

Evaluationsergebnisse und die Wahl statistischer Zwillinge

Nicht Äpfel mit Birnen vergleichen



Aktive Arbeitsmarktpolitik zielt darauf ab, Arbeitslosigkeit und Hilfebedürftigkeit zu vermeiden oder zumindest deren Dauer zu verkürzen. Ob die hierfür aufgewendeten Mittel gut eingesetzt sind, darüber besteht in Politik und Fachöffentlichkeit wenig Einigkeit: Handelt es sich bei der aktiven Arbeitsförderung – wie mitunter behauptet – um ein „Milliardengrab“? Oder vielmehr um ein effektives Instrumentarium zur Wiedereingliederung Arbeitsloser? Die Antwort kann auch vom Design der Evaluationsstudie abhängen – wie eine aktuelle IAB-Studie zeigt.

Auf theoretischer Ebene lassen sich die Effekte von Maßnahmen nicht eindeutig identifizieren. Dank beträchtlicher Fortschritte bei den Datengrundlagen und Untersuchungsmethoden gehört die Evaluation der aktiven Arbeitsmarktpolitik in Deutschland mittlerweile zum Kernbereich der empirischen Arbeitsmarktforschung. Dies hat die Diskussion versachlicht. Dennoch sind die Aussagen wissenschaftlicher Studien nicht immer eindeutig. Ein Beispiel dafür sind verschiedene Untersuchungen zu den Effekten beruflicher Weiterbildungsmaßnahmen in den Jahren 2000 bis 2002 auf die späteren Beschäftigungschancen geförderter Arbeitsloser. Während manche Studien hier insignifikante Effekte feststellen, ermitteln andere signifikant positive Förderwirkungen für alle untersuchten Teilgruppen. Auf den ersten Blick scheint es, als ob sich diese Studien

sogar auf denselben Untersuchungsansatz stützen: Sie schätzen ab, was aus den Teilnehmern einer Maßnahme ohne Förderung geworden wäre. Zu diesem Zweck wird eine Vergleichsgruppe von Nicht-Teilnehmern gebildet, die möglichst gleiche Eigenschaften wie die Teilnehmer aufweisen – sogenannte „statistische Zwillinge“. Das Verfahren wird als statistisches Matching bezeichnet. Ein Zwilling muss dabei nicht eine Person sein, sondern kann auch als „gewichtetes Mittel“ mehrerer Personen gebildet werden.

Eine Sensitivitätsanalyse des IAB verdeutlicht, dass die Definition der Nicht-Teilnehmer an einer Maßnahme ein möglicher Grund für die unterschiedliche Einschätzungen der Förderwirkungen sein kann (vgl. Kasten „Die IAB-Untersuchung zum Evaluationsdesign“). Entscheidend ist, wie die Nicht-Teilnehmer definiert werden. Entweder handelt es sich um Personen, die im genannten Untersuchungszeitraum niemals an einem arbeitsmarktpolitischen Programm teilgenommen haben, oder um Personen, die nach dem Förderbeginn der Teilnehmer ebenfalls in eine arbeitsmarktpolitische Maßnahme eintreten.

Was ist ein „Nicht-Teilnehmer“?

Häufig wählen Evaluationsstudien aufgrund beobach-

teter Merkmale wie Bildung und Erwerbsbiografie eine Vergleichsgruppe „statistischer Zwillinge“ aus den Nicht-Teilnehmern aus; die Ergebnisse der Vergleichsgruppe dienen dann als Schätzung für den Effekt, der ohne Teilnahme an der untersuchten Maßnahme zu erwarten gewesen wäre. Aber wer ist ein Nicht-Teilnehmer? Die Annahme, ein Nicht-Teilnehmer sei eine Person, die während ihrer Arbeitslosigkeit niemals an einem Förderprogramm teilgenommen hat, führt in die Irre – vernachlässigt sie doch die Dynamik des Förderzugangs und die sich daraus ergebenden Folgen für die Zusammensetzung der Vergleichsgruppe.

