

RESEARCH REPORT SERIES

IZA Research Report No. 76

**Machbarkeitsstudie zur Durchführung einer  
Evaluation der arbeitsmarktpolitischen  
Integrationsmaßnahmen für Flüchtlinge**

Forschungsbericht im Auftrag der  
Bundesrepublik Deutschland vertreten durch das  
Bundesministerium für Arbeit und Soziales

Holger Bonin

*IZA und Universität Kassel*

Ulf Rinne

*IZA*

APRIL 2017



# FORSCHUNGSBERICHT

481

## **Machbarkeitsstudie zur Durchführung einer Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge**

# ***Machbarkeitsstudie zur Durchführung einer Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge***

Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA)  
Schaumburg-Lippe-Straße 5-9  
53113 Bonn

Holger Bonin (IZA und Universität Kassel) und Ulf Rinne (IZA)

Bonn, März, 2017

## ***Kurzbeschreibung***

Die Studie klärt die Machbarkeit einer begleitenden Evaluation der arbeitsmarktpolitischen Maßnahmen zur Integration von Flüchtlingen im Zuständigkeitsbereich des Bundesministeriums für Arbeit und Soziales, einschließlich der berufsbezogenen Deutschsprachförderung. Es werden wissenschaftlich fundierte Vorschläge für die inhaltliche Konzeption, zielführende methodische Herangehensweisen und eine sinnvolle organisatorische Umsetzung eines Evaluationskonzepts entwickelt. Dazu werden der Stand der relevanten Forschung aufgearbeitet und bewertet, geeignete Erfolgsmaßstäbe für die Bewertung von Integrationsfortschritten entwickelt, geprüft, welche Methoden zur Offenlegung der Wirkungen von Integrationsmaßnahmen und ihres Zusammenspiels am besten geeignet sind, sowie Datenverfügbarkeit und Datenbedarfe ermittelt.

## ***Abstract***

The study clarifies the feasibility of an accompanying evaluation of labor market policy measures that are aimed at integrating refugees under the responsibility of the Federal Ministry of Labor and Social Affairs, including vocational German language training. It aims at developing scientifically sound proposals for a substantial conceptualization, targeted methodological approaches and a common-sense organizational implementation of an evaluation concept. For this purpose, current and relevant literature is reviewed and appropriate benchmarks for evaluating integration progress are developed. The study also examines which methods are best suited to disclose both direct effects and interactions of the integration measures. In addition, data availability and data requirements are determined.



# Inhalt

<b>Tabellenverzeichnis</b>	<b>7</b>
<b>Abbildungsverzeichnis</b>	<b>8</b>
<b>Abkürzungsverzeichnis</b>	<b>9</b>
<b>Zusammenfassung</b>	<b>11</b>
<b>1. Ausgangslage und Aufgabenstellung</b>	<b>15</b>
<b>2. Forschungsstand</b>	<b>19</b>
2.1 Nachteile von Flüchtlingen auf dem Arbeitsmarkt	20
2.2 Evaluationen integrationspolitischer Maßnahmen	21
<b>3. Vorhandene registrierte Daten</b>	<b>26</b>
3.1 Einleitung	26
3.2 Daten zu Personen im SGB II und III in der Statistik der BA	26
3.3 Geschäftsdaten des BAMF	30
3.4 Schnittstelle zwischen BAMF-Daten und Daten zu Personen im SGB II und III	30
3.5 Fazit	31
<b>4. Leistungen zur Evaluation</b>	<b>32</b>
<b>5. Zielgrößen und ihre Operationalisierung</b>	<b>40</b>
5.1 Einleitung	40
5.2 Arbeitsmarktintegration im eigentlichen Sinn	41
5.3 Weitere Bereiche ökonomisch-soziokultureller Integration	42
5.4 Operationalisierung und Erhebung nicht administrativ erfasster Zielgrößen	50
<b>6. Methodisches Vorgehen</b>	<b>55</b>
6.1 Das Fundamentalproblem der empirischen Wirkungsforschung	55

6.2	Methoden zur Gewinnung kausalanalytischer Aussagen	56
6.2.1	<i>Selektion anhand beobachtbarer Variablen</i>	56
6.2.2	<i>Analysen auf Basis quasi-experimenteller Variation</i>	58
6.2.3	<i>Evaluation von Maßnahmenketten</i>	60
6.3	Qualitativ orientierte Begleitforschung	61
<b>7.</b>	<b>Konzept zur Umsetzung der Evaluation</b>	<b>63</b>
7.1	Vorüberlegungen	63
7.2	Vorschlag zur praktischen Durchführung	66
7.2.1	<i>Modularisierung</i>	66
7.2.2	<i>Wirkungsanalytische Module</i>	67
7.2.3	<i>Befragungsmodul</i>	71
7.2.4	<i>Koordinierung durch eine Geschäftsstelle</i>	74
7.3	Zeitplan	76
	<b>Literaturverzeichnis</b>	<b>80</b>

## Tabellenverzeichnis

<i>Tabelle 5.1</i>	<i>Empirische Studien mit signifikantem Zusammenhang zwischen Integrationsindikator und Arbeitsmarktintegration von Flüchtlingen</i>	<i>47</i>
<i>Tabelle 5.2</i>	<i>Empirische Studien mit signifikantem Zusammenhang zwischen Integrationsindikator und Arbeitsmarktintegration von Zugewanderten</i>	<i>49</i>
<i>Tabelle 5.3</i>	<i>Vorschlag einer Item-Auswahl zur Operationalisierung von nicht administrativ erfassten Zielgrößen des Evaluationsvorhabens mit Hilfe der IAB-BAMF-SOEP-Befragung von Flüchtlingen</i>	<i>52</i>



## Abbildungsverzeichnis

Abbildung 3.1	<i>Schema der von der Statistik der BA für Drittstaaten-Angehörige dargestellten Arten des Aufenthaltsstatus in Deutschland</i>	27
Abbildung 4.1	<i>Bestand an Personen im Kontext von Fluchtmigration in arbeitsmarktpolitischen Instrumenten – Berichtsmonat: September 2016</i>	33
Abbildung 4.2	<i>Bestand an Drittstaaten-Angehörigen in arbeitsmarktpolitischen Instrumenten ohne Personen im Kontext von Fluchtmigration – Berichtsmonat: September 2016</i>	34
Abbildung 5.1	<i>Zentrale Bereiche bei der Integration von Flüchtlingen</i>	43
Abbildung 7.1	<i>Zeitplan und Meilensteine</i>	78

## Abkürzungsverzeichnis

AsylG	Asylgesetz
AZR	Ausländerzentralregister
BA	Bundesagentur für Arbeit
BAMF	Bundesamt für Migration und Flüchtlinge
BMAS	Bundesministerium für Arbeit und Soziales
CAPI	Computer Assisted Personal Interview Computerunterstützte persönliche Befragung
CATI	Computer Assisted Telephone Interview Computerunterstützte telefonische Befragung
ELB	erwerbsfähige Leistungsberechtigte
ESF	Europäischer Sozialfonds
IAB	Institut für Arbeitsmarkt- und Berufsforschung
ID	Identifikator
IEB	Integrierte Erwerbsbiografien
ILO	International Labor Organisation
IV	Instrumentvariable
IZA	Forschungsinstitut zur Zukunft der Arbeit
LATE	Local Average Treatment Effect – lokaler durchschnittlicher Treatment-Effekt
OECD	Organisation of Economic Co-operation and Development
PASS	Panel „Arbeitsmarkt und soziale Sicherung“
SGB II	Sozialgesetzbuch II
SGB III	Sozialgesetzbuch III
SOEP	Sozio-oekonomisches Panel
STEP	Stammdaten-Entwicklungsprojekt
VerBIS	Vermittlungs-, Beratungs- und Informationssystem
zkT	zugelassene kommunale Träger



## Zusammenfassung

Das Ziel der vorliegenden Studie ist es, die Machbarkeit einer wissenschaftlichen Standards genügenden Evaluation der arbeitsmarktpolitischen Maßnahmen zur Integration von Flüchtlingen im direkten Zuständigkeitsbereich des Bundesministeriums für Arbeit und Soziales zu prüfen und ein praktikables Konzept für die geplante systematische Begleitforschung zu diesen Maßnahmen zu entwickeln. Dazu wurden der Stand der einschlägigen Forschung zur Arbeitsmarktintegration von Flüchtlingen aufgearbeitet, die Datenverfügbarkeit im Detail geprüft sowie Vorschläge für geeignete Zielindikatoren und Leitfragen für die Evaluation entwickelt. Die auf dieser Grundlage angestellten konzeptionellen Überlegungen lassen sich dem Grunde nach auch auf andere integrationspolitische Maßnahmen übertragen. Insofern liefert die Studie auch Hinweise für Möglichkeiten zur Gestaltung weiterer denkbarer Evaluationsvorhaben zu anderen der zahlreichen Maßnahmen, die sich unmittelbar an geflüchtete Menschen richten.

Eine auf die Offenlegung von Ursache-Wirkungs-Beziehungen ausgerichtete Begleitforschung ist bei den arbeitsmarktpolitischen Maßnahmen und Programmen des Bundes mittlerweile Standard. Bei integrationspolitischen Maßnahmen für geflüchtete Menschen gestaltet sich die Lösung der Evaluationsaufgabe allerdings besonders schwierig. Die für Wirkungsanalysen benötigten Vergleichsgruppen können sinnvoll nur innerhalb der Gruppe der in jüngerer Zeit neu in Deutschland aufgenommenen Flüchtlinge gebildet werden. Diese bilden jedoch infolge komplexer Flucht-, Bildungs- und Erwerbsbiografien eine erheblich heterogenere und diffusere Gruppe als die herkömmlichen Adressaten der aktiven Arbeitsmarktpolitik.

Um mit Hilfe von quantitativen Analyseverfahren Wirkungen zu identifizieren, die sich ursächlich auf die Teilnahme an einer Maßnahme zurückführen lassen, ist man entweder darauf angewiesen, dass es quasi-experimentelle Zuweisungsprozesse gibt, oder dass sich eine valide Vergleichsgruppe, die sich von den Maßnahmenteilnehmerinnen und Maßnahmenteilnehmern lediglich so gut wie zufällig unterscheidet, wenigstens mit statistischen Methoden konstruieren lässt. Die Studie empfiehlt, das Evaluationsproblem mit einem Mix der machbaren Evaluationsmethoden anzugehen. Dies heißt, sowohl Verfahren einzusetzen, die darauf basieren, systematische Selektionsprozesse bei der Maßnahmenteilnahme anhand von beobachtbaren Größen zu beschreiben, als auch Verfahren, die so gut wie experimentelle Unterschiede beim Maßnahmenzugang ausschöpfen. Wählt man einen offenen Multi-Methoden-Ansatz, statt sich auf ein einziges Verfahren festzulegen, hat man Möglichkeiten, die Robustheit der geschätzten Maßnahmenwirkungen besser einzuschätzen. Führen verschiedene Methoden zu zahlenmäßig ähnlichen, qualitativ übereinstimmenden Befunden, ist dies ein Hinweis darauf, dass die Ergebnisse nicht von den hinter jeder Methode liegenden, letztlich nicht testbaren Annahmen abhängen.

Darüber hinaus empfiehlt die Studie, sich im Rahmen der Evaluation nicht allein auf den Einsatz statistischer Verfahren zu beschränken, sondern vielmehr quantitativ und qualitativ operierende Evaluationsweisen miteinander zu kombinieren. Auch dies kann zur Validierung der statistisch gewonnenen kausalanalytischen Aussagen beitragen; dazu sind mit den Verfahren der qualitativen Sozialforschung die

statistischen Befunde an den Erfahrungen, die in der Praxis von den Stakeholdern gemacht werden, zu spiegeln und plausible Erklärungsmuster dafür zu finden, wie die abgeschätzten Maßnahmenwirkungen zustande kommen.

Die administrative Datenbasis für die Durchführung einer quantitativ angelegten Evaluation der integrationspolitischen Maßnahmen im Zuständigkeitsbereich des Bundesministeriums für Arbeit und Soziales erweist sich als recht tragfähig. Eine valide statistische Abbildung von Personen im Kontext von Fluchtmigration ist in den Arbeitsmarktstatistiken der Bundesagentur für Arbeit inzwischen möglich. Somit bewegt sich das Evaluationsvorhaben von der administrativen Datenbasis her gesehen weitestgehend auf vertrautem Terrain mit etablierten Routinen des Datenzugangs.

Dennoch ist die erfolgreiche Durchführung des Evaluationsvorhabens auf ergänzende Befragungsdaten angewiesen. Grundsätzlich ist der Informationsgehalt der administrativen Daten begrenzt, weil Flüchtlinge erst seit kurzem in Deutschland leben. Somit stehen nur verhältnismäßig wenige amtlich registrierte Informationen zur Verfügung, um systematische Prozesse des Zugangs zu den Integrationsmaßnahmen überzeugend statistisch zu modellieren. Darüber hinaus existiert eine Vielzahl komplementärer Hilfs- und Unterstützungsangebote, die in den Daten der Bundesagentur für Arbeit nicht dokumentiert sind, jedoch für die Inzidenz des Einsatzes der integrationspolitischen Maßnahmen im Bereich der Zuständigkeit des Bundesministeriums für Arbeit und Soziales und deren Wirksamkeit eine Rolle spielen dürften. Schließlich erscheint es konzeptionell sinnvoll, den Katalog der Zielgrößen der Evaluation über die Arbeitsmarktintegration hinaus zu weiten und Integration als multidimensionales Konstrukt aufzufassen. Verbesserungen bei Indikatoren, die diesen anderen Dimensionen der Integration zuzurechnen sind, können nämlich als Frühindikatoren für später nachlaufende Verbesserungen bei der Arbeitsmarktintegration dienen, die im vorgesehenen Untersuchungszeitraum vielfach noch nicht beobachtet werden dürfte. Die zur Charakterisierung sozialer Teilhabe und Integration heranzuziehenden Kennziffern lassen sich nur zu einem geringen Teil aus den vorhandenen administrativen Daten gewinnen.

Die Durchführung der erforderlichen Erhebungen ist wegen der besonderen Zielgruppe eine komplexe Aufgabe. Um den Aufwand in vertretbaren Grenzen zu halten, empfiehlt die Studie erstens eine gezielte Auswahl der Stichprobe aus der Grundgesamtheit der arbeitssuchenden Personen im Kontext von Fluchtmigration. Zum einen sollten nur „statistische Zwillinge“ von Teilnehmenden und Nichtteilnehmenden an den zu evaluierenden Maßnahmen eingezogen werden. Zum anderen sollte die Erhebung auf eine Auswahl von Agenturbezirken der Bundesagentur für Arbeit beschränkt werden, die geeignet sind, das Gesamtbild des Geschehens wiederzugeben. Weiterhin wird empfohlen, sich im Hinblick auf das Fragenprogramm und die Erhebungsweise der für das Evaluationsvorhaben erforderlichen Längsschnitterhebungen an das Vorbild der laufenden IAB-BAMF-SOEP-Befragung von Flüchtlingen zu orientieren; prinzipiell denkbar wäre die Bildung einer spezifischen Ergänzungsstichprobe zu dieser Befragung, wobei aus Gründen der Wirtschaftlichkeit von deren repräsentativer Anlage abgewichen werden müsste. Zudem wären Ergänzungen des Fragenprogramms erforderlich, insbesondere, um nähere Informationen zu den auf die Zielpersonen einwirkenden Maßnahmen bzw. Treatments zu erhalten.

Die Studie wägt ab, welche Leistungen im Rahmen des geplanten Evaluationsvorhabens spezifisch untersucht werden sollen. Hier spricht die geringe Zahl der Flüchtlinge, die an Maßnahmen zur Teilhabe behinderter Menschen teilnehmen, dafür, diese Kategorie aus dem geplanten umfassenden Evaluationsvorhaben herauszulösen und dafür ein eigenständiges, rein qualitativ orientiertes Forschungsvorhaben aufzulegen. Unabhängig von der Entwicklung der Fallzahlen empfiehlt die Studie keine Betrachtung der Kategorie der Freien oder sonstigen Förderung, die unter den arbeitsmarktpolitischen Instrumenten der Bundesagentur für Arbeit allgemein von untergeordneter Bedeutung ist und es sich dabei zudem um eine Sammelkategorie handelt. Die übrigen Förderbereiche der arbeitsmarktpolitischen Regelleistungen der Bundesagentur für Arbeit sollten dagegen jeweils auf ihre spezifischen Wirkungen hin untersucht werden. Darüber hinaus sollten nach Möglichkeit die spezifischen Effekte für Unterkategorien von Instrumenten gemessen werden, die sich inhaltlich sinnvoll zueinander gruppieren lassen. Falls die Fallzahlen eine solche Differenzierung aber nicht erlauben, kann es sinnvoll sein, kleine Leistungen von vornherein nicht für eine quantitativ orientierte Evaluation vorzusehen, statt sie gemeinsam mit den anderen Leistungen eines Förderbereichs zu bewerten. Aussagen zur Bedeutung und potenziellen Wirkung dieser Leistungen wären dann gegebenenfalls im Rahmen der qualitativ angelegten Begleitforschung zu gewinnen.

Um herauszufinden, wie Interaktionen die Wirksamkeit eines Leistungs- oder Maßnahmentyps systematisch beeinflussen, empfiehlt die Studie, Gestaltungsunterschiede und die spezifische Umsetzungsqualität systematisch zu erfassen, wozu Informationen, die über die Daten aus der Förderstatistik der Bundesagentur für Arbeit hinausgehen, durch eine systematische Befragung der Jobcenter sowie beauftragter Maßnahmenträger gesammelt werden müssen. Die gewonnenen Informationen sind dann auf eine eher geringe Zahl von Qualitätsdimensionen zu verdichten, um festzustellen, ob sich die Wirksamkeit einer Maßnahme systematisch verändert, wenn sie mit einer bestimmten Qualität der Umsetzung verbunden ist.

Auf ähnliche Weise kann mit anderen Leistungen zur Integration von Flüchtlingen umgegangen werden. In der deutschen Praxis hat sich ein komplexes Geflecht öffentlich und privat organisierter Unterstützungsangebote für Flüchtlinge herausgebildet. Die jeweils am Ort vorhandenen flüchtlingsbezogenen Hilfen sollten durch neutrale Beobachtung der Gegebenheiten erfasst werden. Darüber hinaus empfiehlt die Studie eine Erhebung der Kenntnisse über die Inanspruchnahme von Angeboten, die nicht zu den Regelleistungen der Bundesagentur für Arbeit zählen, bei den Flüchtlingen selbst, da diese Leistungen schwer überschaubar sind und nicht systematisch registriert werden. So lassen sich Interaktionseffekte schätzen und es kann die Frage beantwortet werden, ob sich die Wirksamkeit der Teilnahme an einer Regelleistung systematisch verändert, wenn die Flüchtlinge zusätzlich bestimmte Beratungs- und Hilfsangebote in Anspruch nehmen.

