGEn. De facto ist eine solche Schlussfolgerung jedoch nur dann zulässig, wenn bestimmte Nebenbedingungen erfüllt sind, deren Gültigkeit in der Untersuchung aber nicht hinreichend überprüft wurden.

Auf den ersten Blick erscheint es zwar naheliegend, dass ein relativ niedriger Anteil von Personen, die im Hilfebezug verbleiben, als Ausdruck für besonders effiziente Vermittlung betrachtet werden kann. Unter bestimmten Voraussetzungen, die nicht sehr weit hergeholt sein müssen, kann jedoch gerade ein hoher Anteil von Langzeitbedürftigen Ausdruck für eine insgesamt effektivere Vermittlung sein. Der Vergleich der Anteile von Langzeitbedürftigen stellt somit kein hinreichendes Kriterium für die Beurteilung der Effektivität der Vermittlung der untersuchten Organisationsformen dar. Im Gegenteil: Ein solcher Vergleich kann unter Umständen sogar zu einer falschen Beurteilung führen. Dies soll im Folgenden anhand einer etwas formaleren Betrachtung beispielhaft erläutert werden.

Übergangswahrscheinlichkeiten statt Bestandsquoten als Messkriterium

Bewegungsprozesse am Arbeitsmarkt lassen sich in vereinfachter Form als Wechsel von Individuen zwischen den beiden Zuständen Arbeitslosigkeit (*A*) und Beschäftigung (*B*) darstellen (vgl. Abb. 1). Die Intensität, mit der Wechsel zwischen beiden Zuständen erfolgen, lässt sich anhand sogenannter bedingter Übergangswahrscheinlichkeiten zwischen den beiden Zuständen beschreiben. In Abb. 1 sind diese als q_{ab}, q_{ba}, r_{ab} und r_{ba} für die beiden prototypischen Organisationsformen *X* und *Y* bezeichnet. So bezeichnet beispielsweise die bedingte Übergangswahrscheinlichkeit q_{ab} die Wahrscheinlichkeit, dass eine Person, die aktuell arbeitslos ist, in der nächsten Periode in eine Beschäftigung eintreten wird. In der Abbildung ist angenommen, dass in den Bereichen der beiden Organisationsformen jeweils spezifische Bedingungen gelten. Die Übergangswahrscheinlichkeiten in Organisationsform *X* (q_{ab} und q_{ba}) müssen daher nicht mit den entsprechenden Übergangswahrscheinlichkeiten in Organisationsform *Y* (r_{ab} und r_{ba}) übereinstimmen.

Abb. 1: Schematische Darstellung von Arbeitsmarktübergängen



Die Wahrscheinlichkeit arbeitslos und damit hilfebedürftig zu sein, bzw. die Wahrscheinlichkeit beschäftigt zu sein, ist in beiden Systemen vollständig von den bedingten Übergangswahrscheinlichkeiten determiniert. So ergibt sich die Wahrscheinlichkeit im Umfeld von Organisationsform *X* arbeitslos zu sein als $P(a|X) = q_{ba} / (q_{ab} + q_{ba})$. Die Wahrscheinlichkeit im Umfeld dieser Organisationsform beschäftigt zu sein ergibt sich als $P(b|X) = q_{ab} / (q_{ab} + q_{ba})$. In analoger Weise ergeben sich die entsprechenden Wahrscheinlichkeiten für Organisationsform *Y*.

Die Effektivität einer Organisationsform lässt sich in erster Linie an der Wahrscheinlichkeit ablesen, mit der der Übergang aus Arbeitslosigkeit in Beschäftigung erfolgt

(q_{ab} bzw. r_{ab}). Bestandswahrscheinlichkeiten bzw. Bestandsquoten sind in dieser Hinsicht nicht eindeutig. Die Untersuchung müsste folglich darauf ausgerichtet sein, q_{ab} und r_{ab} zu messen.

Effektivitätsvergleiche auf der Basis von Bestandsquoten sind dagegen nur unter einer sehr restriktiven Bedingung zulässig: Es muss gewährleistet sein, dass der Maßnahmeneffekt homogen ist, d.h. dass die Wirkung der Organisationsform auf alle Bedürftigen gleich ist. Das würde beispielsweise bedeuten, dass der Unterschied zwischen Organisationsform *X* und Organisationsform *Y* im Hinblick auf die Vermittlung nicht davon abhängt, ob es sich bei den zu vermittelnden Personen um hoch Qualifizierte oder um gering Qualifizierte handelt. Diese Annahme lässt sich zudem nur für die Merkmale überprüfen, die beobachtet wurden.