In Deutschland wie in anderen europäischen Ländern beginnen Förderprogramme nicht zu einem bestimmten Zeitpunkt. Vielmehr werden fortlaufend neue Personen aufgenommen. Überspitzt gesagt: Jede Person, die heute nicht in eine Förderung eintritt, ist bereits morgen ein potenzieller Teilnehmer. Personen, die nie mit einem arbeitsmarktpolitischen Instrument gefördert wurden, dürften damit eine „Positiv-Auswahl“ von Arbeitslosen sein: Sie haben – nach ihrer eigenen Auffassung wie nach Einschätzung der Arbeitsvermittler – auch ohne Förderung überdurchschnittlich gute Chancen, ihre Arbeitslosigkeit zu beenden und eine Beschäftigung zu finden. Das kann etwa daran liegen, dass sie

Die IAB-Untersuchung zum Evaluationsdesign

In der Studie werden die Ergebnisse zur Effektivität ausgewählter Maßnahmen für drei unterschiedliche Untersuchungsdesigns verglichen. Zugrunde liegt eine ausgewählte Gruppe von Personen, die im März 2003 bis zu ein Jahr arbeitslos gemeldet waren, in dieser Zeit noch an keiner Maßnahme teilgenommen hatten und zu Beginn ihrer Arbeitslosigkeit 25 bis 59 Jahre alt waren. Geschätzt werden die Fördereffekte für die Personen, die im März 2003 in verschiedene Maßnahmen kurzer oder mittlerer Dauer (bis zu sechs Monaten) eintraten. Untersucht wurden Personen, die mit einer Arbeitsbeschaffungsmaßnahme, einem

Einarbeitungszuschuss bei erschwelter Vermittlung, einer berufsbezogenen Weiterbildung oder aber einer nicht-betrieblichen bzw. betrieblichen Trainingsmaßnahme zur Kenntnisvermittlung gefördert wurden. Der interessierende Effekt ist die Zahl der kumulierten Tage, die die Geförderten sowie ihre ausgewählten „statistischen Zwillinge“ in den dreieinhalb Jahren nach Förderbeginn in regulärer Beschäftigung verbracht haben. Der Auswahl der Vergleichsgruppen liegen prozessdatenbasierte Informationen über soziodemografische und regionale Merkmale und über die Arbeitsmarkthistorie der Arbeitslosen zugrunde.

besonders motiviert sind oder besondere Talente haben, die in den verfügbaren Daten nicht erfasst sind. Ökonomisch gesprochen tritt das Selektivitätsproblem auf, dass – trotz ähnlicher beobachtbarer Merkmale – nur die besseren „Risiken“ in die Vergleichsgruppe sortiert werden. Vergleicht man eine solche Gruppe mit der Teilnehmergruppe, so wird die tatsächliche Förderwirkung vermutlich unterschätzt.

Empirische Lösungen des Vergleichsgruppenproblems

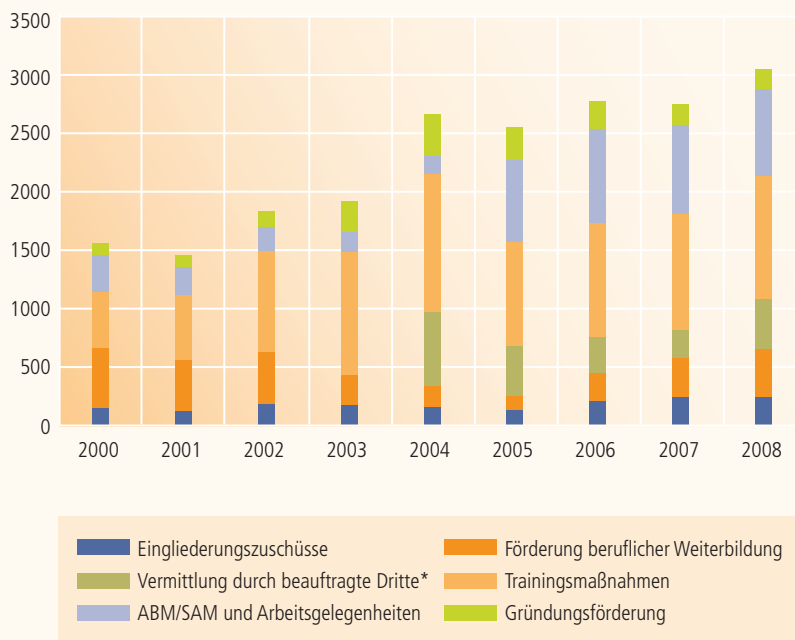
In der europäischen Literatur werden Nicht-Teilnehmer daher meist als Personen definiert, die zwar für den Zeitraum, in dem die Geförderten an einer Maßnahme teilnehmen, nicht gefördert wurden – aber durchaus auf eine eventuelle spätere Förderung „warten“ können. Damit wird letztlich die Wirkung eines Maßnahmein-