Bezüglich der Evaluation der Wirkungen einer Teilnahme an Kursen und Maßnahmen zum Erlernen der deutschen Sprache empfiehlt die Studie, im Rahmen der geplanten Evaluation zwar den Einfluss von Maßnahmen zur Förderung des Erlernens der deutschen Sprache mit zu untersuchen, jedoch keine eigenständige Evaluation einzelner Sprachfördermaßnahmen anzustreben. Vielmehr sollte lediglich geklärt

werden, inwieweit die Teilnahme an Kursen und Maßnahmen zum Spracherwerb die Wirkungen der zu evaluierenden arbeitsmarktpolitischen Instrumente der BA verstärkt. Es sollten also nur Interaktionseffekte von verschiedenen Typen und Verlaufsmustern von Sprachfördermaßnahmen geschätzt werden. Die Studie schlägt darüber hinaus vor, die zur Umsetzung dieses Ansatzes benötigten Informationen zur Wahrnehmung der verschiedenen Möglichkeiten zum Erlernen der deutschen Sprache bei den Flüchtlingen selbst zu erheben.

Was die Organisation der Durchführung einer Evaluation der integrationspolitischen Maßnahmen im Verantwortungsbereich des Bundesministeriums für Arbeit und Soziales angeht, empfiehlt die Studie eine Modularisierung des Vorhabens, also die Zerlegung in eine Anzahl von Teilprojekten, die sich separat durch unterschiedliche Wissenschaftsakteure bearbeiten lassen. Die bei dieser Vorgehensweise zu lösende Aufgabe der Koordination und Steuerung des Gesamtvorhabens kann beispielsweise durch Einrichtung einer Geschäftsstelle gelöst werden. Ein gewünschter Nebeneffekt der Modularisierung ist, dass damit der wissenschaftliche Entdeckungswettbewerb angeregt wird, so dass sich über die im Rahmen des Vergabeverfahrens festzulegenden Vorgaben hinaus methodische Innovationen entwickeln können. Um inhaltliche Überschneidungen zwischen den einzelnen Modulen zu begrenzen und inhaltlich möglichst spezifische Untersuchungen anzuregen, empfiehlt die Studie eine Modularisierung der kausalanalytischen Teile des Vorhabens nach dem Typus der zu bewertenden arbeitsmarktpolitischen Integrationsmaßnahmen. So wären parallel zu bearbeitende Module zur Evaluation der Leistungen zur Aktivierung und beruflichen Eingliederung, der Leistungen zur Berufswahl und Berufsausbildung, der Leistungen zur beruflichen Weiterbildung, sowie der Leistungen zur Aufnahme einer Erwerbstätigkeit und der Beschäftigung schaffenden Maßnahmen denkbar. Weiterhin wird ein eigenständiges Modul zur Evaluation des Konglomerats an Leistungen zur Sprachförderung angeregt, das auf die Bewertung der Wirksamkeit dieser Leistungen in der Kombination mit den integrationspolitischen Maßnahmen und Leistungen der BA fokussiert.

Schließlich empfiehlt die Studie, die Durchführung der Erhebungen zur Gewinnung der für die Wirkungsforschung als Ergänzung der administrativen Daten benötigten Verlaufsinformationen zu Personen im Kontext von Fluchtmigration in ein eigenständiges Modul zu verlagern. Der Vorschlag, ein solches Querschnittsmodul zu schaffen, trägt der Überlegung Rechnung, dass zu dessen Durchführung sehr spezifische Kompetenzen und Kapazitäten auf dem Feld der sozialwissenschaftlichen Umfrageforschung erforderlich sind. Zudem führt diese gebündelte Verfahrensweise zu größeren Fallzahlen. Damit verbundene Lerneffekte, sowie die sinkenden Setup-Kosten pro Fall, dürften gegenüber einer Struktur, bei der in die wirkungsanalytischen Module jeweils eigenständige Verlaufserhebungen integriert sind, die Gesamtkosten der Erhebung deutlich verringern.

## 1. Ausgangslage und Aufgabenstellung

In den Jahren 2015 und 2016 hat Deutschland aus humanitären Gründen so viele Menschen aufgenommen wie nie zuvor. Nach Angaben des Bundesinnenministeriums wurden im Jahr 2015 insgesamt 890.000 neu ankommende Flüchtlinge registriert; im Jahr 2016 waren es 280.000 Menschen. Beim Bundesamt für Migration und Flüchtlinge (BAMF) gingen im Jahr 2015 knapp 480.000 formelle Asylanträge ein, und im Jahr 2016 wurden sogar knapp 750.000 Erst- und Folgeanträge entgegen genommen. Die Gesamtschutzquote bei den Entscheidungen über Asylanträge lag gemäß der Geschäftsstatistik des BAMF im bisherigen Berichtsjahr 2016 bei 63,3 Prozent (BAMF 2016). Der Anteil der Asylsuchenden aus Kriegs- und Bürgerkriegsländern ist hoch, sodass viele Antragstellerinnen und Antragsteller eine Rechtsstellung als Flüchtling oder den subsidiären Schutz gemäß §4 Abs. 1 AsylG erhalten. Da viele der auf diese Weise Schutz findenden Menschen wegen der absehbar fortbestehenden Fluchtgründe länger in Deutschland bleiben dürften, stellt sich die Frage, wie ihre Integration bei uns gut gelingen kann.

Hierbei geht es zum einen um die erfolgreiche Arbeitsmarktintegration der aus humanitären Gründen aufgenommenen Zuwanderinnen und Zuwanderer, von denen viele eher schlechte Voraussetzungen mitbringen, in der hochproduktiven deutschen Wirtschaft Fuß zu fassen. So sind die Neuankömmlinge nach den bislang vorhandenen empirischen Beobachtungen im Durchschnitt geringer beruflich qualifiziert als die einheimische Bevölkerung oder sonstige Migrantinnen und Migranten (IAB 2015; Worbs und Bund 2016). Auch ihre Schulbildung ist stark polarisiert (IAB 2016). Zum anderen geht es darum, die Flüchtlinge, die vielfältige nationale und ethnische Identitäten mitbringen, soziokulturell zu integrieren und gleichberechtigt am gesellschaftlichen Gefüge der Aufnahmegesellschaft teilhaben zu lassen.

Diese beiden Integrationsziele sind nicht unabhängig voneinander zu erreichen. Der Zugang zu einer Beschäftigung ist wegen des gesellschaftlichen Normalbildes der an der Erwerbsarbeit orientierten Lebensführung ein wichtiger Schlüssel zur sozialen Teilhabe (vgl. Kronauer 2006; Europäische Kommission 2004). Zugleich sind viele der Elemente, die zusammen soziale Teilhabe ausmachen – wie etwa die Kompetenzen (Sprachkenntnisse, Qualifikation und Erfahrungen), die physische, psychische und psychosoziale Gesundheit, die Zugehörigkeit zu Netzwerk- und Hilfestrukturen und funktionierende Nahbeziehungen – empirisch relevante Faktoren, die sich systematisch auf die Beschäftigungschancen von zugewanderten Personen auswirken können (vgl. Grin 1996; De Paola und Brunello 2016; Ahonen et al. 2007; Arbel et al. 2012).

Wie die Integrationsprozesse bei den neu von Deutschland aufgenommenen Flüchtlingen verlaufen werden, lässt sich angesichts der historischen Einmaligkeit der momentanen Entwicklung weiterhin nicht verlässlich vorhersagen. Zwar gibt es mittlerweile intensive Bemühungen, Integration hemmende Rechtsunsicherheit durch Verkürzung der Asylverfahren zu verkleinern. Auch wurden die rechtlichen Voraussetzungen für den Zugang zum Arbeits- und Ausbildungsmarkt für die Menschen, die Deutschland aus humanitären Gründen aufnimmt und eine gute Bleibeperspektive haben, in Reaktion auf die Integrationsaufgaben seit dem letzten Jahr in mehreren Schritten deutlich abgesenkt, etwa mit der derzeit in 133 von 156 Arbeitsagenturbezirken befristet für drei Jahre ausgesetzten Vorrangprüfung. Angesichts der ungünstigen individuellen Ausgangsvoraussetzungen vieler Flüchtlinge spricht jedoch einiges dafür, dass sich ihre Integration trotz der Verbesserungen bei den institutionellen Rahmenbedingungen nicht von selbst reibungslos vollziehen wird.

So bedeuten die Rekrutierung und bedarfsgerechte Qualifizierung von Flüchtlingen, die über keine ausreichenden sprachlichen oder fachlichen Kompetenzen verfügen, für die Arbeitgeber einen beträchtlichen Aufwand. Obwohl in manchen Segmenten der Arbeits- und Ausbildungsmärkte in Deutschland Engpässe erkennbar sind, wird sich dieser Aufwand rein betriebswirtschaftlich betrachtet in



vielen Konstellationen nicht rechnen. Die Einstellungsbereitschaft der Unternehmen allein dürfte die Integration der humanitären Zuwanderinnen und Zuwanderer in Beschäftigung somit kaum tragen. Dies macht den begleitenden Einsatz aktiver arbeitsmarktpolitischer Maßnahmen bzw. Instrumente zu einer wichtigen öffentlichen Aufgabe.

Vermittlung von Arbeitsmarkt- und Berufsorientierung, Förderung der Entwicklung allgemeiner beruflicher Fähigkeiten und Kompetenzen inklusive Sprachkenntnisse, sowie Unterstützungsleistungen bei Stellensuche und Berufseinstieg können die Beschäftigungsfähigkeit der Flüchtlinge verbessern und die Arbeitgeber entlasten. Gelingt dies, verbessern sich die Beschäftigungs- und Einkommenschancen der Zielgruppe, und zwar angebots- und nachfrageseitig zugleich. Um Konkurrenzsituationen zum Nachteil der Geringqualifizierten in der schon länger in Deutschland lebenden Bevölkerung zu vermeiden und auch Nachhaltigkeitszielen gerecht zu werden, sind insbesondere aktive Hilfen gefragt, die geeignet sind, die Flüchtlinge auf einen Pfad hin zur Ausübung qualifizierter Tätigkeiten zu bringen. Im Übrigen kann die mit der Bereitstellung aktiver Unterstützungsangebote verbundene individuelle Betreuung durch die relevanten Arbeitsmarktakteure (Agenturen für Arbeit, Jobcenter, BAMF) helfen zu vermeiden, dass Flüchtlinge, etwa aus Mangel an Information oder Kurzfristorientierung heraus, Entscheidungen treffen, die ihre nachhaltige Integration in gute Beschäftigung erschweren, oder sich den an sie gestellten Integrationserwartungen entziehen.

Um eine Integration der Flüchtlinge in möglichst gute und nachhaltige Beschäftigung zu fördern und zu beschleunigen, womit indirekt auch dem Ziel der soziokulturellen Integration gedient wäre, hat die Bundesregierung die gesetzlichen Grundlagen für den Zugang geflüchteter Menschen zu aktiven arbeitsmarktpolitischen Maßnahmen geändert. Diese Maßnahmen werden nun unter Beteiligung der Bundesagentur für Arbeit (BA), des BAMF sowie der Länder und Kommunen implementiert. Abhängig vom individuellen Bedarf kommen hierbei sowohl Standardinstrumente der aktiven Arbeitsmarktpolitik als auch spezifische Kombinationsmaßnahmen, die gesondert für die Zielgruppe der Flüchtlinge ausgestaltet werden, zum Einsatz.

Dass sich aktive arbeitsmarktpolitische Integrationsmaßnahmen als soziale Investition grundsätzlich rentieren können, veranschaulichen Simulationsrechnungen (Bonin 2016). Eine notwendige Bedingung dafür ist aber, dass es wirksame aktive Maßnahmen bzw. Instrumente gibt, mit denen sich Verbesserungen bei Qualität und Tempo der Integration erreichen lassen. Wissenschaftlich fundierte Erkenntnisse zur Wirksamkeit und zur Kosten-Nutzen-Relation der arbeitsmarktpolitischen Integrationsmaßnahmen sind deswegen von großem sozial- und finanzpolitischem Interesse.

Vor diesem Hintergrund ist es das Ziel der vorliegenden Studie, die Machbarkeit einer wissenschaftlichen Standards genügenden Evaluation der arbeitsmarktpolitischen Maßnahmen zur Integration von Flüchtlingen im direkten Zuständigkeitsbereich des Bundesministeriums für Arbeit und Soziales (BMAS) zu prüfen und ein praktikables Konzept für die geplante systematische Begleitforschung zu diesen Maßnahmen zu entwickeln. Damit eine Begleitevaluation steuerungsrelevante Erkenntnisse hervorbringt, muss sie inhaltlich und methodisch darauf ausgerichtet sein, empirisch belastbare Wirkungsaussagen zu einzelnen Interventionen oder auch zu Gruppen zusammenwirkender Interventionen zu erhalten, und zudem fundierte sachliche Begründungen für die gemessenen Wirkungen liefern. Außerdem müssen die zu entwickelnden Vorschläge für ein zielführendes Evaluationsdesign der Randbedingung Rechnung tragen, dass sie in vertretbarer Zeit und mit vertretbarem Ressourceneinsatz umsetzbar sind.

Eine so ausgelegte Begleitforschung ist bei den arbeitsmarktpolitischen Maßnahmen und Programmen des Bundes mittlerweile Standard. Dagegen gibt es – auch international – bislang praktisch keine Erfahrungen mit wirkungsanalytisch angelegten Studien zur Messung der Wirksamkeit und Effizienz aktiver Programme und Maßnahmen zur Verbesserung der Arbeitsmarktintegration für Flüchtlinge (Ott 2013).

Tatsächlich erscheint die Lösung der Evaluationsaufgabe im Kontext der Flüchtlingsmigration anspruchsvoll.

Erstens können die für Wirkungsanalysen benötigten Vergleichsgruppen sinnvoll nur innerhalb der Gruppe der in jüngerer Zeit neu in Deutschland aufgenommenen Flüchtlinge gebildet werden. Es ist nämlich davon auszugehen, dass sich diese Personengruppe im Hinblick auf die Merkmale, die den Integrationsstand beeinflussen können, systematisch von allen anderen Bevölkerungsgruppen unterscheidet. So dürfte der naheliegende Vergleich zwischen den neu aufgenommenen Flüchtlingen und Personen, die bereits früher als Asylsuchende nach Deutschland gekommen sind, zu verzerrten Ergebnissen führen. Zwischenzeitlich haben sich die allgemeine Arbeitsmarktlage und vor allem die institutionellen Voraussetzungen für den Arbeitsmarkt- und Ausbildungszugang substantiell geändert, und Unterschiede in den Startbedingungen zum Zeitpunkt der Zuwanderung können die Entwicklungsverläufe am Arbeitsmarkt von Zugewanderten nachhaltig beeinflussen (Aslund und Rooth 2007). Muss man demnach die Maßnahmeneffekte innerhalb der neu aufgenommenen Flüchtlinge identifizieren, ist man darauf angewiesen, dass es quasi-experimentelle Zuweisungsprozesse im Feld gibt, oder dass sich eine valide Vergleichsgruppe, die sich von der Gruppe der Maßnahmenteilnehmerinnen und Maßnahmenteilnehmer lediglich so gut wie zufällig unterscheidet, mit statistischen Methoden konstruieren lässt.

Zweitens hat sich in der deutschen Praxis ein komplexes Geflecht öffentlich und privat organisierter Unterstützungsangebote für Flüchtlinge herausgebildet, die sich in ihren Wirkungen gegenseitig verstärken oder konterkarieren können. Die zu erwartenden Wirkungsüberlagerungen bedeuten eine erhebliche Erschwernis, wenn es darum geht, die Effekte der Integrationsmaßnahmen im Verantwortungsbereich des BMAS zu isolieren. Hierzu müssen systematische Wirkungsunterschiede der Unterstützungsangebote des Bundes herausgearbeitet werden, je nachdem, wie sie mit anderen Unterstützungsangeboten kombiniert die Flüchtlinge erreichen. Die hierbei zu bewältigenden Herausforderungen sind mit denen vergleichbar, die bei einer systematischen Evaluation der Bundesleistungen im Bereich der aktiven Arbeitsmarktpolitik für Jugendliche unter 25 Jahren auftreten würden (ZEW et al. 2010). Die zu kontrollierenden Nebenleistungen streuen auf der Länderebene und vor allem auf der kommunalen Ebene in quantitativer wie qualitativer Hinsicht erheblich. In Folge der dezentralen, zum Teil gar nicht öffentlich organisierten Angebote fehlt für die Wirkungsforschung zudem eine einheitliche Datenbasis, um festzustellen, welche Unterstützungsleistungen auf der individuellen Ebene in Anspruch genommen werden. Dies zieht einen besonderen Erhebungsaufwand nach sich.

Bedenkt man die Dynamik der Flüchtlingsmigration und der zu ihrer Steuerung und Bewältigung ergriffenen politischen Initiativen und Maßnahmen, könnte sich in der Zukunft eine weitere Erschwernis für die Erforschung der Wirksamkeit der arbeitsmarktpolitischen Integrationsmaßnahmen für die Flüchtlinge daraus ergeben, dass sich Rahmenbedingungen oder der Instrumenteneinsatz selbst während der Laufzeit einer Evaluation gravierend verändern. Ein tragfähiges Konzept für die Begleitforschung sollte deshalb so flexibel angelegt sein, dass das Untersuchungsdesign falls erforderlich im Verlauf des Vorhabens mit vertretbarem Aufwand angepasst werden kann. Dies spricht für einen modularen Aufbau.