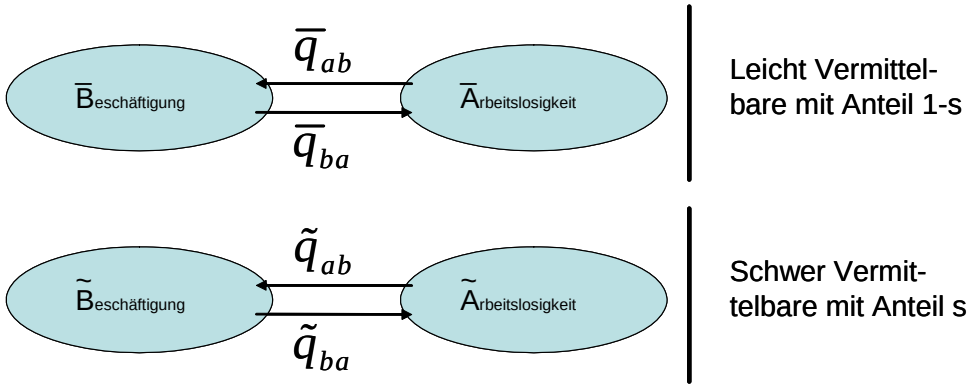
Um die Wirkung der Organisationsform anhand der bedingten Übergangswahrscheinlichkeiten q_{ab} und r_{ab} zu messen sind Verlaufsdaten auf der Grundlage von Zugangskohorten in die Bedürftigkeit erforderlich. Zugangskohorten erlauben eine direkte Schätzung der gesuchten Übergangswahrscheinlichkeiten unter Berücksichtigung von möglicherweise vorhandener unbeobachteter Heterogenität.

Paradoxe Verzerrungen durch Bestandsvergleiche

Werden dagegen – wie im zur Diskussion stehenden Untersuchungdesign – Bestandsvergleiche durchgeführt, können bei heterogenen Maßnahmeeffekten Verzerrungen auftreten, die zu paradoxen Schlussfolgerungen führen. Dies soll im Folgenden anhand eines einfachen Simulationsbeispiels erläutert werden.

Dazu wird angenommen, dass sich die Population der Hilfebedürftigen aus schwer Vermittelbaren und leicht Vermittelbaren zusammensetzt. Der Anteil der schwer Vermittelbaren sei mit *s* bezeichnet. Abb. 2 enthält eine schematische Darstellung dieses Falls am Beispiel von Organisationsform *X*. Statt der Übergangswahrscheinlichkeiten q_{ab} und q_{ba} sind nun die gruppenspezifischen Übergangswahrscheinlichkeiten $\overline{q}_{ab}$ und $\overline{q}_{ba}$ für die leicht Vermittelbaren sowie $\widetilde{q}_{ab}$ und $\widetilde{q}_{ba}$ für die schwer Vermittelbaren zu berücksichtigen.

Abb. 2: Arbeitsmarktübergänge bei heterogenen Teilpopulationen am Beispiel von Organisationsform *X*



Des Weiteren wird angenommen, dass beide Organisationsformen im Hinblick auf schwer Vermittelbare gleich effektiv sind. Organisationsform *Y* sei dagegen im Hinblick auf leicht Vermittelbare effektiver. Nach dem Pareto-Prinzip ist Organisationsform *Y* damit annahmegemäß insgesamt effektiver bei der Überwindung der Hilfebedürftigkeit.

Die konkret angenommenen Simulationsparameter sind Tab. 1 zu entnehmen. Beide Organisationsformen unterscheiden sich nur im Hinblick auf die Übergangswahrscheinlichkeit von leicht Vermittelbaren aus Arbeitslosigkeit in Beschäftigung. Organisationsform *Y* ist hier annahmegemäß doppelt so gut wie Organisationsform *X* ($\overline{r}_{ab}=2\overline{q}_{ab}$). Der Anteil der schwer Vermittelbaren in der Bevölkerung ist in beiden Organisationsformen der Gleiche, d.h. Unterschiede im Ergebnis sind ausschließlich auf Unterschiede in der Performance zurückzuführen.