tritts innerhalb eines begrenzten Zeitraums auf die späteren Arbeitsmarktchancen untersucht.

Eine erste Möglichkeit, das Vergleichsgruppenproblem empirisch zu lösen, besteht darin, nur Maßnahmezugänge aus einem sehr kurzen Zugangszeitraum, etwa einem Kalendermonat, auszuwerten, und als potenzielle Vergleichsgruppenmitglieder alle Arbeitslosen heranzuziehen, die innerhalb dieses Zeitraums nicht in eine Maßnahme eingetreten sind. Diesen Ansatz wählt die IAB-Studie. Je kleiner allerdings die Stichprobe der untersuchten Zugänge ist, desto länger muss der Zugangszeitraum gewählt werden, um überhaupt eine hinreichend große Untersuchungsgruppe zu erhalten. So finden sich in der Evaluationsliteratur für Deutschland durchaus Zugangszeiträume von bis zu 24 Monaten – mit entsprechenden Konsequenzen für die Selektionsprozesse.

Abbildung 1

Zugänge in wichtige arbeitsmarktpolitische Maßnahmen 2000 bis 2008 in Tausend



Anmerkung: Vorläufige Zahlen für 2008.

*Zahlen erst seit 2004 verfügbar.

Quelle: Statistik der Bundesagentur für Arbeit

©IAB

Zweitens erfolgt zur Vermeidung des Selektionsproblems häufig eine Schichtung nach Dauer der Arbeitslosigkeit. Für eine Gruppe von Arbeitslosen, die innerhalb von vier bis sechs Monaten nach Beginn der Arbeitslosigkeit in eine Maßnahme eingetreten sind, wären dann alle Arbeitslosen, die mindestens vier bis sechs Monate arbeitslos waren, ohne in eine Maßnahme einzutreten, eine potenzielle Vergleichsgruppe. Letztere können aber durchaus im siebten Monat nach Beginn der Arbeitslosigkeit an einem Förderprogramm teilnehmen. Die Kehrseite beider Verfahren ist, dass sie zwar das oben genannte Selektivitätsproblem vermeiden, die geschätzten Fördereffekte jedoch nur eingeschränkt aussagekräftig sind: Geschätzt wird nicht der Effekt einer Förderung im Vergleich zu einer Nicht-Förderung, sondern der Effekt eines Förderzugangs zu einem bestimmten Zeitpunkt im Vergleich zu einer Nicht-Förderung zu diesem Zeitpunkt (aber eventuell zu einem späteren Zeitpunkt). Und auch die Teilnehmer können später noch an weiteren Maßnahmen teilnehmen.

Drei Vergleichsgruppenkonzepte

Zu den wichtigsten arbeitsmarktpolitischen Instrumenten zählen die Förderung beruflicher Weiterbildung, Trainingsmaßnahmen, Eingliederungszuschüsse, die Gründungsförderung und die Arbeitsgelegenheiten (vgl. Abbildung 1 auf Seite 88). In der IAB-Untersuchung werden Arbeitslose, die im März 2003 in ausgewählte Maßnahmen eintraten, in einem ersten Konzept mit ähnlichen Arbeitslosen verglichen, die im März 2003 nicht in eine Maßnahme eintraten, aber teils auf einen eventuellen späteren Maßnahmeneintritt „warteten“ (vgl. Abbildung 2). Hierdurch lässt sich das oben genannte Selektivitätsproblem bei der Auswahl der „statistischen Zwillinge“ vermeiden.