Vor diesem Hintergrund soll diese Machbarkeitsstudie eine Reihe von Aufgaben erfüllen. Zunächst erfolgt in Kapitel 2 die Aufarbeitung des Forschungsstands, um zu prüfen, welche Erkenntnisse aus der vorhandenen Forschung für das geplante Evaluationsvorhaben gewonnen werden können. Als Grundlage für die Ausarbeitung eines Daten- und Erhebungskonzepts wird in Kapitel 3 die Verfügbarkeit und Geeignetheit administrativer Daten für die Zwecke des Evaluationsvorhabens geprüft. Anschließend erfolgt in Kapitel 4 eine Prüfung, welche arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge im Aufgabenbereich des BMAS für eine Evaluation in Frage kommen. Die Erarbeitung geeigneter Zielindikatoren, welche der Erwartung Rechnung tragen, dass sich der Übergang in eine (ungeförderte)

Beschäftigung als Oberziel aktiver integrationspolitischer Maßnahmen für Flüchtlinge im direkten Zuständigkeitsbereich des BMAS schleppend vollziehen könnte, und zudem eine Bewertung des Integrationsstands in einem umfassenderen Sinn ermöglichen, wird in Kapitel 5 vorgenommen. Im folgenden Kapitel 6 werden methodisch geeignete Vorgehensweisen zur Erlangung aussagekräftiger Wirkungsanalysen entwickelt. Schließlich erfolgt in Kapitel 7 die Entwicklung eines Vorschlags für die organisatorische Umsetzung einer umfassenden Evaluation der Wirksamkeit der arbeitsmarktpolitischen Integrationsmaßnahmen des Bundes, inklusive einer Abschätzung des mit einer solchen Evaluation verbundenen finanziellen Aufwands.

## 2. Forschungsstand

Die nachstehende Aufarbeitung des einschlägigen Forschungsstands erfolgt mit dem Hauptziel, die wesentlichen Erkenntnisse aus der bereits vorhandenen nationalen und internationalen Forschung zur Wirksamkeit integrationspolitischer Maßnahmen für das Design dieses Evaluationsvorhabens herauszuarbeiten. Die Sichtung der relevanten Forschung erfolgte per Desk Research und berücksichtigt verschiedene Literaturstränge, unter anderem:

- Studien zu spezifischen integrationspolitischen Maßnahmen für Flüchtlinge, wie beispielsweise die Evaluation des ESF-Bundesprogramms zur arbeitsmarktlichen Unterstützung für Bleibeberechtigte und Flüchtlinge mit Zugang zum Arbeitsmarkt (Mirbach et al. 2014) oder die Begleitforschung zum Modellprojekt „Early Intervention“ (Daumann et al. 2015);
- Evaluationsstudien zu allgemeinen integrationspolitischen Maßnahmen, mit besonderem Fokus auf wirkungsorientierte Studien, wie zum Beispiel die flächendeckende Evaluation der Integrationskurse durch das Bundesministerium des Innern und Ramboll-Management (2006);
- Studien zur Wirkung aktiver arbeitsmarktpolitischer Maßnahmen mit Fokus auf Zugewanderte, wie etwa die Evaluation des Bundesprogramms zur berufsbezogenen Sprachförderung von Migrantinnen und Migranten (Walter et al. 2014).

Es werden neben auf Deutschland bezogenen Studien auch wichtige Studien für andere Länder in den Literaturüberblick einbezogen. Neben Erkenntnissen für die Gestaltung einer Begleitforschung zu den aktuellen arbeitsmarktpolitischen Integrationsmaßnahmen, insbesondere in methodischer Hinsicht, werden auch konzeptionelle Schwächen der vorhandenen Studien herausgearbeitet, um auf diese Weise weitere Gestaltungshinweise zu gewinnen. Zentrales Kriterium für die Bewertung der gesichteten Studien ist die interne und externe Validität der empirischen Befunde.

Im Ergebnis zeigt der Literaturüberblick erstens, dass Flüchtlinge in vielen Konstellationen im Vergleich zu anderen Zuwanderergruppen sowie im Vergleich zur schon länger in Deutschland lebenden Bevölkerung am Arbeitsmarkt im Nachteil sind – auch wenn beobachtbare Merkmale mit Einfluss auf die Arbeitsmarktposition berücksichtigt werden („Refugee Gap“). Daraus folgt für das Design des Evaluationsvorhabens, dass Wirkungsanalysen bevorzugt auf Kontrollgruppen innerhalb der Flüchtlinge aufbauen sollten. Die Literatur verweist in Zusammenhang mit dem „Refugee Gap“ auch auf eine ganze Reihe individueller und Kontextgrößen, die systematisch mit der Arbeitsmarktposition von Flüchtlingen korrelieren. Die diesbezüglichen empirischen Erkenntnisse fließen später in unsere Vorschläge zu den intermediären Integrationszielen der Evaluation ein, die als Vorlaufvariable für die Integration in den Arbeitsmarkt angesehen werden können.

Der Literaturüberblick zeigt zweitens gravierende Lücken im Bereich der Evaluation integrationspolitischer Maßnahmen für Flüchtlinge („Literature Gap“). Die Anzahl der zu dieser Thematik verfügbaren Studien ist nach wie vor klein, und die vorliegenden Studien verwenden häufig Vorgehensweisen, die nicht geeignet sind, um Ursache-Wirkungs-Beziehungen offen zu legen. Deswegen ergeben sich aus dem Design dieser Studien auch insgesamt nur relativ wenige Anhaltspunkte für die Lösung der anstehenden Evaluationsaufgabe.

## 2.1 Nachteile von Flüchtlingen auf dem Arbeitsmarkt

Die einschlägige Forschung zeigt, dass Flüchtlinge in vielen Konstellationen deutliche Nachteile auf dem Arbeitsmarkt haben und insgesamt schlechtere Arbeitsmarktergebnisse aufweisen – sowohl im Vergleich zur einheimischen Bevölkerung als auch im Vergleich mit anderen Zuwanderergruppen (vgl. Gangiano 2012). Dieser Nachteil („Refugee Gap“) bleibt in der Regel auch bestehen, wenn beobachtbare Merkmale berücksichtigt werden, welche die Arbeitsmarktergebnisse typischerweise beeinflussen (wie etwa Sprachkenntnisse, Bildungsniveau, Arbeitserfahrung; vgl. OECD 2015).

So bestehen für Flüchtlinge zunächst grundsätzlich erhebliche Schwierigkeiten, in den Arbeitsmarkt einzutreten und eine Beschäftigung aufzunehmen. Das ist auch international der Fall: In Großbritannien zeigt sich diesbezüglich etwa eine beachtliche Diskrepanz zwischen recht ausgeprägten Erwerbsaspirationen der geflüchteten Menschen und einer nur geringen realisierten Erwerbsbeteiligung (Evans und Murray 2009; Platts-Fowler und Robinson 2011). Auch in Deutschland zeigt eine erste Auswertung der IAB-BAMF-SOEP-Befragung von Geflüchteten, dass rund 80 Prozent der nichterwerbstätigen Flüchtlinge angeben, künftig „ganz sicher“ arbeiten zu wollen (IAB, 2016). Gleichzeitig waren zum Befragungszeitpunkt von den in den Jahren 2015 und 2016 zugezogenen Geflüchteten 13 Prozent erwerbstätig, von den im Jahr 2014 zugezogenen geflüchteten 22 Prozent und von den im Jahr 2013 und zuvor zugezogenen Geflüchteten 31 Prozent. Bei früheren Kohorten von Flüchtlingen betrug die Erwerbstätigkeitsquote nach 5 Jahren etwa 50 Prozent (vgl. Brücker et al. 2015a).

Auch europaweit zeigt sich, dass humanitäre Migranten erst nach gut 5 Jahren zu anderen Gruppen von Migranten wie etwa Familiennachzülern aufschließen (vgl. OECD, 2016, S. 11). Es gelingt Flüchtlingen in dieser Betrachtung aber über den gesamten Beobachtungszeitraum von knapp 20 Jahren nicht, mit ihrer Beschäftigungsquote zur Gruppe der Migranten, die zu Erwerbszwecken oder für ein Studium zuwandern, aufzuschließen. Anfängliche Nachteile von Flüchtlingen auf dem Arbeitsmarkt können grundsätzlich im Zeitverlauf also geringer werden und sogar verschwinden (vgl. OECD 2014). Dieser Aufholprozess erstreckt sich jedoch typischerweise über einen recht langen Zeitraum.

Der „Refugee Gap“, der sich nur sehr allmählich und gegenüber einigen Zuwanderergruppen auch gar nicht schließt, hängt mit verschiedenen Faktoren und Kontextgrößen zusammen. Auch wenn sich zum Beispiel die rechtlichen Rahmenbedingungen für den Arbeitsmarktzugang in Deutschland in letzter Zeit mehrfach geändert haben und Verbesserungen wie etwa eine Verkürzung des Arbeitsverbotes für Asylsuchende in Kraft getreten sind (vgl. Thränhardt 2015), so wirkte sich der eher restriktive Arbeitsmarktzugang tendenziell negativ auf die Arbeitsmarktergebnisse der geflüchteten Menschen aus. Die größten Hürden für Flüchtlinge auf dem Arbeitsmarkt sind jedoch weiterhin fehlende Sprachkenntnisse und eine geringe (formale) Qualifikation bzw. die Anerkennung von Qualifikationen.

In vielen Fällen steht eine geringe Qualifikation oder eine fehlende formale Qualifikation dem Einstieg in den Arbeitsmarkt entgegen, und die Anerkennung von im Ausland erworbenen Abschlüssen ist ebenfalls weiterhin als hohe Hürde anzusehen. Und auch wenn diese Hürde überwunden ist, zeigt sich zwar einerseits, dass nach der Anerkennung von Abschlüssen die Beschäftigungschancen grundsätzlich zunehmen (Lemaître 2007). Andererseits ist damit aber häufig auch eine steigende Anspruchshaltung verbunden, die möglicherweise eine verlängerte Suchperiode nach einer „passenden“, also qualifikationsäquivalenten, Beschäftigung nach sich zieht (Correa-Valez et al. 2013).

Fehlende Sprachkenntnisse sind ebenfalls als ein zentraler Faktor für die unterdurchschnittliche Arbeitsmarktintegration von geflüchteten Menschen und zugewanderten Personen anzusehen (Chiswick und Miller, 2007, 2009). So zeigen zum Beispiel Studien für Großbritannien, dass die Beherrschung der englischen Sprache dort mit einem Anstieg der Beschäftigungswahrscheinlichkeit um bis zu 30 Prozent

einhergeht (Dustmann und Fabbri 2003; Shields und Wheatley Price 2001, 2002). Auch in Deutschland zahlen sich Sprachkenntnisse auf dem Arbeitsmarkt aus: Dustmann (1994) ermittelt eine Lohnprämie von gut 7 Prozent, wenn männliche Zuwanderer gute schriftliche Kenntnisse der deutschen Sprache vorweisen können.

Des Weiteren erschweren in der Gruppe der geflüchteten Menschen vergleichsweise häufig Einschränkungen der Beschäftigungsfähigkeit durch körperliche und psychische Beeinträchtigungen und Krankheiten den Eintritt in den Arbeitsmarkt. Traumatisierung, Einsamkeit und Depression spielen in diesem Zusammenhang eine Rolle, auch wenn diesbezüglich Forschungsbedarf besteht (etwa wie hoch das genaue Ausmaß der Inzidenz unter den geflüchteten Menschen tatsächlich ist; vgl. IAB, 2016). Auch mangelnde Arbeitserfahrungen und Kenntnisse des deutschen Arbeitsmarktes erschweren eine Arbeitsaufnahme. Vor allem geflüchtete Personen ohne Berufs- oder Hochschulbildung suchen ihre erste Arbeitsstelle auch deshalb in Deutschland häufig über soziale Kontakte (IAB, 2016), die jedoch erst etabliert werden müssen.

In Zusammenhang mit der Stellensuche wirkt sich schließlich auch Diskriminierung negativ auf die Arbeitsmarktergebnisse der Flüchtlinge aus. So zeigt etwa die Untersuchung von Weichselbaumer (2016), dass selbst in Deutschland aufgewachsene Bewerberinnen mit besten Deutschkenntnissen und „deutscher“ Bildungs- und Ausbildungsbiografie mit erheblichen Benachteiligungen konfrontiert sind, wenn sie einen türkisch klingenden Namen haben und noch dazu ein Bewerbungsfoto mit Kopftuch vorlegen. Kopftuchtragende Musliminnen müssen nach dieser Studie für eine Einladung zum Bewerbungsgespräch – bei identischer Qualifikation – mehr als viermal so viele Bewerbungen verschicken. Merkmale wie die Firmengröße, eine internationale Ausrichtung des Unternehmens, gute Deutschkenntnisse als Einstellungsvoraussetzung oder ein direkter Mitarbeiter- oder Kundenkontakt machten bei den Ergebnissen dieses Feldversuchs keinen Unterschied. Allerdings nahm das Ausmaß der Diskriminierung mit steigendem Qualifikationsniveau der ausgeschriebenen Stelle noch zu.

Gelingt trotz aller Hürden der Einstieg in den Arbeitsmarkt, so sind geflüchtete Menschen häufig (zunächst) in Arbeitsverhältnissen mit geringer Entlohnung und/oder niedrigem beruflichen Ansehen anzutreffen (Evans und Murray 2009). Auch eine Konzentration in einzelnen Sektoren und Branchen wird festgestellt (Ott 2011). Verschiedene Studien konstatieren auch für Deutschland, dass geflüchtete Menschen insbesondere im Niedriglohnsektor beschäftigt sind (vgl. die Verweise in SVR 2016, S. 18). Dabei können Konkurrenzsituationen mit schon länger im Aufnahmeland lebenden Menschen entstehen. Allerdings kann Flüchtlingszuwanderung auch eine höhere Mobilität in dieser Gruppe bewirken und schon länger im Land lebende Geringqualifizierte gelingt dadurch möglicherweise ein Aufstieg in komplexere Tätigkeiten (vgl. Foged und Peri, 2016).

## **2.2 Evaluationen integrationspolitischer Maßnahmen**

Rigore Evaluationsstudien integrationspolitischer Maßnahmen sind nach wie vor eine Seltenheit. Es findet diesbezüglich vor allem keine systematische Vorgehensweise statt. Die wenigen existierenden Studien, die wissenschaftlichen Standards genügen und Ursache-Wirkungs-Beziehungen offenlegen, sind überwiegend eigenständige Forschungsarbeiten. Deshalb existieren gravierende Lücken („Literature Gap“) und es fällt aufgrund der unvollständigen und wenig systematischen Forschung äußerst schwer, allgemeine Schlussfolgerungen zu ziehen oder konkrete Handlungsempfehlungen abzuleiten. Es zeigt sich allerdings in vielen Kontexten, dass integrationspolitische Maßnahmen und Programme mit direktem Arbeitsmarktbezug vergleichsweise positive Effekte für Zuwanderergruppen generieren (vgl. Rinne, 2013). Insbesondere werden in einer größeren Zahl von Evaluationsstudien für Lohnsubventionen im privaten Sektor positive Beschäftigungseffekte ermittelt (vgl. Butschek und Walter, 2013).

Eine nennenswerte Anzahl von relevanten Studien existiert im skandinavischen Kontext. So wurden zum Beispiel die Effekte von Integrationskursen („Introduction Programs“) relativ eingehend in Schweden analysiert – nicht zuletzt aufgrund der dortigen vergleichsweise langen Tradition dieser Maßnahmen, die bis in die späten 1960er Jahre reicht. Neuzugewanderten Personen soll mit diesen Programmen recht kurz nach ihrer Ankunft der Einstieg in den Arbeitsmarkt und ihre gesellschaftliche Integration erleichtert werden. Die Programme setzen sich in der Regel aus verschiedenen Elementen zusammen, wozu häufig Sprachkurse und Maßnahmen der aktiven Arbeitsmarktpolitik wie etwa Weiterbildung, Unterstützung bei der Arbeitsplatzsuche oder subventionierte Beschäftigung gehören.

Svantesson und Aranki (2006) analysieren in einem einfachen Regressionsansatz den Zusammenhang zwischen dem Einsatz der verschiedenen Elemente dieser Integrationskurse in Schweden auf den kurzfristigen Arbeitsmarkterfolg der Teilnehmer. Dabei zeigen sich heterogene Effekte der einzelnen Elemente mit tendenziell positiveren Ergebnissen für Einzelmaßnahmen, die Erfahrungen auf dem ersten Arbeitsmarkt beinhalten. Diese Resultate stellen jedoch lediglich Korrelationen, nicht aber kausale Wirkungszusammenhänge dar, denn die Selektivität der Programmteilnahme findet in dieser Untersuchung methodisch keine entsprechende Berücksichtigung.

Durch eine andere Vorgehensweise werden in der Studie von Andersson Joona und Nekby (2012) Ursache-Wirkungs-Beziehungen für diese Programme ermittelt. Dabei wird die Tatsache ausgenutzt, dass es in einigen schwedischen Regionen allein dem Zufall überlassen wurde, ob neuzugewanderte Personen an regulären Integrationskursen oder an „neuen“ Kursen mit veränderter Ausrichtung teilnahmen. Die auf diese Weise berechneten Ergebnisse geben Aufschluss über die relative Effektivität der beiden Programme, wobei sich Vorteile der neuen Kurse mit einer intensiveren und individuellen Beratung sowie flexibleren Sprachkursen zeigen. Diese Vorteile machten sich nicht nur in einer höheren Erwerbsbeteiligung der Programmteilnehmer bemerkbar, sondern diese nahmen im weiteren Zeitverlauf auch häufiger an anderen Weiterbildungsangeboten und Programmen teil. Offen bleibt, ob dies letztlich als ein positives Ergebnis zu bewerten ist, also ob diese anschließenden Programmteilnahmen zu einer besseren Arbeitsintegration führen, oder ob dies lediglich ein Hinweis auf sogenannte „Programmkarrieren“ ohne entsprechende Arbeitsmarktergebnisse ist. Denn wie bei vielen anderen Untersuchungen stehen die Ergebnisvariablen auch hier nur für einen recht kurzen Beobachtungszeitraum nach der Programmteilnahme zu Verfügung, sodass keine Aussagen über mittel- bis langfristige Effekte getroffen werden können. Die Frage nach der Nachhaltigkeit dieser Programme bleibt somit unbeantwortet.

Verschiedene Programme der „regulären“ aktiven Arbeitsmarktpolitik ohne spezifische Zielgruppenorientierung sind grundsätzlich ebenfalls als integrationspolitische Maßnahmen anzusehen, denn Zuwanderer nehmen regelmäßig an ihnen teil (und häufig bereits in einer recht frühen Phase nach ihrer Ankunft). Zum Teil werden derartige Maßnahmen auch explizit für die Zielgruppen von Zuwanderern geplant und durchgeführt. Aus diesem Grund lassen sich Studien zur Wirkung aktiver arbeitsmarktpolitischer Maßnahmen mit Fokus auf Zugewanderte in zwei Kategorien einteilen: Es gibt erstens Analysen, welche die Effektheterogenität derartiger Maßnahmen untersuchen und dabei (auch) die Gruppe der Zuwanderer berücksichtigen. Zweitens gibt es Studien, welche die kausalen Effekte von speziellen Maßnahmen untersuchen, an denen ausschließlich Zuwanderer teilnehmen.