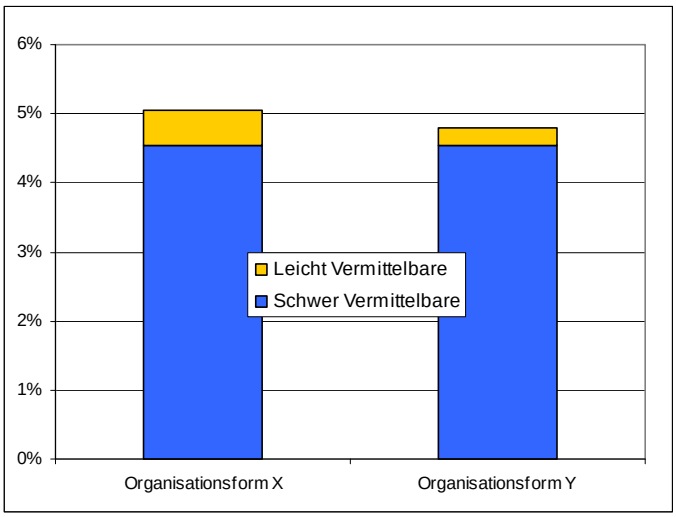
Tab. 1: Angenommene Simulationsparameter

Organisationsform *X*		Organisationsform *Y*	
$\overline{q}_{ab}$	0,100	$\overline{r}_{ab}$	0,200
$\overline{q}_{ba}$	0,001	$\overline{r}_{ba}$	0,001
$\tilde{q}_{ab}$	0,020	$\tilde{r}_{ab}$	0,020
$\tilde{q}_{ba}$	0,002	$\tilde{r}_{ba}$	0,002
Anteil der schwer Vermittelbaren (*s*): 0,500			

Für schwer Vermittelbare ergibt sich aus der gewählten Parameterkonstellation sowohl im Umfeld von Organisationsform *X* als auch im Umfeld der Organisations-

form *Y* eine Arbeitslosenquote von gut 9%. Für leicht Vermittelbare beträgt die entsprechende Arbeitslosenquote dagegen knapp 1% in Organisationsform *X* und knapp 0,5% in Organisationsform *Y*. Im Bevölkerungsdurchschnitt ergibt sich für beide Organisationsformen eine Arbeitslosenquote von etwa 5% mit einem leichten Vorteil zugunsten von Organisationsform *Y* (vgl. Abb. 3).

Abb. 3: Simulierte Arbeitslosenquoten im Systemgleichgewicht



Quelle: Berechnungen auf der Grundlage der Annahmen in Tab. 1

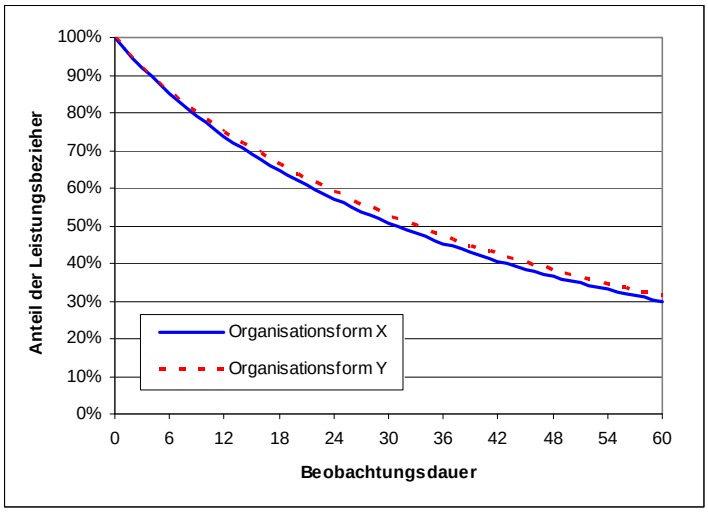
Abb. 3 macht deutlich, dass schwer Vermittelbare im Bestand der Arbeitslosen – gemessen an ihrem Bevölkerungsanteil – deutlich überrepräsentiert sind. Im Bestand von Organisationsform *X* beträgt er 90%, im Bestand von Organisationsform *Y* sogar knapp 95%. Für dieses Phänomen ist ihr längerer Verbleib in der Arbeitslosigkeit verantwortlich. Man spricht daher auch von der Dauerverzerrung von Beständen.

Bemerkenswert ist Folgendes: Obwohl die Arbeitslosenquote in Organisationsform *Y* geringer ist als in Organisationsform *X*, ist der Anteil der schwer Vermittelbaren – und damit der Anteil der Langzeitbedürftigen – unter den Arbeitslosen in Organisationsform *Y* höher als in Organisationsform *X*. In diesem scheinbaren Widerspruch liegt auch die Quelle für mögliche Fehlinterpretationen. Die annahmegemäß bessere Performance von Organisationsform *Y* bei den leicht Vermittelbaren führt dazu, dass dort der Anteil dieser Gruppe relativ gering ist. Die Kennziffern für den verbleibenden Bestand werden von den schwer Vermittelbaren dominiert. So ist in dieser Konstellation beispielsweise auch die durchschnittliche Dauer des Hilfebezugs im Bestand der Arbeitslosen in Organisationsform *Y* höher als in Organisationsform *X*, und zwar gerade weil in Organisationsform *Y* eine effektivere Vermittlung betrieben wird.