Im zweiten Konzept werden als potenzielle Vergleichspersonen nur Arbeitslose herangezogen, die über den gesamten Betrachtungszeitraum nicht gefördert wurden. Damit wird geprüft, ob sich das theoretisch begründete Selektivitätsproblem tatsächlich in niedrigeren Schätzungen der Fördererfolge auswirkt.

Dies ist sicherlich ein Extremfall, der sich in der Literatur so nicht findet. Üblicherweise liegt der Zeitraum, in dem die Nicht-Teilnehmer an keiner Maßnahme teilgenommen haben, zwischen einem Monat und zwei Jahren. Dennoch lassen sich hiermit die prinzipiellen Unterschiede in den Ergebnissen gut herausarbeiten.

Abbildung 2

Drei Vergleichsgruppenkonzepte

1. Konzept	März 2003	April 2003 bis September 2006
Geförderte	Eintritt in Förderung	keine Restriktion
wartende Vergleichsgruppe	kein Eintritt in Förderung	keine Restriktion

2. Konzept	März 2003	April 2003 bis September 2006
Geförderte	Eintritt in Förderung	keine Restriktion
nie geförderte Vergleichsgruppe	kein Eintritt in Förderung	kein Fördereintritt

3. Konzept	März 2003	April 2003 bis September 2006
einmal geförderte	Eintritt in Förderung	kein weiterer Fördereintritt
nie geförderte Vergleichsgruppe	kein Eintritt in Förderung	kein Fördereintritt



Im dritten Konzept wird berücksichtigt, dass auch die Maßnahmenteilnehmer später in weitere Förderprogramme eintreten können – daher beschränkt sich der letzte Vergleich auf „einmal geförderte“ und „nie geförderte“ Personen. Hierdurch sollten die beiden verglichenen Gruppen ähnlicher sein als beim zweiten Konzept (und die Schätzergebnisse sollten wieder positiver ausfallen).

Maßnahmekarrieren sind keine Ausnahme

Das gewählte Konzept hat auf die Auswahl der potenziellen Vergleichspersonen und der Teilnehmer einen erheblichen Einfluss (vgl. Tabelle): Der Anteil der beim ersten Konzept ausgewählten „wartenden“ Vergleichspersonen, die später an einer Maßnahme teilnehmen, liegt bei 33 Prozent. Bei den untersuchten Maßnahmenteilnehmern beträgt der Anteil derjenigen, die später noch an einem weiteren Programm partizipieren, in der Regel sogar über 50 Prozent. Eine Ausnahme sind lediglich diejenigen Personen, die eine mit einem Eingliederungszuschuss geförderte Beschäftigung aufgenommen haben – hier nehmen „nur“ 31 Prozent aller potenziellen Vergleichsgruppen im Beobachtungszeitraum später noch mindestens einmal an einem weiteren Förderprogramm teil.

Für die Gesamtgruppe der Teilnehmer und der beim ersten Konzept ausgewählten Vergleichspersonen lässt sich noch einmal aufgliedern, welches die nächste Maßnahme war (vgl. Abbildung 3). Zwar machen die (kurzen) Trainingsmaßnahmen einen relativ hohen Anteil der weiteren Fördermaßnahmen aus. Dennoch sind auch andere „Maßnahmekarrieren“ nicht eben selten. Um so mehr erstaunt, dass hierzu für Deutschland bisher kaum wissenschaftliche Analysen vorliegen.