Dieser zweiten Kategorie von Untersuchungen lässt sich zum Beispiel die Studie von Aslund und Johansson (2011) zuordnen, die ein speziell auf die Zielgruppe von Einwanderern und Flüchtlingen zugeschnittenes Programm analysieren, das qualifizierte Personen aus dieser Zielgruppe durch verschiedene Angebote bei der Arbeitsplatzsuche unterstützt. In einem Differenz-in-Differenzen-Ansatz nutzen die Autoren hier regionale Variation im Einsatz dieses Programmes aus. Sie können auf diese Weise zeigen, dass Zuwanderer in Regionen, in denen dieses Programm implementiert wurde, eine höhere

Übergangswahrscheinlichkeit in Beschäftigung haben als in Regionen, in denen dieses Programm (noch) nicht eingesetzt wurde.

Auch Sarvimäki und Hämäläinen (2010) analysieren eine aktive arbeitsmarktpolitische Maßnahme mit einem expliziten Fokus auf die Zielgruppe von Zuwanderern. Das untersuchte Programm in Finnland besteht aus einer individuellen Sequenz von Weiterbildungselementen und geförderter Beschäftigung, wobei eine Nichtteilnahme mit Sanktionen der Unterstützungsleistungen bestraft wird. Die Autoren nutzen zur Identifikation kausaler Effekte der Programmteilnahme eine institutionelle Besonderheit aus: Die zuvor freiwillige Programmteilnahme wurde ab einem bestimmten Datum verpflichtend und obligatorisch für alle Zuwanderer, die nach einem bestimmten Datum eingereist waren. Als diese Neuerung beschlossen wurde, lag das relevante Einreisedatum jedoch bereits in der Vergangenheit, weshalb sich Zuwanderergruppen, die kurz vor bzw. kurz nach diesem Datum herum eingereist sind, weder in beobachteten und noch in unbeobachteten Merkmalen unterscheiden dürften – allein die verpflichtende Programmteilnahme sollte sie unterscheiden. Im Ergebnis stellen die Autoren fest, dass die Maßnahme sowohl die Beschäftigungschancen als auch die Löhne der Teilnehmer erhöht und sich darüber hinaus negativ auf die Inanspruchnahme sozialer Unterstützungsleistungen auswirkt.

Clausen et al. (2009) untersuchen hingegen keine spezifische Maßnahme für Zuwanderer, sondern sie betrachten in ihrer Untersuchung vielmehr die relativen Effekte einer Teilnahme an verschiedenen Programmen der aktiven Arbeitsmarktpolitik für Zuwanderer kurz nach ihrer Ankunft in Dänemark. Als Methodik zur Identifikation kausaler Effekte verwenden sie einen timing-of-events-Ansatz und berücksichtigen dabei auch die Teilnahme an Sprachkursen sowie Fortschritte in den vorhandenen Sprachkompetenzen.<sup>1</sup> Sie konstatieren in ihren Ergebnissen erhebliche Lock-in-Effekte der Teilnahme an fast allen Programmen. Einzige Ausnahme sind in dieser Hinsicht Lohnsubventionen, die sich (wie auch Sprachkurse) zudem positiv auf die Übergangswahrscheinlichkeit in reguläre Beschäftigung auswirken. Auch diese Evaluationsstudie betrachtet die Ergebnisvariablen nur über einen eher kurzen Zeithorizont nach der Programmteilnahme und kann somit keine Aussagen zur längerfristigen Effektivität und Nachhaltigkeit treffen.

Eine Untersuchung der Wirkung von eher kürzeren Maßnahmen der aktiven Arbeitsmarktpolitik in Deutschland stammt von Thomsen et al. (2013). Dabei vergleichen die Autoren die Effektivität dieser Maßnahmen für Einheimische und Zuwanderer unter Verwendung statistischer Matching-Verfahren. Im Ergebnis zeigt sich, dass sich Maßnahmen zur Eignungsfeststellung positiv auf die Beschäftigungschancen sowohl von Einheimischen als auch von Zuwanderern auswirken. Dies trifft auch auf die Vermittlung von Kenntnissen und Fertigkeiten zu, sofern zum Zeitpunkt der Maßnahmenteilnahme der Leistungsbezug erst relativ kurz andauert. Die Autoren stellen hingegen keine positiven Effekte von Bewerbertraining und kombinierten Programmen fest. Schließlich zeigt die Untersuchung, dass sich etwaige Unterschiede in der Effektivität der Programmteilnahme zwischen Einheimischen und Zuwanderern allein auf beobachtbare Merkmale zurückführen lassen – derartige Unterschiede sind also nicht in der jeweiligen Gruppenzugehörigkeit begründet, sondern hängen ausschließlich mit sozio-demografischen Faktoren wie etwa dem Alter, der Bildung oder der bisherigen Arbeitsmarkterfahrung zusammen.

Evaluationsstudien, die Programme und Maßnahmen mit der expliziten Zielgruppe geflüchteter Menschen analysieren, sind sehr selten. Bei der Evaluation des ESF-Bundesprogramms zur Unterstützung für Bleibeberechtigte und Flüchtlinge mit Zugang zum Arbeitsmarkt von Mirbach et al. (2014) handelt es sich jedoch um eine Studie zu einer spezifischen integrationspolitischen Maßnahmen für Flüchtlinge im deutschen Kontext. Diese Untersuchung unterschied sich somit von den bisher diskutierten Arbeiten, da Flüchtlinge darin lediglich eine Teilgruppe der betrachteten Zuwanderergruppen waren – wenn

---

<sup>1</sup> Heinesen et al. 2011 verwenden eine ähnliche Methodik ebenfalls im dänischen Kontext.



überhaupt. Ein weiterer methodischer Unterschied besteht darin, dass Mirbach et al. (2014) keine kausalanalytische Vorgehensweise anwenden. Im Ergebnis werden verschiedene Aspekte wie etwa Programmsteuerung und Implementierung vor allem vor dem Hintergrund betrachtet, ob Zielvorgaben erreicht wurden. Darüber hinaus werden Vermittlungen in Arbeit auch quantitativ betrachtet, wobei Vergleiche mit im Vorfeld gesetzten Zielvorgaben und zwischen den beiden Förderrunden des Programms durchgeführt werden. Der kausale Effekt der Programmteilnahme wird jedoch nicht isoliert betrachtet. Insofern handelt es sich um eine deskriptive und eher qualitative Evaluation, die vor allem feststellt, dass das ESF-Bundesprogramms zur Unterstützung für Bleibeberechtigte und Flüchtlinge mit Zugang zum Arbeitsmarkt erfolgreich umgesetzt worden ist. Aussagen über die Effektivität und Effizienz der Umsetzung bleiben vor diesem Hintergrund jedoch sehr vage. So heißt es zum Beispiel in der Zusammenfassung der Studie (vgl. Mirbach et al. 2014, S. 74): „Zweitens legen die Befunde der Teilnehmendenerhebungen die Einschätzung nahe, dass mit den in der zweiten Förderrunde erreichten Übergängen in Arbeit bzw. Ausbildung das Potential der unter den gegebenen rechtlichen Bedingungen und der Aufnahmebereitschaft des Arbeitsmarktes Vermittelbaren in großem Umfang ausgeschöpft ist.“ Die Autoren stellen in diesem Zusammenhang außerdem fest, dass Übergänge in Arbeit von geflüchteten Menschen vor allem durch ihre Arbeitsmarktnähe und institutionelle Faktoren beeinflusst werden.

Daumann et al. (2015) beschreiben Erkenntnisse der qualitativen Begleitforschung zum Modellprojekt „Early Intervention“ basierend auf Interviews in sechs Modellagenturen. Diese Arbeit hat ebenfalls eher den Charakter einer Prozessevaluation und stellt keine kausale Wirkungsforschung im engeren Sinne dar, liefert aber einige Hinweise für das Design dieses Evaluationsvorhabens. So hat sich beispielsweise die schriftliche Datenerhebung in den Agenturen mit Abfragen zur formalen Ausbildung und zur Berufserfahrung wegen Sprachproblemen und mangelnder Erfahrung der geflüchteten Menschen im Umgang mit Formularen als wenig geeignet erwiesen (Daumann et al. 2015, S. 4). Bei den Zuweisungskriterien in die Maßnahme kam es zudem zu unterschiedlichen Auslegungen der Vorgaben durch einzelne Agenturen. In der Beratungs- und Vermittlungsarbeit stellten sich fehlende Sprachkenntnisse als recht hohe Hürde dar, weshalb Deutschkurse als wichtiges und unabdingbares Förderinstrument angesehen werden. Fehlende Sprachkompetenz wirkt sich auch in anderen Bereichen wie etwa der Kompetenzfeststellung aus. Weitere Probleme bereiten die rechtlichen Rahmenbedingungen. Die Autoren vermuten einen fundamentalen Zielkonflikt zwischen dem gewünschten schnellen Vermittlungserfolg auf der einen Seite und den langwierigen Prozessen der Vermittlung von Deutschkenntnissen, der Kompetenzfeststellung und der Anerkennung von Bildungs- und Berufsabschlüssen auf der anderen Seite. Schließlich werden als Grundvoraussetzungen für eine erleichterte Arbeitsmarktintegration der Zielgruppe eine umfassende Betreuung, Deutschkurse, eine geeignete Kompetenzfeststellung und darauf aufsetzende Unterstützungsleistungen angesehen.

Ergebnisse der Evaluation der Integrationskurse nach dem Zuwanderungsgesetz finden sich in Bundesministerium des Innern und Ramboll-Management (2006). Darin kommen die Autoren vor dem Hintergrund ihrer Analysen zu dem Schluss, dass die Integrationskurse grundsätzlich eine deutliche qualitative Verbesserung der deutschen Integrationspolitik darstellen. Wesentliche Verbesserungen werden in der Zusammenführung von Instrumenten in einem Ansatz, in der Bündelung von Zuständigkeiten und in der Etablierung einer strukturierten Förderlandschaft gesehen. Allerdings werden auch verschiedenen „Optimierungsmöglichkeiten“ identifiziert. Unter anderem kommen spezielle Zielgruppenkurse und Kurse für Teilnehmende mit unterschiedlichen Voraussetzungen (zu) häufig nicht zustande. Die daraus resultierende Heterogenität der Teilnehmerstruktur verringere wiederum den Lernerfolg – allerdings lässt sich der Lernerfolg nicht eindeutig ermitteln, denn für eine derartige Evaluation fehlen Daten von verpflichtenden Abschlusstests bzw. Testverfahren und somit die zentralen Ergebnisvariablen. Der Integrationserfolg der Teilnehmenden lässt sich daher nur mit Einschätzungen der Kursträger und freiwilligen Abschlusstests abschätzen. Dies dürfte ein verzerrtes Bild ergeben.

Die Wirksamkeit und Nachhaltigkeit von Integrationskursen wird zu einem späteren Zeitpunkt auch in Schuller et al. 2011 untersucht. Zu diesem Zeitpunkt konnten bereits Verbesserungen vor dem Hintergrund der Anregungen der vorherigen Evaluationsergebnisse umgesetzt werden und die Untersuchung konnte auf Daten des so genannten „Integrationspanels“ zurückgreifen. In dieser repräsentativen Datenbasis werden Teilnehmende in insgesamt vier Wellen befragt. Zudem stehen Daten über Personen zu drei Befragungszeitpunkten zur Verfügung, die nicht an den Integrationskursen teilgenommen haben. Die Konstruktion und Rekrutierung dieser Kontrollgruppe zur Ermittlung der kausalen Effekte der Programmteilnahme stellte sich als große methodische Herausforderung dar, denn eine Teilnahme an den Integrationskursen ist laut Gesetz für alle Neuzuwanderer mit nicht ausreichenden Deutschkenntnissen verpflichtend. Insofern bleiben Zweifel bestehen, ob sich die Teilnehmenden von den Mitgliedern der Kontrollgruppe tatsächlich nur in der Programmteilnahme unterscheiden, nicht aber in weiteren beobachteten und unbeobachteten Eigenschaften. Die Autoren konstatieren selbst eine eingeschränkte Vergleichbarkeit der beiden Gruppen, die sie unter anderem mit Matching-Verfahren adressieren. Vorbehaltlich dieser Einschränkungen zeigen die Wirkungsanalysen, dass sich die Teilnahme an den Integrationskursen gleich auf eine ganze Reihe von Integrationsaspekten nachhaltig positiv auswirkt. Diese Integrationsaspekte beinhalten unter anderem Deutschkenntnisse, Arbeitsmarktpartizipation, Kontakte zu Deutschen und Verbundenheitsgefühle mit Deutschland.

Walter et al. (2014) haben das Bundesprogramm zur berufsbezogenen Sprachförderung von Migrantinnen und Migranten evaluiert. Im Rahmen dieses Programms wurde die Förderung von Deutsch als Zweitsprache für möglichst alle Erwachsenen mit Migrationshintergrund und Sprachförderbedarf an einer Stelle gebündelt. Es sollte ein kombiniertes Angebot geschaffen werden, das neben der Förderung der berufsbezogenen Deutschkenntnisse auch fachliche Kompetenzen vermitteln sollte, um bestehende Qualifikationshemmnisse zu überwinden und insgesamt die Chancen einer Integration in den ersten Arbeitsmarkt oder in Ausbildung zu verbessern. Auch vor diesem Hintergrund ist die Integration in Fort- und Weiterbildungsmaßnahmen bedeutsam. Neben dem Sprachunterricht waren auch Qualifizierungselemente (Fachunterricht, Praktikum, Betriebsbesichtigungen) Bestandteil der Kurse. Vor Kursbeginn findet zur Ermittlung des individuellen Sprachförder- und Qualifizierungsbedarfs eine Kompetenzfeststellung bei den Teilnehmenden statt. Die Kurse werden grundsätzlich mit einer Abschlussprüfung beendet. Neben einer eingehenden Analyse der Struktur der Teilnehmenden und einer Implementationsanalyse führen Walter et al. (2014) auch eine Wirkungsmessung dieses Programms durch. Zur Identifikation der kausalen Effekte der Teilnahme nutzen sie Sozialdaten der BA und wenden Matching-Verfahren an – mit der großen methodischen Herausforderung, eine adäquate Kontrollgruppe zu konstruieren. Denn die verwendeten Daten sind zwar grundsätzlich reich an Informationen und beinhalten eine große Zahl an Individuen, aber es sind darin keine Informationen zur zentralen Voraussetzung der Teilnahme – dem Sprachförderbedarf – für potenzielle Kontrollpersonen enthalten. Deshalb musste eine recht starke Annahme getroffen werden: Es wurde angenommen, dass eine potenzielle Kontrollperson im Durchschnitt über einen vergleichbaren Sprachförderbedarf verfügt, wenn sie auch in anderen soziodemografischen und erwerbsbiografischen Eigenschaften mit den Teilnehmenden vergleichbar ist. Mit dieser Vorgehensweise ermitteln die Autoren als Ergebnis ihrer Wirkungsanalyse innerhalb des recht kurzen Beobachtungszeitraums von 18 Monaten nach Programmbeginn einen tendenziell negativen bzw. gegen Null tendierenden Beschäftigungseffekt.

## 3. Vorhandene registrierte Daten

### 3.1 Einleitung

Die Mindestvoraussetzung für die Machbarkeit einer Evaluation der Integrationsmaßnahmen für Flüchtlinge ist der Zugang zu administrativen Individualdaten, die mindestens die Information enthalten, wer an den zu evaluierenden Leistungen teilnimmt und wer nicht, sowie brauchbare Kontaktdaten, um für den Zweck der Evaluation fehlende Informationen per Befragung ergänzen zu können. Ohne eine erste Information zur Maßnahmenteilnahme müsste eine aufwändige Screening-Erhebung durchgeführt werden, um überhaupt erst die Zielgruppe der Untersuchung einzugrenzen. Auf diese Weise ließe sich das Evaluationsvorhaben kaum wirtschaftlich durchführen. Bei fehlenden Kontaktdaten müsste auf Adressdatenbanken ausgewichen werden, die als Grundlage für Bevölkerungsbefragungen zwar durchaus zur Verfügung stehen, die Grundgesamtheit der in jüngster Zeit nach Deutschland neu hinzugezogenen Flüchtlinge bislang aber nicht adäquat abbilden.

Aus diesem Grund musste in einem vorbereitenden Arbeitsschritt zur Konkretisierung des Evaluationskonzepts zunächst die Verfügbarkeit – und Zugänglichkeit – von für den Evaluationszweck geeigneten, amtlich erhobenen Individualdaten geprüft werden. Dabei wurden die Geschäftsstatistiken der BA (Daten der Agenturen für Arbeit und der Jobcenter) und des BAMF als den beiden mit der Administration der Asylsuchenden bzw. arbeitssuchenden Flüchtlingen zentral befassten Institutionen in den Blick genommen. Dieses Kapitel fasst die im Hinblick auf die amtlichen Datengrundlagen einer Evaluation der Integrationsmaßnahmen wesentlichen Erkenntnisse und Schlussfolgerungen zusammen.

### 3.2 Daten zu Personen im SGB II und III in der Statistik der BA

Lange Zeit war eine belastbare Abbildung von Flüchtlingen in den Statistiken der BA nicht möglich, da Angaben zum aufenthaltsrechtlichen Status von Ausländerinnen und Ausländern in den operativen Verfahren nur mangelhaft erfasst wurden. Dieser Zustand hat sich in jüngster Zeit grundlegend verändert, so dass die Qualität der bei der Statistik der BA zu den Flüchtlingen verfügbaren Individualdaten im Hinblick auf das Evaluationsvorhaben als insgesamt tragfähig und günstig beurteilt werden kann.