Das in der BMAS-Studie für die Beurteilung der Effektivität herangezogene Untersuchungsdesign besteht im Prinzip darin, zu einem Zeitpunkt t_0 eine Bestandsstichprobe von Hilfebedürftigen zu ziehen und zu einem späteren Zeitpunkt t_1 den Status der Hilfebedürftigkeit erneut zu messen. Der Anteil der verbleibenden Hilfebedürftigen wird dann als zentrales Beurteilungskriterium für die Effektivität der Vermittlungsarbeit im Zuständigkeitsbereich der jeweiligen Organisationsform herangezogen. Unter den getroffenen Annahmen (vgl. Tab. 1) führt ein solches Evaluationsdesign in vielen Fällen paradoxerweise dazu, Organisationsform X besser aussehen zu lassen als Organisationsform *Y*, obwohl Organisationsform *Y* objektiv eine bessere Performance aufweist.

Abb. 4 zeigt für beide Organisationsformen die Entwicklung des Anteils der verbleibenden Hilfebedürftigen in Abhängigkeit vom zeitlichen Abstand zwischen Bestandserhebung und Zweitmessung (Beobachtungsdauer). Auch wenn der Unterschied hier nicht dramatisch ausfällt, ist doch zu erkennen, dass der Anteil der Hilfebedürftigen im Laufe der Zeit in Organisationsform *Y* weniger stark abgebaut wird als in Organisationsform *X*. Die Ursache für diesen scheinbaren Widerspruch liegt in der unterschiedlichen Ausgangslage, die im gewählten Evaluationsdesign definitionsgemäß außer Betracht bleibt: In Organisationsform *Y* liegt im Ausgangszustand annahmegemäß ein höherer Anteil von schwer Vermittelbaren vor, gerade weil sie im Hinblick auf leicht Vermittelbare effektiver sind. Daher ist sie Organisationsform *X* im Zeitvergleich scheinbar unterlegen, weil letztere davon profitiert, dass dort im Ausgangszustand ein höherer Anteil von leicht Vermittelbaren anzutreffen ist, der schneller abgebaut wird als der Anteil der schwer Vermittelbaren. Nur wenn in beiden Organisationsformen die gleiche Ausgangsverteilung von schwer und leicht Vermittelbaren anzutreffen wäre, würde sich die Überlegenheit von Organisationsform *Y* in den angestellten Zeitvergleich eindeutig zeigen. Da die Ausgangsbestände aber ihrerseits das Ergebnis von Performance-Unterschieden sind, ist die Voraussetzung von gleichen Ausgangsverteilungen im Regelfall nicht gegeben.

Abb. 4: Verbleib im Leistungsbezug bezogen auf den Ausgangsbestand



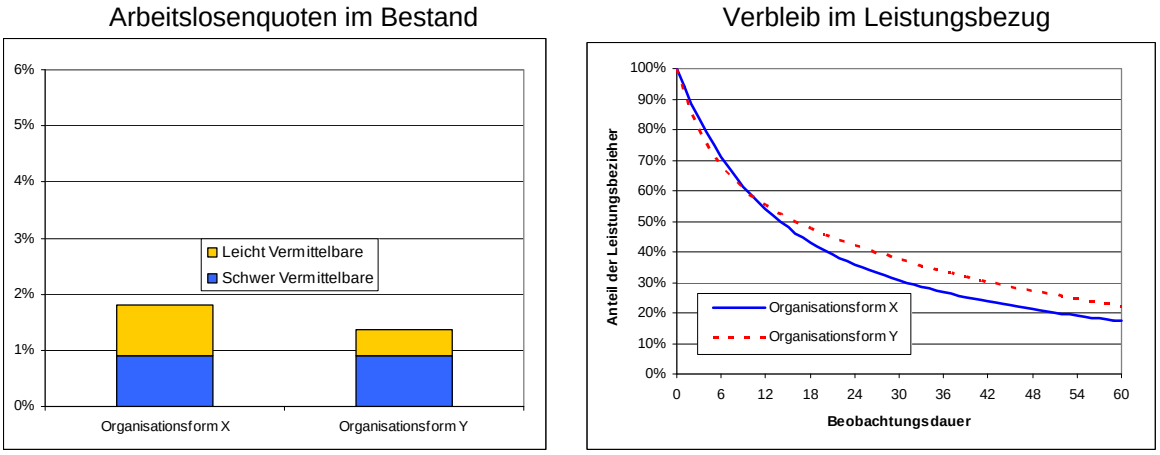
Quelle: Berechnungen auf der Grundlage der Annahmen in Tab. 1