Förderwirkungen variieren

Wie die hohe Zahl an Maßnahmekarrieren bereits vermuten lässt, differieren die geschätzten Fördereffekte auf die Teilnehmer zwischen den drei Untersuchungskonzepten erheblich. Dies zeigt Abbildung 4 auf Seite 92. Je nach gewähltem Vergleichsgruppenkonzept differieren zum Beispiel für Maßnahmen der beruflichen Weiterbildung die geschätzten Wirkungen zwischen einem Beschäftigungsplus von (insignifikanten) 22 Tagen und (signifikanten) 138 Tagen. Die geschätzten Fördereffekte fallen beim zweiten Konzept – bei dem Geförderte mit nie geförderten Arbeitslosen verglichen werden – stets am geringsten aus. Immerhin: Bei den betrieblichen Maßnahmen – Eingliederungszuschüsse oder betriebliche

Tabelle

Anteil der in (weiteren) Maßnahmen geförderten Personen

	Geförderte	davon in Maßnahmen ab April 2003	
	Anzahl	Anzahl	in Prozent
Weiterbildung	1.414	701	50 %
Nicht-betriebliches Training	5.716	3.018	53 %
Betriebliches Training	3.135	1.633	52 %
Eingliederungszuschuss	734	228	31 %
Arbeitsbeschaffung	449	263	59 %
	Potenzielle Vergleichspersonen		
keine Maßnahme	144.406	48.253	33 %

Anmerkung: Berücksichtigt wurden Geförderte in Maßnahmen mit einer Dauer von bis zu sechs Monaten.