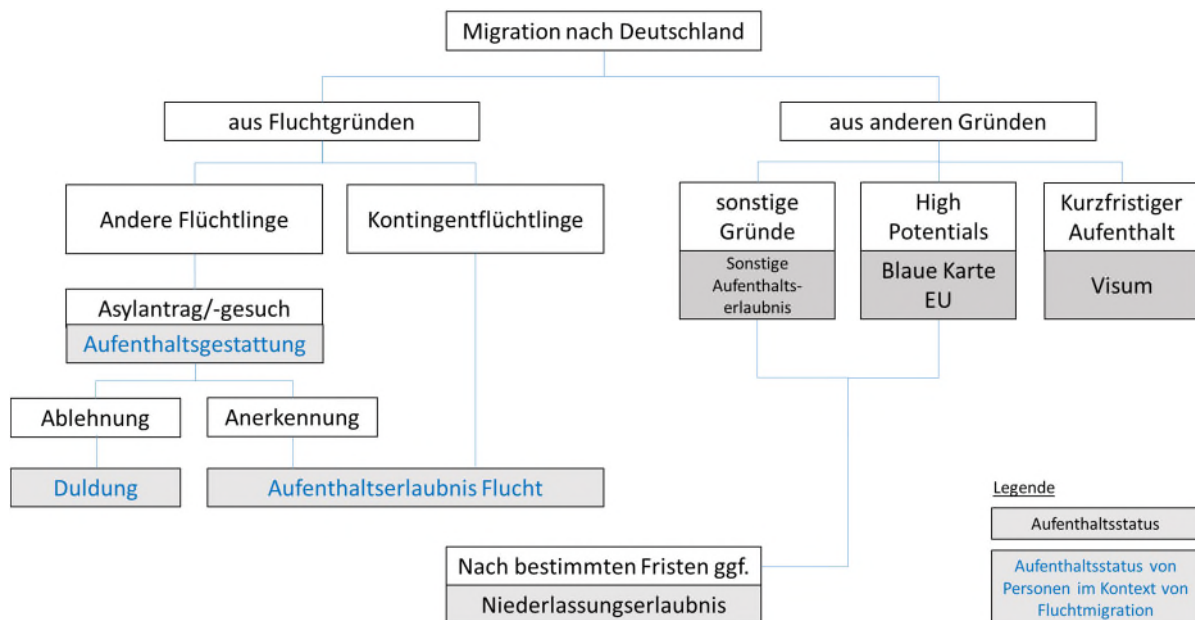
Der Grund hierfür liegt in einer grundlegenden Überarbeitung der Erfassung des Aufenthaltsstatus von Drittstaaten-Angehörigen durch die Arbeitsagenturen und Jobcenter und der auf dieser Information aufbauenden statistischen Verarbeitung. Als Drittstaaten-Angehörige zählen alle Ausländerinnen und Ausländer, die weder Staatsangehörige eines Mitgliedsstaats der Europäischen Union, Islands, Liechtensteins, Norwegens oder der Schweiz sind oder Staatenlose sind, und die für ihren Aufenthalt in Deutschland einen Aufenthaltstitel benötigen. Somit sind praktisch alle Flüchtlinge in der Gruppe der Drittstaaten-Angehörigen erfasst.

Seit Jahresbeginn 2016 mussten die Arbeitsagenturen und Jobcenter alle von der Veränderung betroffenen Datensätze neu erfassen. Zur Jahresmitte waren die Datenfelder zu 94 Prozent gefüllt. Selbst wenn man gewisse Datenunvollkommenheiten in Folge nicht auszuschließender Umstellungs- und Anfangsfehler in Rechnung stellt, liegt seitdem eine Datenbasis vor, die nach Einschätzung der Statistik der BA hinreichend belastbar ist. Die regelmäßige Berichterstattung über Personen im Kontext von Fluchtmigration, die bei Arbeitsagenturen und Jobcentern gemeldet sind, hat mit dem Berichtsmonat Juni 2016 begonnen. Der „Migrations-Monitor Arbeitsmarkt: Personen im Kontext der Fluchtmigration“, der ausgewählte Strukturinformationen zu Arbeitsuchenden, Arbeitslosen und erwerbsfähigen Leistungsberechtigten (ELB) für Deutschland, die Länder, sowie Agentur- und Jobcenterbezirke enthält und auch über den Bestand an

Teilnehmenden in ausgewählten arbeitsmarktpolitischen Maßnahmen informiert, wird monatlich aktualisiert.

Unter dem Begriff „Personen im Kontext der Fluchtmigration“ fasst die Statistik der BA Drittstaaten-Angehörige zusammen, die sich in Deutschland aufgrund einer Aufenthaltsgestattung als Asylbewerber, aufgrund einer „Aufenthaltserlaubnis Flucht“ als anerkannt Schutzberechtigte (der Begriff umfasst anerkannte Asylbewerber, anerkannte Flüchtlinge und subsidiär Schutzberechtigte) oder aufgrund einer Duldung als abgelehnte Asylbewerber aufhalten. Diese Personengruppen lassen sich anhand der Kategorien des Aufenthaltsstatus identifizieren, wie sie in VerBIS bzw. in der Datenlieferung über XSozial-BA-SGB II nun für alle Drittstaaten-Angehörigen abgelegt werden.<sup>2</sup> Abbildung 3.1 zeigt in stark vereinfachter Form das dabei von der Statistik der BA für die Zwecke der statistischen Berichterstattung zugrunde gelegte Schema.

**Abbildung 3.1** Schema der von der Statistik der BA für Drittstaaten-Angehörige dargestellten Arten des Aufenthaltsstatus in Deutschland



Quelle: Eigene Darstellung in Anlehnung an Statistik der BA (2016).

Erläuterungen: Drittstaaten-Angehörige sind Personen, die weder Staatsangehörige eines Mitgliedstaates der Europäischen Union, des Europäischen Wirtschaftsraums (EU zzgl. Island, Liechtenstein, Norwegen) oder der Schweiz sind, noch Staatenlose.

Mit dem Konzept der Personen im Kontext von Fluchtmigration kommt eine spezifische Abgrenzung der Gruppe der Flüchtlinge bzw. Geflüchteten zustande, die nicht unbedingt mit anderen Abgrenzungen oder Definitionen dieser Zielgruppe deckungsgleich ist. Die von der Statistik der BA verwendete Definition orientiert sich stark an der Beteiligung im Asylverfahren – unabhängig von der Frage, ob es zu einer Anerkennung kommt oder welcher Anerkennungsgrund zum Schutz führt. Diese Vorgehensweise bringt

<sup>2</sup> Mittelfristig ist in der BA geplant, die Information zum Aufenthaltsstatus in den Personen-Stammdaten über STEP zu hinterlegen. Das seit mehreren Jahren betriebene Stammdaten-Entwicklungsprojekt STEP soll allgemein die Datenstrukturen im Bereich der Personen- und Betriebe-Stammdaten sowie der Stammattribute aller Fachverfahren konsolidieren und harmonisieren. Auf die Datenlieferungen der zugelassenen kommunalen Träger (zkt) über XSozial-BA-SGB II hat dies keine Auswirkungen.

gewisse Einschränkungen mit sich. Insbesondere fallen anerkannte Schutzsuchende, die eine Niederlassungserlaubnis haben oder eingebürgert werden, aus der Gruppe der Personen im Kontext von Fluchtmigration heraus. Wichtiger noch erscheint, dass Angehörige von anerkannten Schutzsuchenden, die über den Familiennachzug Aufenthalt in Deutschland erhalten, nicht berücksichtigt werden, da sie keine Aufenthaltserlaubnis Flucht, sondern einen anderen Aufenthaltstitel bekommen. Dies könnte perspektivisch problematisch werden, falls die Zahl der nachziehenden Familienangehörigen zunimmt und diese Zugang zum Arbeits- oder Ausbildungsmarkt erhalten.

Trotz dieser leichten Einschränkungen empfiehlt es sich, für den Zweck des Evaluationsvorhabens das Konzept der Personen im Kontext von Fluchtmigration zur Abgrenzung der Zielgruppe von der Statistik der BA zu übernehmen. Eine solche Festlegung hat den wichtigen Vorteil, dass in der Evaluation auf Arbeitsmarktstatistiken der BA zu „Flüchtlingen“ Bezug genommen werden kann, ohne dass Kompatibilitätsprobleme entstehen.

Durch die Erfassung des individuellen Aufenthaltsstatus in VerBIS bzw. in der Datenlieferung über XSozial-BA-SGB II können auch die daran anknüpfenden Daten in der Statistik der BA spezifisch für die Zielgruppe der Flüchtlinge betrachtet werden. Hierzu zählen insbesondere die für die Durchführung des Evaluationsvorhabens zentralen Informationen zur Teilnahme an Maßnahmen der aktiven Arbeitsmarktförderung nach SGB III sowie zur Teilnahme an Maßnahmen aus dem Rechtskreis SGB II, soweit diese über die IT-Verfahren der BA erfasst werden. Im Bereich der zugelassenen Träger (zkt) wird die Maßnahmenteilnahme ebenfalls erfasst, allerdings teilweise in einer unterschiedlichen Tiefengliederung, was jedoch für die Evaluation im Regelfall ausreichend sein dürfte.

Weiterhin stehen, da die Flüchtlinge in der Statistik der BA völlig analog zu anderen Personen erfasst werden, alle auch sonst gängigen personenbezogenen Merkmale zur Verfügung, darunter etwa das Geschlecht, das Alter, die Schulbildung, die Berufsausbildung und das Anforderungsniveau des Zielberufs. Dagegen ist nicht gesichert, dass Angaben zur Muttersprache vorliegen – eine Information, die eine effiziente Ansprache im Rahmen einer ergänzenden persönlichen Befragung der Zielpersonen erleichtern könnte. Zwar können sehr gute Fremdsprachenkenntnisse beispielsweise von den Vermittlungsfachkräften in VerBIS durch die Auswahlmöglichkeit „verhandlungssicher“ erfasst werden; hierbei handelt es sich jedoch lediglich um eine „Kann-Möglichkeit“, die in den statistischen Daten nicht ausgewertet wird. Die in den statistischen Daten enthaltene Information zur Staatsbürgerschaft liefert zur Muttersprache angesichts der Sprachenvielfalt in vielen der Herkunftsländer ebenfalls nur näherungsweise Anhaltspunkte. Ob und wie gute Kenntnisse der deutschen Sprache vorhanden sind, wird in den administrativen Daten nur unsystematisch in Fällen erfasst, in denen die Information für die Arbeit der zuständigen Vermittlerin oder des zuständigen Vermittlers relevant ist – und auch dann nur grob. Informationen zum Sprachstand, der ein wesentlicher Faktor für die Arbeitsmarkt- und Ausbildungschancen der Flüchtlinge sein dürfte, müssten also aus anderer Quelle gewonnen werden.

Im Rahmen des Vermittlungsgesprächs werden Angaben zum bisherigen Erwerbsverlauf abgefragt und in VerBIS nach Werdegangsarten – Beschäftigung, Selbständigkeit, Elternzeit, Berufserfahrungen und Bildung, Erwerbstätigkeit im Ausland während der letzten sieben Jahre – gegliedert abgelegt. Damit lassen sich gewisse Informationen gewinnen, für die keine anderen Prozessdaten vorliegen, etwa zu einer früheren Beschäftigung im Herkunftsland. Solche Informationen hat das IT- und Informationsmanagement am IAB 2015 mit der Entwicklung der sogenannten Werdegangshistorie aus VerBis (WGH) als Zusatzbaustein zur IEB grundsätzlich erschlossen. Allerdings lässt sich die Qualität der Daten der WGH bislang noch nicht verlässlich abschätzen. Dies gilt für die WGH generell. Darüber hinaus lassen sich spezifische Qualitätsprobleme bei der retrospektiven Erfassung des Werdegangs im Rahmen der Vermittlungsgespräche mit arbeitslosen Personen im Kontext von Fluchtmigration nicht ausschließen. Trotzdem besteht hier ein Potenzial, aus den registrierten Daten zusätzliche Informationen zu gewinnen,

die infolge ihrer Relevanz für die Vermittlung ebenfalls für die Zuweisung in integrationspolitische Maßnahmen und deren Wirksamkeit systematisch relevant sein dürften. Von daher sollten die Informationen aus der WGH als Ergänzungsbaustein zur IEB bei der Durchführung der Evaluation ebenfalls Beachtung finden, dabei allerdings wegen der ungeklärten Qualität aber mit der erforderlichen Vorsicht behandelt werden.

In den Daten der Statistik der BA sind auch aktuelle Kontaktinformationen abgelegt, die grundsätzlich eine individuelle Ansprache ermöglichen sollten. Wie gut sich die vorhandenen Kontaktdaten in der Praxis für die Durchführung einer Befragung eignen, lässt sich allerdings nicht mit Bestimmtheit sagen, da hierzu für die Zielgruppe der Flüchtlinge keine Erfahrungswerte vorliegen. Es sollte jedoch mit erhöhten Antwortausfällen gerechnet werden, weil die Zielgruppe mobiler ist und auch ihre Telefonnummern häufiger wechseln dürfte als andere Personengruppen im SGB II und III. Besondere Schwierigkeiten mit der (Wieder-)Erreichbarkeit sind an den späteren Messzeitpunkten im Verlauf einer Panelerhebung einzuplanen. Für diejenigen, die bereits in den Arbeitsmarkt integriert sind und daher keine aktuellen Kundinnen und Kunden der Agenturen der Arbeit oder der Jobcenter mehr sind, werden die in den Daten der BA hinterlegten Kontaktdaten nicht aktualisiert. Daher könnte es zu im Hinblick auf die Zielgrößen systematischen Ausfällen im Verlauf der Wiederholungsbefragung kommen, in Folge derer die Effekte der Maßnahme unterschätzt werden. Um solchen Panelausfällen vorzubeugen, muss die Anfangsstichprobe der nicht-teilnehmenden Personen im Kontext von Fluchtmigration deutlich größer dimensioniert werden als bei den teilnehmenden Personen.

In der Beschäftigtenstatistik, die sich aus den Arbeitgebermeldungen zur Sozialversicherung speist, werden Flüchtlinge auch in Zukunft nicht identifizierbar sein, da hier wie bisher keine Daten zum Aufenthaltsstatus erfasst werden. Hieraus ergeben sich gewisse Einschränkungen. Bei der Verknüpfung der Informationen aus den operativen Systemen der Arbeitsagenturen und der Jobcenter mit den Informationen der Beschäftigungsstatistik in der Statistik der BA zu Integrierten Erwerbsbiografien (IEB) lässt sich zwar die Information zum Aufenthaltsstatus bewahren. Diese Information bleibt allerdings auf dem letzten in den operativen Systemen erfassten Stand eingefroren. Hieraus ergeben sich für das Evaluationsvorhaben aber keine wesentlichen Probleme, weil hierfür der Status als Flüchtling beim Eintritt in die zu untersuchenden Maßnahmen zentral ist. Nicht adäquat untersuchen ließen sich etwa Veränderungen des Aufenthaltsstatus als Zielgröße (etwa als Indikator für Integration), oder in anderer Richtung die Einflüsse von Veränderungen beim Aufenthaltsstatus nach Aufnahme einer Beschäftigung auf die – langfristige – Maßnahmenwirksamkeit. Solche Aspekte gehören aber nicht zum Kern des Erkenntnisinteresses.

Durch Rückgriff auf Integrierte Erwerbsbiografien lässt sich das ganze Spektrum an Zielindikatoren für die Bewertung der Arbeitmarkteffekte der Integrationsmaßnahmen für Flüchtlinge analysieren, wie es im Kontext der Evaluation aktiver arbeitsmarktpolitischer Maßnahmen gängig ist (vgl. Abschnitt 5.2).

Der Zugriff auf die zur Durchführung des Evaluationsvorhabens notwendigen Personendaten aus dem SGB II und III sowie der darauf aufbauenden Integrierten Erwerbsbiografien könnte gemäß den für Evaluationen arbeitsmarktpolitischer Maßnahmen mittlerweile gut etablierten Routinen erfolgen, also auf Antrag über das Datenservicezentrum am IAB und unter Einhaltung der Gebote des Datenschutzes und der Datensparsamkeit. Da etablierten Routinen gefolgt wird, sollten die im ersten Schritt erforderlichen Daten relativ schnell nach Start des Evaluationsprozesses – günstigenfalls innerhalb von 2-3 Monaten – verfügbar sein.

### 3.3 *Geschäftsdaten des BAMF*

Die Geschäftsdaten des BAMF können anders als die Daten im SGB II und III aus Gründen des Datenschutzes nicht für die Durchführung des geplanten Evaluationsvorhabens verwendet werden. Damit stehen insbesondere keine personenbezogenen Verlaufsdaten zur Verfügung, um Ereignisse zwischen der Ankunft der Flüchtlinge in Deutschland und ihrem Übergang in den Zuständigkeitsbereich der Rechtskreise des SGB II und III zu erfassen, die für die Wirksamkeit der zu evaluierenden Integrationsmaßnahmen bedeutsam sein können. Dies betrifft etwa Details zum Verlauf und zur Dauer des Asylverfahrens. Ebenfalls nicht aus BAMF-Quellen zu beziehen sind personenbezogene Daten zur Teilnahme an den in die Zuständigkeit des BAMF fallenden Integrations- und Sprachfördermaßnahmen. Diese Informationen können allenfalls in den IEB erscheinen, da in den Daten der Statistik der BA auch Fremdförderungen mit Beginn- und Enddatum ausgewiesen werden. Diese Angaben sind aber möglicherweise ungenau und unvollständig, insbesondere was BAMF-Förderungen angeht, die vor dem Übergang der Personen im Kontext von Fluchtmigration in den Zuständigkeitsbereich der Agenturen für Arbeit bzw. der Jobcenter stattfinden.

Offenbar werden im Rahmen von Vermittlungsgesprächen im SGB II und III teilweise Daten erhoben, welche sich auf die Vorgeschichte der Flüchtlinge im Zuständigkeitsbereich des BAMF beziehen. Diese Informationen werden jedoch lediglich als Freitext abgelegt und stehen deshalb für Auswertungen nicht zur Verfügung. Angesichts dessen ist bis auf Weiteres nicht damit zu rechnen, dass sich Datenlücken, die sich in Folge des fehlenden Zugangs zu den beim BAMF registrierten Individualdaten ergeben, verlässlich über Informationen auffüllen lassen, die in den administrativen Daten im SGB II und III erfasst werden. Größen, die für die Wirksamkeit der zu evaluierenden Integrationsleistungen relevant sein könnten, beispielsweise die zeitliche Distanz zwischen Einreise und Asylerstantragstellung, die Dauer des Asylverfahrens oder die Inanspruchnahme von Integrationsleistungen des BAMF, müssten deshalb bei den Flüchtlingen selbst nacherhoben werden.