Die Widersprüche zwischen Messung und Realität treten umso deutlicher zum Vorschein, je größer der mehr oder weniger willkürlich gewählte zeitliche Abstand zwischen Bestandserhebung und Zweitmessung ist. Darüber hinaus wird das Ergebnis davon beeinflusst, wie hoch der Anteil der schwer Vermittelbaren in der Bevölkerung ist. Um Letzteres zu illustrieren sind in Abb. 5 die Ergebnisse für eine entsprechend modifizierte Situation dargestellt. Die Simulationsparameter unterscheiden sich von denen für Abb. 4 nur dadurch, dass der Anteil der schwer Vermittelbaren in der Bevölkerung mit 10% statt wie zuvor mit 50% veranschlagt wird.

Hier wird zunächst besonders deutlich, dass es einen Unterschied macht, wann die Zweitmessung erfolgt. Nach vier Monaten würde man beispielsweise noch einen leichten Vorteil von Organisationsform *Y* feststellen, der jedoch spätestens nach 18 Monaten in einen gravierenden Nachteil umschlägt. Auf den ersten Blick mag es überraschen, dass die Paradoxie zwischen Messung und Realität gerade dann so hervortritt, wenn in der vorliegenden Konstellation der Anteil der schwer Vermittelbaren eher gering ist. Der Grund dafür ist Folgender: Wenn der Bevölkerungsanteil der schwer Vermittelbaren geringer ist, ist *c.p.* auch ihr Anteil unter den Arbeitslosen geringer. Im vorliegenden Fall beträgt er in Organisationsform *X* 50% und in Organisationsform *Y* 67%. Entsprechend höher ist der Anteil der leicht Vermittelbaren. Zunächst schlägt deshalb bei Organisationsform *Y* noch der angenommene Performance-Vorteil bei den leicht Vermittelbaren zu Buche, solange deren statistische Masse noch hinreichend groß ist. Ist dieser Vorteil aufgezehrt, tritt das bereits bekannte Phänomen in Aktion: Der höhere Ausgangsbestand an leicht Vermittel-

baren lässt Organisationsform *X* auf Dauer besser aussehen, weil sich der Abbau der Hilfebedürftigkeit in erster Linie bei den leicht Vermittelbaren abspielt.

Abb. 5: Ausgangsbestand im Gleichgewicht und Verbleib im Leistungsbezug bezogen auf den Ausgangsbestand bei geringerem Anteil von schwer Vermittelbaren in der Bevölkerung

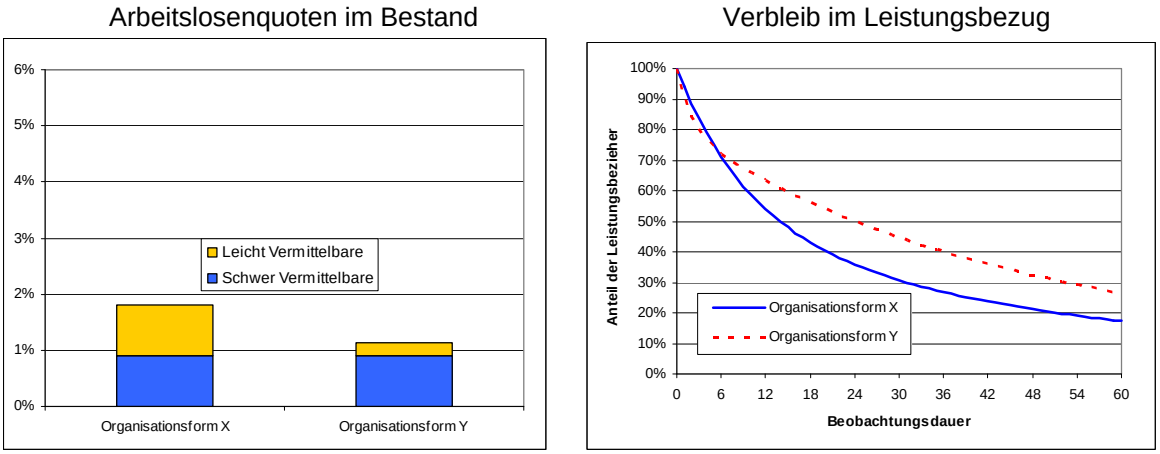


Quelle: Berechnungen auf der Grundlage der Annahmen in Tab. 1; für den Parameter s wird jedoch statt eines Wertes von 50% ein Wert von 10% angenommen