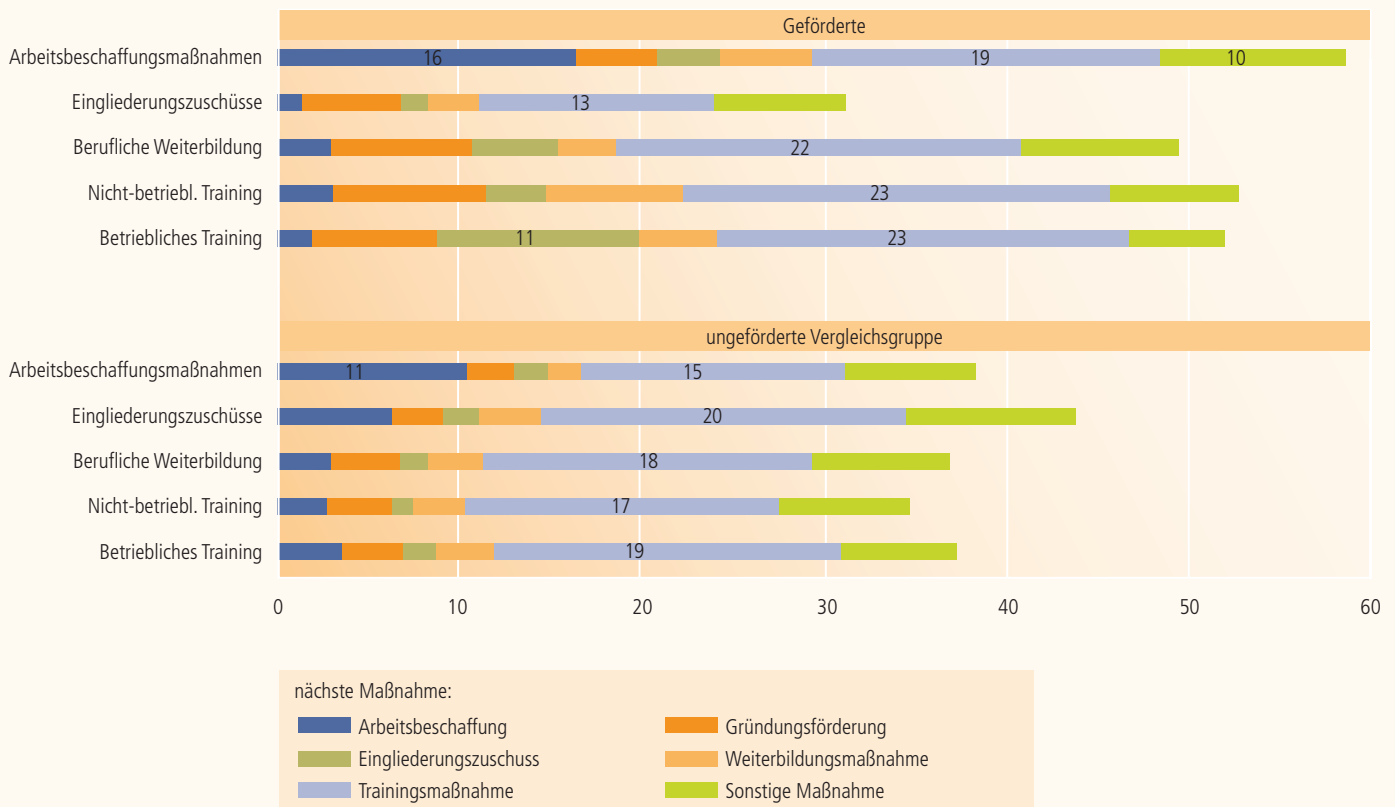
Quelle: Eigene Berechnungen

© IAB

Abbildung 3

Übergang von Geförderten und ihren statistischen Zwillingen in die nächste Maßnahme

in Prozent aller Geförderten bzw. der ungeförderten Vergleichsgruppe



Lesehilfe: Aus der Gruppe der Ungeförderten, die als Vergleichsgruppe zu den Teilnehmern einer betrieblichen Trainingsmaßnahme dient, absolvieren 19 Prozent später eine Weiterbildungsmaßnahme.
 Anmerkung: Erste Maßnahme 2003. Zahlenwerte nur bei Anteilen ab 10 Prozent.

Quelle: Eigene Berechnungen

©IAB

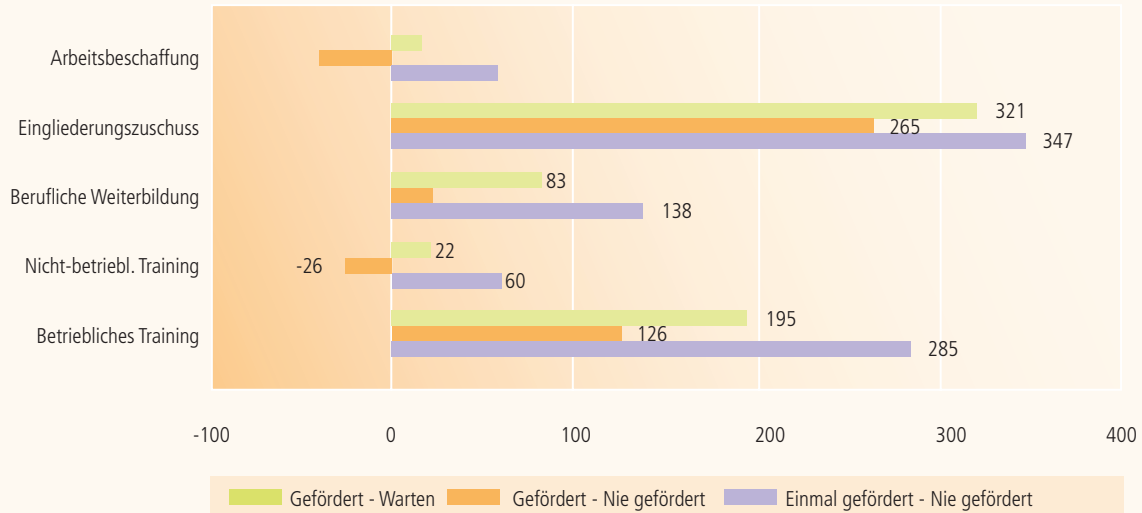
Trainingsmaßnahmen – sind die Geförderten nach dreieinhalb Jahren stets signifikant länger regulär beschäftigt gewesen als die ausgewählten Vergleichspersonen. Auch bei den Arbeitsbeschaffungsmaßnahmen ist der Unterschied zwischen der Teilnehmergruppe und der Vergleichsgruppe in keinem Fall signifikant – unabhängig vom gewählten Vergleichsgruppenkonzept lässt sich kein Effekt nachweisen. Anders bei Maßnahmen mit relativ schwachen Fördereffekten wie der berufsbezogenen Weiterbildung und den nicht-betrieblichen Trainingsmaßnahmen: Hier hängt es vom jeweiligen Vergleichsgruppenkonzept ab, ob die Geförderten signifikant länger oder kürzer in Beschäftigung waren als die Vergleichsgruppe.