### 3.4 *Schnittstelle zwischen BAMF-Daten und Daten zu Personen im SGB II und III*

Nach Schaffung der dazu erforderlichen Voraussetzungen durch das Datenaustauschverbesserungsgesetz vom 2. Februar 2016 hat die BA daran gearbeitet, die Erfassung der Personendatensätze von Flüchtlingen durch funktionale Erweiterungen in den IT-Verfahren zu vereinfachen. Dazu wurde zur Stammdatenanlage in STEP die Möglichkeit zur Anzeige von Daten aus einem IT-Verfahren des BAMF anhand einer ID geschaffen. Bei dieser ID handelt es sich um die Nummer aus dem Ausländerzentralregister (AZR-Nummer), die zentral bei der erstmaligen Registrierung jeder Ausländerin und jedes Ausländers vergeben wird und einen Personendatensatz bundesweit eindeutig identifiziert. Inwiefern auch bei den zugelassenen kommunalen Trägern Datenschnittstellen zum BAMF hergestellt werden (sollen), ist nicht bekannt.

Flüchtlinge, deren Daten noch nicht in STEP erfasst sind, sollen anhand dieser AZR-Nummer gesucht werden. Auf diese Weise kann gewährleistet werden, dass es sich bei dem Trefferergebnis um die ausweisliche Person handelt. Die bereits beim BAMF erfassten Personenstammdaten (Name, Geschlecht, Geburtsdatum und -ort, Staatsangehörigkeit, Titel und Familienstand) werden in STEP automatisch eingeblendet und können nach Überprüfung der Angaben im Ausweisdokument der Kundin oder des



Kunden sofort gespeichert werden. Die gespeicherten Angaben werden automatisiert in VerBIS übergeben.<sup>3</sup>

Soweit dieser Mechanismus angewendet wird, was in der Vermittlungspraxis allerdings die Ausnahme und nicht die Regel ist, führt er dazu, dass die bei Arbeitsagenturen oder Jobcentern in gemeinsamer Trägerschaft und BAMF abgelegten Personenstammdaten exakt übereinstimmen. Somit besteht eine eindeutige Verbindung zwischen den beiden Datensphären. Es existiert also zumindest technisch die Möglichkeit, zu jedem Flüchtling, für den Stammdaten bei den Arbeitsagenturen oder Jobcentern in gemeinsamer Trägerschaft angelegt wurden, Daten aus dem AZR zuzuspielen. Dies wäre eine ressourcensparsame Vorgehensweise, um im Rahmen einer Wiederholungsbefragung aktuelles und verlässliches Adressmaterial von Flüchtlingen zu erhalten. Derzeit dürfte die Umsetzung dieser wünschenswerten Lösung allerdings an den grundsätzlichen Vorgaben des Datenschutzes scheitern. Das BMAS sollte allerdings prüfen, ob eine solche Verknüpfung vorbehaltlich der ausdrücklichen Einwilligung der Betroffenen möglich wäre. Diese Zustimmung ließe sich im Rahmen einer anhand von Adressbeständen der Agenturen für Arbeit oder Jobcenter realisierten Erstbefragung von Flüchtlingen einholen, so dass die Stammdaten-Schnittstelle zum AZR zur Gewinnung des Adressmaterials für spätere Wellen einer Panelbefragung genutzt werden könnte.

Dass die Datenbestände des BAMF grundsätzlich für die Legung eines Feldzugangs zum Zweck einer persönlichen Befragung geeignet wären, belegt das Beispiel der laufenden IAB-BAMF-SOEP-Befragung von Flüchtlingen, für die die Stichprobenauswahl auf Basis des AZR getroffen wurde. Eine entsprechende Nutzung der AZR-Daten setzt eine Forschungs Kooperation mit dem Forschungszentrum des BAMF und darüber hinaus die Zustimmung des Bundesministeriums des Innern voraus.

### 3.5 Fazit

Die Lage, was die administrativen Datengrundlagen des geplanten Evaluationsvorhabens angeht, erweist sich als relativ günstig: Durch die initiierten Maßnahmen zur Verbesserung der statistischen Erfassung des Aufenthaltsstatus bei Drittstaaten-Angehörigen lassen sich die Flüchtlinge in geeigneter Weise in den Daten der Statistik der BA eingrenzen. Somit bewegt sich das Evaluationsvorhaben von der Datenbasis her gesehen weitestgehend auf vertrautem Terrain. Für einen zügigen Datenzugang dürfte sorgen, dass es hierfür lang und gut etablierte Routinen gibt.

Trotz der guten Verfügbarkeit administrativer Daten wird dieses Evaluationsvorhaben im Vergleich zu anderen Studien zur Wirkung arbeitsmarktpolitischer Maßnahmen stark auf ergänzende Befragungsdaten angewiesen sein. Dies liegt erstens daran, dass der Informationsgehalt der administrativen Daten naturgemäß begrenzt ist, weil die Flüchtlinge erst seit kurzem in Deutschland leben. Zweitens gestalten sich die Prozesse des Zugangs zu den Integrationsmaßnahmen im Bereich der Grundsicherung für Arbeitsuchende und der Arbeitsförderung komplex, denn parallel dazu existiert eine Vielzahl komplementärer Hilfs- und Unterstützungsangebote, die in den Daten nicht dokumentiert sind. Beide Informationsdefizite müssen durch Erhebungen ausgeglichen werden. Dabei sind sowohl der Feldzugang zu den Flüchtlingen als auch ihre Befragung in Folge ihrer ethnischen Diversität mit besonderem Aufwand verbunden.

---

<sup>3</sup> Darüber hinaus werden bei der Abfrage in STEP aus dem BAMF-System Angaben zur Schulbildung und zur Branche der letzten Tätigkeit eingeblendet; diese Informationen sind jedoch schlecht gefüllt und unzuverlässig, so dass sie nicht automatisch gespeichert werden. Die entsprechenden Felder dürften in der Praxis normalerweise durch die Vermittlerin oder den Vermittler mit präziseren Informationen gefüllt werden.



## 4. Leistungen zur Evaluation

Als Teil der vorbereitenden konzeptionellen Arbeiten an der Erstellung eines machbaren und wirtschaftlichen Konzepts zur Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen des Bundes für Flüchtlinge galt es zu klären, welche Leistungen im Rahmen des geplanten Evaluationsvorhabens spezifisch untersucht werden sollen. Wie im vorangegangenen Kapitel erläutert, ist eine valide statistische Abbildung von Flüchtlingen in den Arbeitsmarktstatistiken der BA inzwischen möglich. Auf dieser Basis veröffentlicht die BA-Statistik inzwischen aggregierte Ergebnisse zu arbeitsuchenden und arbeitslosen Personen im Kontext von Fluchtmigration, das heißt, zu denjenigen Drittstaaten-Angehörigen (Personen, die nicht Staatsangehörige eines Mitgliedstaates der Europäischen Union, Islands, Liechtensteins, Norwegens oder der Schweiz sind oder Staatenlose), die über eine Aufenthaltsgestattung, eine Aufenthaltserlaubnis Flucht oder eine Duldung verfügen.<sup>4</sup>

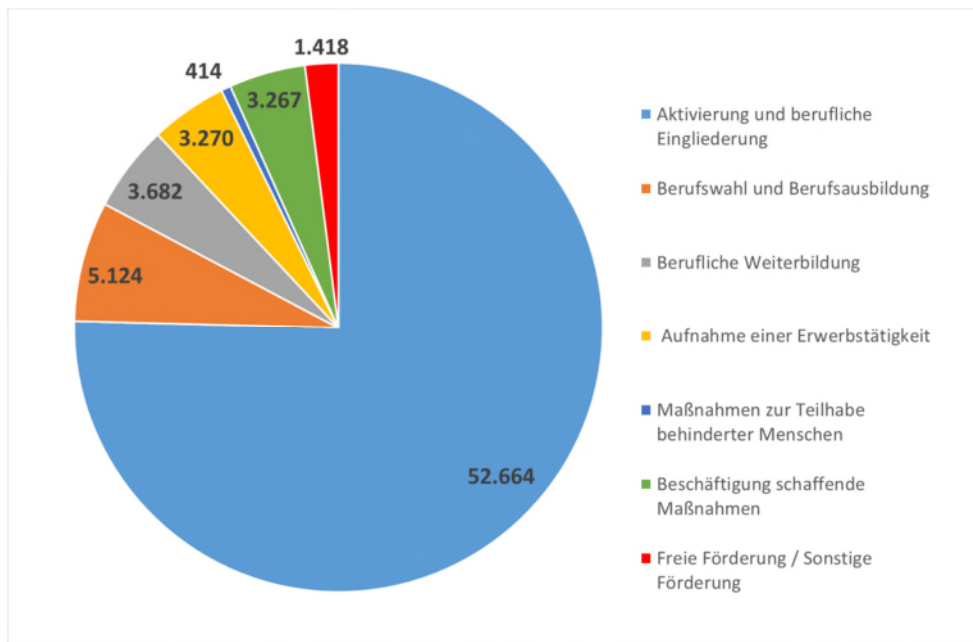
Die Förderstatistik der BA erfasst dabei die Förderungen bzw. Teilnahmen von Personen an Maßnahmen der aktiven Arbeitsförderung (§ 3 SGB III Abs. 2) und die Leistungen zur Eingliederung (§ 16 SGB II) des Bundes. Zudem besteht grundsätzlich die Möglichkeit, auch Fremdförderungen zu erfassen, worunter beispielsweise auch Integrationskurse im Bereich der Zuständigkeit des BAMF fallen; diese Sammelkategorie dürfte aber nicht vollständig gefüllt sein. Eine Förderung, die im Rahmen der Standardberichterstattung der Förderstatistik nachgewiesen wird, liegt vor, wenn für eine Person bzw. im Rahmen der Teilnahme an einer Maßnahme der aktiven Arbeitsförderung eine Zahlung geleistet wird. Die Kategorien der Instrumente, über deren Einsatz sich der Nachweis in der Förderstatistik erstreckt, sind die Leistungen zur Aktivierung und beruflichen Eingliederung, Maßnahmen zur Unterstützung bei Berufswahl und Berufsausbildung, die Förderung der beruflichen Weiterbildung, der Einsatz beschäftigungsbegleitender Leistungen bei Aufnahme einer Erwerbstätigkeit, die Maßnahmen zur Teilhabe behinderter Menschen, Beschäftigung schaffende Maßnahmen, sowie die Nutzung vorhandener freier Fördermöglichkeiten.

Wie Abbildung 4.1 zeigt, kommen sämtliche dieser Leistungen und Instrumente in der Zielgruppe der Personen im Kontext von Fluchtmigration – im Folgenden zur besseren Lesbarkeit auch kurz als Flüchtlinge bezeichnet – zum Einsatz. Im Berichtsmonat September 2016 wurden insgesamt 69.839 Flüchtlinge in arbeitsmarktpolitischen Instrumenten beobachtet. Damit wurde von den 210.617 nicht arbeitslosen, aber arbeitsuchenden Flüchtlingen ein Drittel (33,2 Prozent) durch arbeitsmarktpolitische Leistungen und Instrumente erreicht.

---

<sup>4</sup> Der Status als Arbeitssuchender oder als Arbeitsloser richtet sich nach den Kriterien des Sozialgesetzbuchs. Arbeitssuchend ist, wer eine Beschäftigung als Arbeitnehmer sucht. Arbeitslos ist, wer als Arbeitssuchender keine Beschäftigung hat, dem Arbeitsmarkt zur Verfügung steht und nicht an einer arbeitsmarktpolitischen Maßnahme teilnimmt. Die Zahl der Arbeitssuchenden ist also systematisch größer als die Zahl der Arbeitslosen.

**Abbildung 4.1 Bestand an Personen im Kontext von Fluchtmigration in arbeitsmarktpolitischen Instrumenten – Berichtsmonat: September 2016**



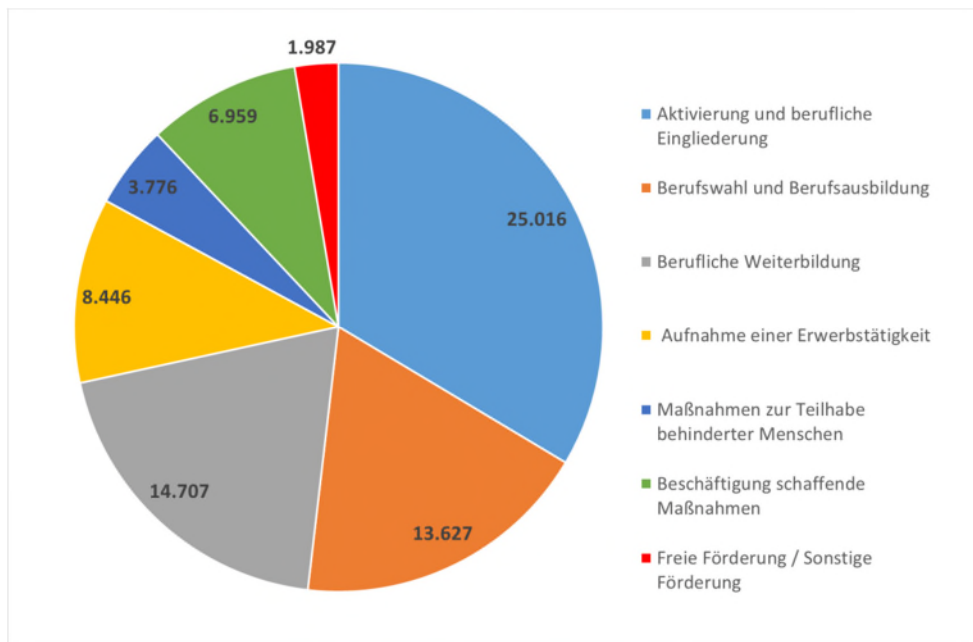
Darstellung des IZA auf Grundlage der Statistik der BA (2016).

Erläuterungen: Personen im Kontext von Fluchtmigration ist ein statistischer Begriff, für den über das Asylverfahren hinaus der Bezug zum Arbeitsmarkt ausschlaggebend ist. Zusammengefasst ist ein Personenkreis mit ähnlichen Problemlagen im Hinblick auf den Arbeits- und Ausbildungsmarkt. Er umfasst Ausländer mit einer Aufenthaltsgestattung, einer Aufenthaltserlaubnis Flucht oder einer Duldung. Personen, die im Rahmen eines Familiennachzugs (§§29ff AufenthG) zu geflüchteten Menschen nach Deutschland migrieren, zählen dagegen nicht dazu.

Im Vergleich mit anderen Bevölkerungsgruppen – dies veranschaulicht beispielhaft die Verteilung der sonstigen Drittstaaten-Angehörigen in arbeitsmarktpolitischen Maßnahmen auf die verschiedenen Bereiche der Förderung (Abbildung 4.2) – zeigt sich eine besondere Förderstruktur bei den Flüchtlingen. Das für diese Personengruppe weit überdurchschnittlich häufig eingesetzte arbeitsmarktpolitische Instrument sind Maßnahmen zur Aktivierung und beruflichen Eingliederung nach § 16 SGB II in Verbindung mit § 45 SGB III. Hierauf entfielen im Berichtsmonat September 2016 mehr als Dreiviertel (75,4%) der von der BA registrierten Förderfälle von Flüchtlingen. Bei den übrigen Drittstaaten-Angehörigen betrug dieser Anteil nur 33,6%, und bei den nicht aus Drittstaaten kommenden Personen sogar nur 19,9%.

Bei den Maßnahmen zur Aktivierung und beruflichen Eingliederung kann es sich beispielsweise um Bewerbungstraining, Kompetenzfeststellung oder Vermittlungsunterstützung handeln, bei Flüchtlingen mit Arbeitsmarktzugang – für die Asylbewerberinnen und Asylbewerber aus Eritrea, Irak, Iran, Somalia und Syrien gemäß der Sonderregel des § 131 SGB III befristet bis zum 31. Dezember 2018 auch schon vor Ende der Wartefrist – auch um eine berufliche Kenntnisvermittlung bis zu acht Wochen, insbesondere im Rahmen von Praktika. Die Maßnahmen zur Aktivierung und beruflichen Eingliederung bei einem Arbeitgeber, die auch mit Maßnahmen zur berufsbezogenen Sprachförderung kombiniert werden können, spielen allerdings mit 4,5% der Fälle in diesem Förderbereich bislang nur eine untergeordnete Rolle.

**Abbildung 4.2 Bestand an Drittstaaten-Angehörigen in arbeitsmarktpolitischen Instrumenten ohne Personen im Kontext von Fluchtmigration – Berichtsmonat: September 2016**



Darstellung des IZA auf Grundlage der Statistik der BA (2016).

Erläuterungen: Drittstaaten-Angehörige sind Personen, die weder Staatsangehörige eines Mitgliedstaates der Europäischen Union, des Europäischen Wirtschaftsraums (EU zzgl. Island, Liechtenstein, Norwegen) oder der Schweiz sind, noch Staatenlose. Personen im Kontext von Fluchtmigration ist ein statistischer Begriff, für den über das Asylverfahren hinaus der Bezug zum Arbeitsmarkt ausschlaggebend ist. Zusammengefasst ist ein Personenkreis mit ähnlichen Problemlagen im Hinblick auf den Arbeits- und Ausbildungsmarkt. Er umfasst Ausländer mit einer Aufenthaltsgestattung, einer Aufenthaltserlaubnis Flucht oder einer Duldung. Personen, die im Rahmen eines Familiennachzugs (§§29ff AufenthG) zu geflüchteten Menschen nach Deutschland migrieren, zählen dagegen nicht dazu.