Vollends paradox wird es, sollte man in Organisationsform *Y* auf die Idee kommen, diesem Effekt durch Steigerung der Effektivität zu begegnen. Das Ergebnis gleicht bildlich gesprochen dem Wettlauf zwischen Hase und Igel: Je größer der Effektivitätsunterschied bei der Vermittlung von leicht Vermittelbaren gegenüber Organisationsform *X*, desto schlechter sieht Organisationsform *Y* am Ende aus.

Abb. 6 stellt dazu beispielhaft eine Situation dar, bei der es in Organisationsform *Y* gelingt, eine viermal höhere Übergangswahrscheinlichkeit aus Arbeitslosigkeit in Beschäftigung zu realisieren als in Organisationsform *X*. Die übrigen Parameter entsprechen denen aus Abb. 5. Im Ergebnis sieht es so aus als ob die Verbleibswahrscheinlichkeit im Hilfebezug in Organisationsform *Y* um bis zu 15 Prozentpunkte höher ist als in Organisationsform *X*.

Abb. 6: **Ausgangsbestand im Gleichgewicht und Verbleib im Leistungsbezug bezogen auf den Ausgangsbestand bei deutlich besserer Performance von Organisationsform *Y***



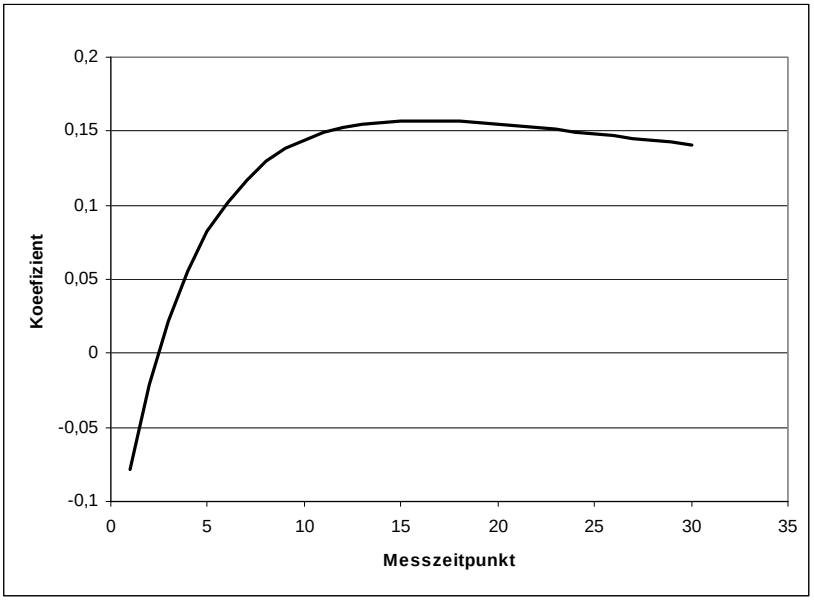
Quelle: Berechnungen auf der Grundlage der Annahmen in Abb. 5; für den Parameter $\bar{r}_{ab}$ wird jedoch statt eines Wertes von 20% ein Wert von 40% angenommen

Solange sich schwer Vermittelbare und leicht Vermittelbare anhand beobachtbarer Kriterien unterscheiden lassen, kann eine entsprechend differenzierte Analyse vor möglichen Fehleinschätzungen schützen. Wendet man das beschriebene Untersuchungsdesign getrennt auf die beiden Subpopulationen von schwer und leicht Vermittelbaren an, treten die beschriebenen Probleme nicht auf. Es ergibt sich korrekterweise, dass der Anteil der schwer Vermittelbaren in beiden Organisationsformen gleichmäßig zurückgeht. Ebenso korrekt tritt zutage, dass der Anteil der leicht Vermittelbaren in Organisationsform *Y* relativ stärker zurückgeht als in Organisationsform *X*. Ist eine getrennte Betrachtung nach Subpopulationen jedoch nicht möglich und/oder wird sie – wie im zur Debatte stehenden Untersuchungsdesign – nicht berücksichtigt, kann es zu schwerwiegenden Fehleinschätzungen kommen.