Die Ergebnisse bestätigen die These, dass Personen, die nie an einer Maßnahme teilnehmen, tatsächlich eine „Positiv-Selektion“ von Arbeitslosen darstellen. Sie haben häufig keine Unterstützung durch eine Maßnahme benötigt, um wieder in Beschäftigung zu gelangen. Ein ähnliches Argument lässt sich für die Gruppe der Personen anbringen, die „nur“ einmal gefördert wurden – bei ihnen war offensichtlich keine zweite Förderung erforderlich. Ein zweites (triviales) Argument kommt hinzu: Maßnahmeteilnehmer müssen erst einmal den Einbindungseffekt (oder Lock-in-Effekt) der Förderung wieder aufholen: Solange die Förderung andauert, können sie nicht ungefördert beschäftigt sein.

Abbildung 4

Drei Vergleichskonzepte und ihre geschätzte Förderwirkung auf die Tage in regulärer Beschäftigung dreieinhalb Jahre nach Förderbeginn

in Tausend



Lesehilfe: Nach dem ersten Konzept „Gefördert - Warten“ waren Teilnehmer an der Maßnahme „Berufliche Weiterbildung“ in den dreieinhalb Jahren nach Förderbeginn 83 Tage mehr regulär beschäftigt als die Vergleichspersonen.

Anmerkung: Angabe Zahlenwert nur bei statistisch signifikantem ($\alpha = 0,05$) Fördereffekt. Berücksichtigt wurden Maßnahmen mit einer Dauer von bis zu sechs Monaten.

Quelle: Eigene Berechnungen

©IAB

Fazit

Wer an den Fördereffekten arbeitsmarktpolitischer Maßnahmen interessiert ist, kommt nicht umhin, sich mit Vergleichsgruppenkonzepten auseinanderzusetzen. Denn wie oben gezeigt wurde, können etwa bei berufsbezogener Weiterbildung und nicht-betrieblichen Trainingsmaßnahmen die Ergebnisse je nach gewähltem Konzept ganz unterschiedlich ausfallen.

Grundsätzlich gilt bei laufenden Zugängen in die Förderung: Die Methodik der „statistischen Zwillinge“ kann die Wirkung einer einzelnen Maßnahme im Vergleich zu einer Situation ganz ohne Förderung nicht zuverlässig ermitteln. Zwar scheint sich der Maßnahmeneffekt umso besser isolieren zu lassen, je länger potenzielle Vergleichspersonen in keiner Maßnahme gewesen sind. Umso größer ist aber die Gefahr, Äpfel und Birnen miteinander zu vergleichen. Denn wenn sich

die Vergleichsgruppe auf lange nicht geförderte Personen beschränkt, steigt die Gefahr, dass sich Teilnehmer und ausgewählte Vergleichspersonen systematisch in Bezug auf (für den Forscher) unbeobachtbare Merkmale unterscheiden. Evaluationsstudien für Deutschland verfolgen daher überwiegend das bescheidenere Ziel, die Wirkung eines Maßnahmeeintritts innerhalb eines begrenzten Zeitraums auf die folgende Erwerbsbiografie abzuschätzen – spätere Maßnahmeteilnahmen nicht ausgeschlossen.

Literatur

Gesine Stephan (2008), The Effects of Active Labor Market Programs in Germany – an Investigation Using Different Definitions of Non-treatment, *Jahrbücher für Nationalökonomie und Statistik* 228, 586-611.

Die Autorin



Prof. Dr. Gesine Stephan
ist Leiterin des Forschungs-
bereichs „Arbeitsförderung
und Erwerbstätigkeit“ am IAB.
gesine.stephan@iab.de