Die verbleibenden Förderungen von Flüchtlingen verteilten sich im September 2016 relativ gleichmäßig auf vier Kategorien:

- die verschiedenen Leistungen zur Unterstützung der Berufswahl, wie Berufseinstiegsbegleitung, assistierte Ausbildung, berufsvorbereitende Bildungsmaßnahmen, Einstiegsqualifizierung und ausbildungsbegleitende Hilfen (Gesamtanteil an allen Maßnahmenteilnahmen: 7,3%);
- die Maßnahmen zur Förderung der beruflichen Weiterbildung, wie Anpassungsqualifizierung, die Vermittlung von Teilqualifikationen oder die Umschulung bei einem Träger, wobei für die Teilnahme bei Personen mit Arbeitsmarktzugang im Rechtskreis SGB II für eine erfolgreiche Bewältigung ausreichende Sprachkenntnisse (in der Regel B1 GER) Voraussetzung sind (Gesamtanteil an allen Maßnahmenteilnahmen: 5,3%);
- Beschäftigung schaffende Maßnahmen, wobei fast ausschließlich Arbeitsgelegenheiten, die Flüchtlinge bereits niedrigschwellig an den deutschen Arbeitsmarkt heranzuführen, zum Tragen kommen (Gesamtanteil an allen Maßnahmenteilnahmen: 4,7%); Teilnehmerinnen und Teilnehmer des am 1. August 2016 gestarteten Programms „Flüchtlingsintegrationsmaßnahmen (FIM)“ zur Schaffung von 100.000 Arbeitsgelegenheiten für Flüchtlinge vor Abschluss des Asylverfahrens tauchen in den Daten allerdings nicht auf, da es sich um eine Maßnahme nach dem Asylbewerberleistungsgesetz handelt, die von der BA lediglich administriert wird;

- die Maßnahmen zur Förderung der Erwerbstätigkeit, wobei es sich prinzipiell sowohl um abhängige als auch selbständige Beschäftigung handeln kann; mit 2.374 von 2.439 Förderfällen dominieren jedoch Leistungen zur Förderung der abhängigen Beschäftigung, und hierbei wiederum mit 1.943 Förderfällen das Instrument des Eingliederungszuschusses, das auf einen Minderleistungsausgleich für Arbeitgeber abzielt (Gesamtanteil an allen Maßnahmenteilnahmen: 4,7%).

Ziemlich kleine Anteile an Förderfällen entfallen derzeit auf die Sammelkategorie der Leistungen der freien Förderung (2,0%) und die Maßnahmen zur Teilhabe behinderter Menschen (0,6%). Letztere sind derzeit von der Anzahl der teilnehmenden Flüchtlinge bei weitem zu schwach besetzt, um belastbare quantitative Wirkungsanalysen durchzuführen. Es gilt daher weiter zu beobachten, ob die Anzahl der Teilnehmerinnen und Teilnehmerinnen noch eine für entsprechende Untersuchungen notwendige Mindestgröße (etwa ab 5.000 Fälle aufwärts) erreicht, wenn die Zahl der arbeitssuchenden Flüchtlinge, wovon angesichts der vergangenen Entwicklung der Asylantragstellerzahlen auszugehen ist, in nächster Zeit zunimmt. Dass dann genügend Fälle erreicht werden, ist aber keineswegs gewährleistet. Dies zeigt etwa der Blick auf den Bestand an Drittstaaten-Angehörigen in arbeitsmarktpolitischen Instrumenten ohne die Personen im Kontext von Fluchtmigration (Abbildung 4.2).

Sollte die Zahl der Flüchtlinge, die an Maßnahmen zur Teilhabe behinderter Menschen teilnehmen, zu gering bleiben, ist zu empfehlen, diese Kategorie aus dem geplanten umfassenden Evaluationsvorhaben herauszulösen und dafür ein eigenständiges, rein qualitativ orientiertes Forschungsvorhaben aufzulegen. Diese Vorgehensweise würde auch eine spezifische Ausrichtung der Untersuchung ermöglichen, wodurch den nochmals besonderen Problemlagen in der Gruppe von Flüchtlingen mit Behinderung, die eine doppelte Integrationsherausforderung zu bewältigen haben, in angemessener Weise Rechnung getragen werden könnte.

Unabhängig von der Entwicklung der Fallzahlen nicht empfohlen wird eine Betrachtung der Kategorie der Freien oder sonstigen Förderung, die unter den arbeitsmarktpolitischen Instrumenten allgemein von untergeordneter Bedeutung ist und es sich dabei zudem um eine Sammelkategorie handelt. Würde man die so zusammengefassten Maßnahmen als Ganzes evaluieren, wären die Resultate nicht sehr aussagekräftig. Der gemessene Maßnahmeneffekt wäre dann ein gewogener Durchschnitt der möglicherweise ganz unterschiedlichen Effekte der dahinter stehenden einzelnen Maßnahmen. Eine Evaluation der einzelnen Maßnahmen erscheint im Verhältnis zum Gewicht der Freien und sonstigen Förderung im Instrumentenkasten der BA jedoch unter wirtschaftlichen Gesichtspunkten nicht angemessen und ließe sich in Anbetracht kleiner Teilnahmezahlen praktisch ohnehin kaum angemessen realisieren.

Die übrigen Förderbereiche der arbeitsmarktpolitischen Regelleistungen sollten dagegen jeweils auf ihre spezifischen Wirkungen hin untersucht werden. Eine vergleichende Evaluation der Förderbereiche nach einem möglichst einheitlichen Standard kann Informationen liefern, um die künftige Allokation von Ressourcen auf die unterschiedlichen Instrumente unter Gesichtspunkten der Wirksamkeit und Effizienz zu justieren.

Bei der Evaluation der einzelnen Förderbereiche gilt es im Blick zu behalten, dass sie zum Teil sachlich unterschiedliche Maßnahmen zusammenfassen. So können die Maßnahmen zur Förderung der abhängigen Beschäftigung und zur Förderung der Selbständigkeit, die beide unter das Dach der Leistungen zur Aufnahme einer Erwerbstätigkeit fallen, ganz unterschiedliche Wirkungen entfalten. Nach Möglichkeit sollten die spezifischen Effekte für Unterkategorien von Instrumenten gemessen werden, die sich inhaltlich sinnvoll zueinander gruppieren lassen. Falls die Fallzahlen eine solche Differenzierung aber nicht erlauben, kann es sinnvoll sein, diese von vornherein nicht für eine quantitativ orientierte Evaluation vorzusehen, statt sie gemeinsam mit den anderen Leistungen eines Förderbereichs zu

bewerten. Aussagen zur Bedeutung und potenziellen Wirkung dieser Leistungen wären dann gegebenenfalls im Rahmen der qualitativ angelegten Begleitforschung zu gewinnen. Im erwähnten Beispiel müssten nach heutigem Stand angesichts der verschwindend geringen Teilnehmezahlen die Leistungen zur Aufnahme einer selbständigen Beschäftigung herausgenommen werden, so dass es bei einer Evaluation der Leistungen zur Aufnahme einer abhängigen Beschäftigung bliebe.

Aus diesen Überlegungen folgt, dass die Entwicklung der Struktur der teilnehmenden Flüchtlinge in den arbeitsmarktpolitischen Maßnahmen in der Zeit bis zum Start des geplanten Evaluationsvorhabens weiter beobachtet werden muss, um das Tableau der in das Evaluationsvorhaben einzubeziehenden Regelleistungen der BA im Detail gegebenenfalls noch zu justieren.

Zusätzlich zum Typ einer Leistung oder Maßnahme bildet ihre Qualität eine Differenzkategorie, der im Rahmen der geplanten Evaluation Aufmerksamkeit geschenkt werden sollte. So ist damit zu rechnen, dass ein bestimmtes arbeitsmarktpolitisches Instrument für Flüchtlinge durch die Akteure vor Ort in ganz unterschiedlicher Weise umgesetzt wird. Dies kann etwa die Inhalte der Maßnahme betreffen oder die Merkmale der für die Erbringung einer Leistung ausgewählten Träger. Leistungen können ganz spezifisch auf Flüchtlinge zugeschnitten sein, oder eine gemischte Gruppe von Teilnehmerinnen und Teilnehmern aufweisen.

Herauszufinden, wie solche Gestaltungsunterschiede die Wirksamkeit eines Leistungs- oder Maßnahmentyps systematisch beeinflussen, ist Teil der Evaluationsaufgabe. Hierzu müssen diese Unterschiede zunächst erfasst werden. Diese lassen sich teilweise aus der Förderstatistik der BA konstruieren. So ist es prinzipiell möglich, die Teilnehmerinnen und Teilnehmer, die an einer bestimmten Maßnahme gemeinsam teilnehmen, miteinander zu verknüpfen, so dass dem Einfluss von Teilnehmerstrukturen nachgegangen werden kann. Zum Teil müssten Informationen, die eine Beschreibung der Qualität der Umsetzung von Maßnahmen vor Ort ermöglichen, jedoch erst gewonnen werden, etwa durch eine Befragung der Jobcenter oder beauftragter Maßnahmenträger. Solche Informationen wären dann mit statistischen Verfahren zu einer Typologie unterschiedlicher Qualitäten einer Maßnahme zu verdichten.

Im Prinzip konstituiert jede Überkreuzung von Maßnahmentypen und damit verbundenen Qualitätstypen ein spezifisches Treatment. Es geht dann um die Bewertung von Interaktionseffekten. Lässt sich feststellen, dass die Wirksamkeit der Teilnahme an einer Maßnahme bei den Teilnehmerinnen und Teilnehmern systematisch besser oder schlechter ist, wenn sie mit einer bestimmten Qualität der Umsetzung verbunden ist? Für eine belastbare Identifikation von Interaktionseffekten müssen für die möglichen Überkreuzungen von Maßnahmen- und Qualitätstypen genügend Beobachtungen vorliegen. Aus Gründen der Wirtschaftlichkeit sollte sich das geplante Evaluationsvorhaben daher auf eine eher geringe Zahl von Qualitätsdimensionen fokussieren. Ihre Auswahl sollte sich an den spezifischen Erkenntnisinteressen des BMAS orientieren.

Die Wirksamkeit der Regelleistungen kann nicht nur durch deren spezifische Umsetzungsqualität, sondern auch durch die Interaktion mit anderen Leistungen zur Integration von Flüchtlingen verändert werden. Dieser Aspekt ist im vorliegenden Kontext von besonderer Bedeutung, da Länder und Kommunen, aber auch zahlreiche private Träger und ehrenamtliche Initiativen auf lokaler Ebene vielfältige Beratungs- und Hilfsangebote für Flüchtlinge machen. Diese Leistungen sind schwer überschaubar und sie werden auch nicht systematisch registriert. Um der Frage nachzugehen, inwiefern weitere Beratungs- und Hilfsangebote auf die Wirksamkeit der zu evaluierenden Regelleistungen einen systematischen Einfluss entfalten, müssen diese daher zunächst erfasst werden. Dazu werden umfassende Erhebungen der auf der kommunalen Ebene vorhandenen Angebote benötigt.

Weil sich auf dieser Ebene aber nicht klären lässt, welche Flüchtlinge genau von welchen komplementären Hilfs- und Beratungsleistungen erreicht werden, sollten darüber hinaus die Kenntnisse über sowie die Inanspruchnahme von Angeboten, die nicht zu den Regelleistungen zählen, bei den Flüchtlingen selbst erfasst werden. Dass die Flüchtlinge durchaus in der Lage sind, auf entsprechende Fragen Auskunft zu geben, zeigen erste Resultate der laufenden IAB-BAMF-SOEP-Befragung von Flüchtlingen (IAB 2016). In diesem Rahmen wurde nach der Bekanntheit und Inanspruchnahme verschiedener Beratungsstellen und -angebote für Geflüchtete gefragt, sowie nach der Inanspruchnahme von Hilfsangeboten für verschiedene Problemlagen und der allgemeinen Arbeitsmarkt- und Berufsberatung durch Arbeitsagenturen und Jobcenter. Diese Informationen ermöglichen zumindest eine gewisse Differenzierung der Wahrnehmung von Unterstützungsangeboten durch Flüchtlinge, wobei allerdings potenziell relevante Aspekte zu Trägerschaft und zur Angebotsqualität auf diesem Weg nur sehr bedingt zu erfassen sind.

Soweit sich die neben den Regelleistungen der Arbeitsagenturen und Jobcentern herlaufenden Unterstützungsangebote für die Flüchtlinge auf diese Weise typisieren lassen, lassen sich, wie oben beschrieben, erneut Interaktionseffekte schätzen. Dies dient der empirischen Klärung der Frage, ob die Wirksamkeit der Teilnahme an einer Regelleistung bei den Teilnehmerinnen und Teilnehmern systematisch besser oder schlechter ausfällt als im Durchschnitt, wenn die Flüchtlinge zusätzlich bestimmte Beratungs- und Hilfsangebote in Anspruch nehmen.

Ein letztes wichtiges Feld, das im Rahmen des geplanten Evaluierungsvorhabens große Aufmerksamkeit erfordert, betrifft den Einfluss, den die Teilnahme an Kursen und Maßnahmen zum Erlernen der deutschen Sprache auf die Integrationsziele hat. Es ist davon auszugehen, dass die Teilnahme bzw. die damit verbundenen systematischen Fortschritte beim Sprachstand die mit den Regelleistungen der BA erzielten Wirkungen wesentlich verändern. Dies zeigt sich schon daran, dass Teilnahme- bzw. Sprachstandsnachweise Voraussetzung für den Zugang in einige arbeitsmarktpolitische Instrumente sind.

Soweit man die Frage stellt, wie die Teilnahme an Kursen oder Maßnahmen zum Erlernen der deutschen Sprache die Wirksamkeit der zu evaluierenden arbeitsmarktpolitischen Instrumente verändert, geht es erneut um die Bewertung von Interaktionseffekten. Die methodische Vorgehensweise zu ihrer Beantwortung ist dann dem Grunde nach dieselbe wie bei den beiden zuvor diskutierten Konstellationen.

Eine analytisch anders gelagerte Frage wäre, welchen spezifischen Beitrag die Teilnahme an Deutschkursen zu den Integrationszielen und insbesondere zur Beschäftigungsentwicklung bei den Flüchtlingen leistet. Diese integrationspolitisch durchaus relevante Frage lässt sich allerdings kaum adäquat beantworten. Konzeptionell sind die Sprachkurse ein integraler Bestandteil der Maßnahmenketten zur Unterstützung von Flüchtlingen; von daher dürfte es am Ende nur wenige neu ankommende Flüchtlinge geben, die überhaupt nicht von einem der vielfältigen Angebote zum Erlernen der deutschen Sprache erreicht werden. Hierbei dürfte es sich dann nicht um eine zufällige, sondern um eine hochgradig selektierte Personengruppe handeln. Es ist damit zu rechnen, dass sich unter denjenigen, die überhaupt nicht an Sprachkursen teilnehmen, überdurchschnittlich viele Personen mit ausreichenden Ausgangssprachkenntnissen oder aber mit fehlender Integrationsbereitschaft sammeln.

Eine nochmals anders gelagerte Frage wäre, wie sich die Teilnahme an einer bestimmten Maßnahme zum Erlernen der deutschen Sprache, etwa an den Integrationskursen des BAMF, auswirkt. Diese Frage ließe sich grundsätzlich beantworten, ohne die besondere Gruppe derjenigen, die an überhaupt keinem Sprachkurs teilnehmen, als Vergleichsgruppe heranzuziehen. Es könnten nämlich auch Teilnehmerinnen und Teilnehmer, die andere Maßnahmen zum Erlernen der deutschen Sprache durchlaufen, zum Vergleich der jeweils erreichten Zielgrößen herangezogen werden – vorausgesetzt, man beschränkt sich dabei auf solche Teilnehmerinnen und Teilnehmer, die mit genauso großer Wahrscheinlichkeit auch an der zu evaluierenden Maßnahme hätten teilnehmen können. Im Ergebnis erhält man dann allerdings nur

Aussagen zur relativen Wirksamkeit verschiedener Arten von Sprachkursen. Somit lässt sich nur aussagen, ob die zu evaluierende Maßnahme im Vergleich zur Alternative zu besseren Zielwerten führt, aber nicht beurteilen, ob die Zielwerte durch die Maßnahme verbessert oder verschlechtert werden.

Im vorliegenden Kontext dürfte es zudem nicht leicht sein, die Prozesse, die Flüchtlinge dazu bringen, ein ganz bestimmtes Angebot an Sprachkursen in Anspruch zu nehmen, hinreichend allein anhand der beobachtbaren Daten zu beschreiben. Es müsste vielmehr für jede zu bewertende Sprachfördermaßnahme ganz gezielt nach quasi-zufälligen Prozessen gesucht werden, die Teilnehmerinnen und Teilnehmer der zu untersuchenden Maßnahme von vergleichbaren Flüchtlingen in anderen Maßnahmen abgrenzen. Dies ist eine nicht leicht zu bewältigende Aufgabe. Darüber hinaus gilt es auch hier, die Schwierigkeiten im Auge zu behalten, die sich daraus ergeben, dass die Flüchtlinge vielfach im Zeitablauf mehrere Angebote zum Erlernen der deutschen Sprache durchlaufen. Damit sind die Zielgrößen, die bei den Teilnehmerinnen und Teilnehmern beobachtet werden können, nicht nur vom Effekt der eigentlich zu evaluierenden Sprachfördermaßnahme beeinflusst, sondern auch von den Effekten gegebenenfalls vorher oder nachher durchlaufener Maßnahmen. Sich deshalb bei der Evaluation nur auf die Teilnehmerinnen oder Teilnehmer zu konzentrieren, die (bislang) lediglich die zu untersuchende Maßnahme durchlaufen haben, würde erneut Selektivitätsfragen aufwerfen.

Eine generelle Schwierigkeit, mit der umgegangen werden muss, wenn im Rahmen des Evaluationsvorhabens der Einfluss von Angeboten zum Erlernen der deutschen Sprache mit untersucht werden soll, ist die Unvollständigkeit der Daten. Teils liegt die Verantwortlichkeit für die Durchführung von Kursen und Maßnahmen beim BAMF und geht erst später auf die Agenturen für Arbeit oder Jobcenter über. Die BAMF-Aktivitäten sind in den für die Evaluation verfügbaren administrativen Daten aber nur relativ grob dokumentiert. So gut wie gar nicht dokumentiert werden Teilnahmen an den vielfältigen, häufig dezentral organisierten Sprachprogrammen außerhalb des Aufgabenbereichs von BA und BAMF. Folgt man den vorläufigen Ergebnissen der IAB-BAMF-SOEP-Befragung von Flüchtlingen, könnten diese Formen jedoch in der Praxis eine große Rolle spielen. Immerhin gaben dort 38 Prozent an, an einem nicht näher spezifizierten Sprachkurs teilgenommen zu haben. Die Teilnahmequote an den Integrationskursen lag dagegen nur bei 33 Prozent.