Anfällig dafür sind insbesondere die üblichen nicht-parametrischen Verfahren zur Ermittlung des sogenannten Average-Treatment-Effects wie sie in der BMAS-Studie zur Anwendung kamen. Gleiches gilt aber auch für parametrische Verfahren, selbst wenn die relevanten Merkmale berücksichtigt werden. Um letzteres zu illustrieren, werden im Folgenden auf der Basis der Parameterkonstellation aus Abb. 5 die Schätzungen für ein Logit-Modell für den Verbleib in der Arbeitslosigkeit präsentiert, bei dem zunächst nicht zwischen leicht und schwer Vermittelbaren unterschieden wird. Die Schätzungen wurden für verschiedene Messzeitpunkte durchgeführt. Die Arbeitslosigkeitswahrscheinlichkeit von leicht Vermittelbaren im Zuständigkeitsbereich von Organisationsform *X* dient jeweils als Referenzkategorie. Ein positiver

Koeffizient besagt, dass die Arbeitslosigkeitswahrscheinlichkeit in Organisationsform *Y* höher ist als in Organisationsform *X*. Ein negativer Koeffizient indiziert entsprechend den umgekehrten Fall.

Abb. 6: Geschätzter Y-Koeffizient für die Wahrscheinlichkeit des Verbleibs in Arbeitslosigkeit in Abhängigkeit vom Messzeitpunkt



Quelle: Berechnungen auf der Grundlage von Logit-Schätzungen auf der Grundlage der Annahmen in Abb. 5.

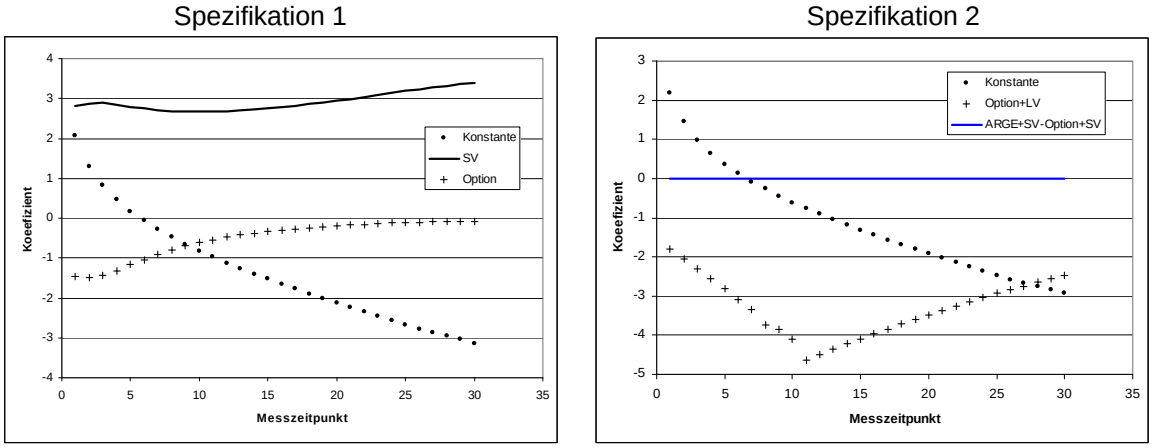
Die geschätzten Parameter sind anfänglich negativ, was bedeutet, dass die Arbeitslosigkeitswahrscheinlichkeit in Organisationsform *Y* niedriger ist als in Organisationsform *X*. Bereits ab der dritten Messperiode schlägt das Vorzeichen jedoch um, was zu der falschen Interpretation führen würde, dass Organisationsform *Y* bei der Vermittlung weniger effizient arbeitet als Organisationsform *X*.

Das dargestellte Problem lässt sich leider auch nicht in befriedigender Weise dadurch lösen, dass im Rahmen von multivariaten Modellen für den Verbleib im Hilfebezug explizit zwischen leicht Vermittelbaren und schwer Vermittelbaren differenziert wird. In diesem Fall kommt es entscheidend auf die Modellspezifikation an, aber selbst im günstigsten Fall sind die Parameter nur qualitativ, nicht aber kausal quantitativ interpretierbar.

Um dies zu illustrieren werden im Folgenden die Ergebnisse von weiteren simulierten Schätzungen präsentiert. In der ersten Spezifikation wurden zwei Haupteffekte berücksichtigt. Der eine (*Y*) bildet die Zugehörigkeit einer Person zu Organisationsform *Y*, der andere indiziert, ob die betreffende Person schwer vermittelbar ist (*SV*).

Wie zuvor bilden leicht Vermittelbare im Zuständigkeitsbereich von Organisationsform *X* die Referenzkategorie. Der negative Koeffizient für Organisationsform *Y* in Spezifikation 1 zeigt für frühe Messzeitpunkte immerhin korrekterweise deren Vorteil im Hinblick auf den Vermittlungserfolg an. Der Effekt verschwindet jedoch, je später die Messung erfolgt. Die Interpretation hängt somit auch hier davon ab, zu welchem Zeitpunkt die Messung erfolgt. In der Alternativspezifikation 2 wurden die Effekte für schwer Vermittelbare in Organisationsform *X*, für leicht Vermittelbare in Organisationsform *Y* und für schwer Vermittelbare in Organisationsform *Y* explizit modelliert. Immerhin ergibt sich hier korrekt und unabhängig vom Messzeitpunkt, dass sich die Arbeitslosigkeitswahrscheinlichkeit von schwer Vermittelbaren in beiden Organisationsformen nicht unterscheidet. Die Differenz beider Effekte ergibt exakt Null. Der Effekt für leicht Vermittelbare in Organisationsform *Y* ist ebenfalls korrekt negativ. Gleichwohl schwankt die Größenordnung erheblich in Abhängigkeit vom Messzeitpunkt, was wiederum die Interpretierbarkeit beeinträchtigt.

Abb. 7: Geschätzte Koeffizienten für die Wahrscheinlichkeit des Verbleibs in Arbeitslosigkeit in Abhängigkeit vom Messzeitpunkt



Quelle: Berechnungen auf der Grundlage von Logit-Schätzungen unter den Annahmen in Abb. 5. X steht für Zugehörigkeit zu Organisationsform *X*, Y steht für Zugehörigkeit zu Organisationsform *Y*, SV für schwer Vermittelbare und LV für leicht Vermittelbare.

Fazit

Das keineswegs unrealistische Beispiel einer heterogenen Maßnahmewirkung zeigt, dass das in der Wirkungsanalyse des BMAS verwendete Evaluationsdesign keine eindeutigen Schlussfolgerungen im Hinblick auf die eigentliche Fragestellung zulässt. Das verwendete Evaluationsdesign kann dazu führen, dass die effektivere Organisationsform scheinbar schlechter abschneidet. Die Frage „Wer kann es besser?“ muss damit nach wie vor als unbeantwortet gelten.

Es kann zwar sein, dass die ARGEn tatsächlich die bessere Arbeit leisten, aber die vorliegende BMAS-Studie liefert dafür keine hinreichende Rechtfertigung. Damit wird keineswegs behauptet, die Optionskommunen leisteten in Wahrheit die bessere Arbeit. Die Simulationsbeispielen sind zwar bewusst in diesem Sinne so gewählt, haben aber lediglich illustrierenden Charakter.

Festzustellen ist, dass das in der BMAS-Studie verwendete Evaluationsdesign beim Vorliegen heterogener Maßnahmeneffekte paradoxe Resultate hervorrufen kann, die umso ausgeprägter sind

- je effektiver die überlegene Organisationsform arbeitet,
- je größer der Bevölkerungsanteil der Gruppe ist, bei dem sich die Effektivitätsunterschiede bemerkbar machen und
- je später die Messung erfolgt.

Im konkreten Fall liegt die entscheidende Ursache für das Problem darin, dass relative Änderungen im Verhältnis zu einem Ausgangsbestand betrachtet werden, ohne die mögliche Dauerverzerrung des Bestandes zu berücksichtigen. Eine sinnvolle Strategie zur Vermeidung dieses Problems hätte darin bestanden, Zugangskohorten in den Hilfebezug zu erheben statt einer Bestandsstichprobe. Auf der Grundlage von Zugangskohorten lassen sich bedingte Übergangswahrscheinlichkeiten schätzen, die zur Beschreibung von Wirkungsunterschieden wesentlich geeigneter sind als Bestandswahrscheinlichkeiten und die gegenüber Verzerrungen der beschriebenen Art robust sind. Allerdings ist dabei zu berücksichtigen, dass die Analyse auf der Grundlage von Zugangskohorten einen längeren Beobachtungszeitraum erfordert hätte als er seitens der Politik zugestanden worden war. Dies zeigt, in welchem Dilemma evidenzbasierte Politikberatung häufig steht. Rasche Ergebnisverfügbarkeit und belastbares Untersuchungsdesign schließen sich nicht selten aus. Dies im Einzelfall zu erkennen und zu vermitteln wird auf absehbare Zeit eine Herausforderung für beide Seiten bleiben.