Angesichts dieser vielfältigen Schwierigkeiten ist zu empfehlen, im Rahmen der geplanten Evaluation zwar den Einfluss von Maßnahmen zur Förderung des Erlernens der deutschen Sprache mit zu untersuchen, jedoch keine eigenständige Evaluation einzelner Sprachfördermaßnahmen anzustreben. Vielmehr sollten lediglich Interaktionseffekte untersucht werden, also geklärt werden, inwieweit die Teilnahme an Kursen und Maßnahmen zum Spracherwerb die Wirkungen der zu evaluierenden arbeitsmarktpolitischen Instrumente der BA verstärken. Hierbei sollte nach Möglichkeit der Einfluss verschiedener Typen und Verlaufsmuster von Sprachfördermaßnahmen herausgearbeitet werden. Um mit der inkonsistenten Teilnahmedatenbasis in diesem Bereich umzugehen, wird vorgeschlagen, die zur Umsetzung dieses Ansatzes benötigten Informationen zur Wahrnehmung der verschiedenen Möglichkeiten zum Erlernen der deutschen Sprache bei den Flüchtlingen selbst zu erheben.

Im Rahmen dieser Befragung sollten folgende Aspekte erfasst werden:

- die Teilnahme an spezifischen Kursangeboten und Maßnahmen von BA und BAMF (insbesondere Integrationskurs des BAMF, ESF-BAMF-Kurs, BA-Einstiegskurs, berufsbezogene Sprachförderung) sowie die Teilnahme an Deutschkursen außerhalb dieser Angebote, wobei Mehrfachnennungen wegen der zu erwartenden Maßnahmenketten zugelassen sein müssen;

- der jeweilige zeitliche Umfang (Gesamtzahl der Unterrichtsstunden) und die Intensität (Anzahl der Unterrichtsstunden im Verhältnis zur Laufzeit des Sprachkurses in Tagen, gemessen an der Differenz von End- und Anfangsdatum) der besuchten Kurse;
- die Grundausrichtung des jeweiligen Kurses: allgemeiner oder berufsorientierter Spracherwerb;
- die Selbsteinschätzung der Qualität des jeweiligen Kurses und der damit erzielten Sprachfortschritte.

Die Landschaft der Sprachförderung für Flüchtlinge befindet sich momentan stark in Bewegung, insbesondere durch die laufende Umsetzung des Kombi-Programms KomPas (Kompetenzfeststellung, frühzeitige Aktivierung und Spracherwerb) und des Gesamtprogramms Sprache (GPS). Daher muss die weitere Entwicklung der Praxis in den kommenden Monaten bei der Finalisierung des Erhebungsinstruments für das Evaluationsvorhaben im Blick behalten werden.



## 5. Zielgrößen und ihre Operationalisierung

### 5.1 Einleitung

Es liegt nahe, den Erfolg der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge an den damit erzielten Verbesserungen der Position der Geförderten am Arbeitsmarkt zu messen. Diese Zieldimension hat auch hier, wie bei den meisten anderen Maßnahmen der aktiven Arbeitsmarktpolitik – eine Ausnahme bildet etwa das an die Heranführung an ein Arbeitsleben ausgerichtete Bundesprogramm für Langzeitarbeitslose „Soziale Teilhabe am Arbeitsmarkt“ – selbstverständlich besonderes Gewicht.

Dennoch erscheint es konzeptionell sinnvoll, in dem geplanten Evaluationsvorhaben die Perspektive über die Arbeitsmarktintegration hinaus zu weiten und Integration als multidimensionales Konstrukt aufzufassen, worin die im Vergleich zur Aufnahmegesellschaft gleichberechtigte Teilhabe am Arbeitsmarkt nur eines von mehreren Elementen ist. Hierfür spricht, dass sich Fortschritte in der Arbeitsmarktintegration und in anderen ökonomisch-soziokulturellen Dimensionen der Integration wechselseitig beeinflussen. Somit können Verbesserungen bei Indikatoren, die diesen anderen Dimensionen der Integration zuzurechnen sind, als Frühindikatoren für später nachlaufende Verbesserungen bei der Arbeitsmarktintegration dienen.

Eine Betrachtung von Frühindikatoren wiederum erscheint angesichts der begrenzten Laufzeit des geplanten Evaluationsvorhabens sinnvoll, da davon auszugehen ist, dass die Arbeitsmarktintegration der Flüchtlinge angesichts ihrer Startvoraussetzungen auch dann ein lang andauernder Prozess bleibt, wenn die Hilfen der Arbeitsmarktpolitik wirksam sind. So zeigt die erste vorläufige Auswertung der neu begonnenen IAB-BAMF-SOEP-Befragung von in den letzten Jahren aufgenommenen Geflüchteten, dass unter den im Jahr 2013 und vorher zugezogenen Geflüchteten lediglich 31 Prozent als Vollzeit- und Teilzeitkräfte, Selbständige, Auszubildende oder Praktikanten erwerbstätig sind, und unter den im Jahr 2014 zugezogenen Geflüchteten nur 22 Prozent (IAB 2016).

Ein weiteres Argument für eine Weitung der Zielperspektive ist, dass mit den Hilfen für Flüchtlingen, die Deutschland aus humanitären Gründen aufgenommen hat, nicht ausschließlich ökonomische Ziele verfolgt werden, sondern auch gesellschaftliche Ziele des sozialen Zusammenhalts. Zudem würde diese Vorgehensweise dem Stand der wissenschaftlichen wie im Übrigen auch der politischen Diskussion, was gute Integration von Zugewanderten ausmacht, besser gerecht (vgl. Werth et al. 1997; Siegert 2006; Filsinger 2008).

Eine zu beachtende Hürde ist allerdings, dass sich die dafür benötigten Informationen nur zu einem geringen Teil aus den für die Evaluation zur Verfügung stehenden administrativen Daten gewinnen lassen. Um das geplante Evaluationsvorhaben nicht zu überfrachten und dabei wirtschaftlich vorzugehen, ist darum eine Prioritätensetzung erforderlich. Bei der Auswahl von solchen Zielgrößen, die Arbeitsmarktintegration nicht direkt abbilden, sollte darauf geachtet werden, dass sie für die langfristige Arbeitsmarktposition von Geflüchteten eine prädiktive Qualität aufweisen und sich außerdem mit vertretbarem Aufwand durch ergänzende Umfragen gewinnen lassen. Die Basis für diesen Auswahlprozess liefert ein Überblick über einschlägige empirisch orientierte Untersuchungen zu Faktoren, die in einem systematischen positiven Zusammenhang zur Arbeitsmarktintegration von Flüchtlingen in den Aufnahmeländern stehen.

Der folgende Abschnitt 5.2 befasst sich zunächst mit einer Reihe von adäquaten Zielindikatoren für die Bewertung der Arbeitsmarktintegration von Flüchtlingen im eigentlichen Sinn. Abschnitt 5.3 reflektiert anschließend über weitere Bereiche ökonomisch-soziokultureller Integration, die als empirisch relevante

Treiber des längerfristigen Arbeitsmarkterfolgs ein vollständigeres Bild liefern könnten. Abschnitt 5.4 diskutiert, wie diese in den für den Evaluationszweck zugänglichen administrativen Individualdaten praktisch nicht abgedeckten Bereiche im Rahmen ergänzender Erhebungen bei den Zielpersonen operationalisiert werden könnten.

## 5.2 *Arbeitsmarktintegration im eigentlichen Sinn*

Um Fortschritte bei der Arbeitsmarktintegration im eigentlichen Sinn zu beobachten, können auf Grundlage der IEB-Daten ohne größeren Aufwand eine Reihe von Indikatoren erstellt werden, wie sie bei Evaluationen zu aktiven arbeitsmarktpolitischen Maßnahmen in Deutschland gängig sind. Hierzu gehören insbesondere:

- die Wahrscheinlichkeit der Aufnahme einer (ungeförderten) sozialversicherungspflichtigen Beschäftigung,
- die Wahrscheinlichkeit der Aufnahme einer geringfügig entlohnten Beschäftigung im Minijob,
- die Wahrscheinlichkeit der Aufnahme einer Teilzeitbeschäftigung,
- die Wahrscheinlichkeit, die Arbeitslosigkeit zu verlassen, möglichst differenziert nach Grund des Abgangs (neben Aufnahme einer Beschäftigung insbesondere Aufnahme eines Studiums, Übergang in eine Maßnahme oder Rückzug vom Arbeitsmarkt),
- die zeitliche Dauer bis zur Aufnahme einer Beschäftigung,
- die Stabilität der aufgenommenen Beschäftigung, gemessen an der Dauer des Arbeitsverhältnisses,
- die zeitliche Dauer der Arbeitslosigkeit und des Hilfebezugs,
- die Höhe des erzielten (Tages-)Arbeitsentgelts,
- die Wahrscheinlichkeit der nicht qualifikationsadäquaten Tätigkeit sowie die Wahrscheinlichkeit der Tätigkeit in einer Niedriglohn-Branche oder in einem Niedriglohn-Beruf.<sup>5</sup>

Somit lassen sich anhand der administrativen Daten recht differenzierte Einschätzungen zur Entwicklung der Partizipations- und Beschäftigungsraten der Geflüchteten, aber auch zur Qualität der erreichten Beschäftigung gewinnen. Die Informationen zur Dauer des Leistungsbezugs und zum Erwerbseinkommen liefern darüber hinaus eine Basis für zumindest näherungsweise Kosten-Nutzen-Bewertungen, wie sie für Analysen zur Effizienz des Instrumenteneinsatzes benötigt werden.

---

<sup>5</sup> Die dafür erforderliche Kategorisierung der Branchen und Berufe kann anhand von volkswirtschaftlichen Daten (etwa Beschäftigtenstatistik, Verdienststrukturerhebung) vorgenommen werden, indem das mittlere Lohnniveau der Beschäftigten in einer Branche oder einem Beruf ins Verhältnis zum mittleren Lohnniveau aller Beschäftigten gesetzt wird. Folgt man dabei den Konventionen zur Abgrenzung des Niedriglohnssektors durch internationale Organisationen wie OECD und ILO, beginnt der Niedriglohnsektor dort, wo das Entgelt niedriger ist als zwei Drittel des mittleren Entgelts. Die konventionelle Grenzziehung ist allerdings nicht wissenschaftlich präzise begründet, so dass für Sensitivitätsanalysen auch andere Grenzen möglich wären. Im Hinblick auf die Beurteilung der qualifikationsadäquaten Tätigkeit ist zu beachten, dass formale Qualifikationsniveaus im Kontext der Personen mit Fluchthintergrund nicht unbedingt aussagekräftig sind. Auch die Selbstauskünfte beschäftigter Personen in der Zielgruppe sind möglicherweise unzuverlässiger als normal, weil sie mit den in Deutschland üblichen beruflichen Anforderungen wenig vertraut sind.

Um spezifischeren Gesichtspunkten der Integration Rechnung zu tragen, könnten anhand der Berufs- und Wirtschaftszweiginformationen zudem Segregationsprozesse in den Blick genommen werden. Dazu wäre zu analysieren, wie hoch die Wahrscheinlichkeit ist, dass Flüchtlinge in Bereichen des Arbeitsmarkts beschäftigt werden, in denen überdurchschnittlich viele Menschen aus den Hauptherkunftsgebieten der Asylsuchenden (alternativ: Ausländerinnen und Ausländer allgemein oder aus Drittstaaten) tätig sind, bzw. die Wahrscheinlichkeit, dass Flüchtlinge in besonders von deutschen Arbeitskräften besetzten Bereichen beschäftigt werden.

Weitere relevante Informationen, die den Arbeitsmarktstatus i.e.S. der Flüchtlinge betreffen, könnten nur per Befragung ergänzt werden. Dies betrifft insbesondere die Chancen auf qualifikations- oder fähigkeitsadäquate Beschäftigung, die bei Zuwanderinnen und Zuwanderern tendenziell geringer sind als bei der übrigen Bevölkerung (Leuven und Osterbeek 2011). Im Kontext der qualifikationsadäquaten Beschäftigung interessiert insbesondere das Phänomen der Überqualifizierung, also die Situation, in denen Beschäftigte über ein höheres formales Qualifikationsniveau verfügen, als es für die von ihnen ausgeübte Beschäftigung benötigt wird – bildlich gesprochen der Fall des Taxi fahrenden Ingenieurs. Dagegen geht es bei der Frage der fähigkeitsadäquaten Beschäftigung darum, inwieweit für die ausgeübte Tätigkeit die erlernten Fähigkeiten verwendet werden (Quintini 2011). Dies kann auch dann der Fall sein, wenn eine qualifikationsadäquate Beschäftigung ausgeübt wird; ein Beispiel wäre der Fall eines ausgebildeten Ingenieurs, der als Geschäftsführer mit betriebswirtschaftlichen Aufgaben befasst ist. Ein gängiger Weg, um qualifikations- oder fähigkeitsadäquate Beschäftigung zu erfassen, ist, Selbsteinschätzungen der Beschäftigten einzuholen. Die gemessene Mismatch-Häufigkeit kann dabei sensitiv auf die Art der Fragestellung reagieren (Sloane 2003). Im Evaluationskontext, in dem es um maßnahmeninduzierte Veränderungen in der Mismatch-Häufigkeit geht, können Messfehler in den Niveaugrößen allerdings herausdifferenziert werden.<sup>6</sup>

Weitere nur per Befragung zu gewinnende Informationen zum Stand der Arbeitsmarktintegration betreffen die – mutmaßlich nicht selten anzutreffende – vorübergehende Erwerbstätigkeit im Praktikum sowie die nicht abhängige Beschäftigung in der Selbstständigkeit. Ebenfalls nur per Befragung zu gewinnen wären Angaben zu informellen Beschäftigungsverhältnissen, was aber in Folge dieser rechtlich problematischen Konstellation erhebungstechnisch mit Risiken verbunden und daher eher nicht zu empfehlen ist.

### **5.3 Weitere Bereiche ökonomisch-soziokultureller Integration**

Es dürfte breiter Konsens bestehen, dass die Arbeitsmarktintegration nur ein Baustein dessen ist, was die Integration von Geflüchteten (und allgemeiner die Integration von Menschen mit Migrationshintergrund) ausmacht. Allerdings gibt es trotz der weiten Verbreitung des Begriffs kein einheitliches Verständnis über die Elemente, die Integration im Ganzen ausmachen. Abbildung 5.1 zeigt den auf Grundlage eines Überblicks über den Stand der Literatur und praktischer Erfahrungen unternommenen Versuch einer ganzheitlichen konzeptionellen Synthese (Ager und Strang 2008).

Auf Basis der fundamentalen Rechte, die den Aufenthaltsstatus sowie die Handlungsmöglichkeiten und -grenzen der Flüchtlinge in der Aufnahmegesellschaft festlegen, stehen eine Reihe unterschiedlicher Gebiete, die in der Gesamtheit Integration ausmachen. Diese Gebiete haben sämtlich eine Doppelnatur: einerseits kann die individuelle Position in einem Gebiet eine bessere oder schlechtere Integration

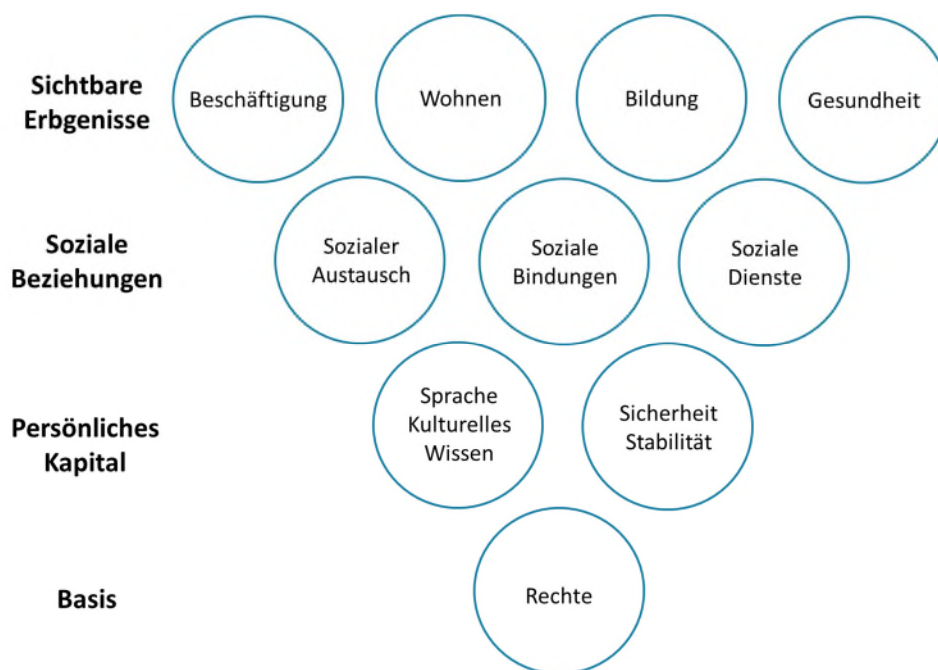
---

<sup>6</sup> Problematische Verzerrungen könnten allerdings dann auftreten, wenn die Teilnahme an einer zu evaluierenden Maßnahme selbst die Wahrscheinlichkeit, dass die Zielgröße fehlerhaft gemessen wird, verändert. Nicht ausgeschlossen werden kann etwa, dass Teilnehmerinnen und Teilnehmern maßnahmenbedingt zu einer anderen Selbsteinschätzung über die eigenen Fähigkeiten gelangen.

anzeigen; andererseits handelt es sich aber auch um Inputs, die Integration voranbringen oder auch behindern können. Beispielsweise können gute Sprachkenntnisse und gutes kulturelles Wissen über die Aufnahmegesellschaft ein Ausdruck dafür sein, dass bereits Integrationsfortschritte erzielt wurden, denn Flüchtlinge, die sich weit von ihrer Heimat aufhalten, bringen entsprechendes Wissen typischerweise nicht von Anfang an mit. Zugleich ist entsprechendes kulturelles Humankapital jedoch ein wichtiger Faktor, um etwa soziale Beziehungen aufzubauen oder am Arbeitsmarkt Fuß zu fassen.

Hieraus ergibt sich, dass die verschiedenen in Abbildung 5.1 skizzierten Bereiche nicht abgegrenzt voneinander zu sehen sind, sondern sich wechselseitig beeinflussen und durchdringen. Diese Abhängigkeiten sind auch für die Vorgehensweise der geplanten Evaluation bedeutsam. Um isolieren zu können, wie sich die zu evaluierenden Instrumente auf Zielgrößen der Evaluation auswirken, ist es notwendig, den Ausgangsstand der Integration in den verschiedenen Bereichen zu kennen und die Veränderungen bei den Integrationszielgrößen bei Teilnehmerinnen und Teilnehmern den Veränderungen gegenüberzustellen, die bei Nichtteilnehmerinnen und Nichtteilnehmern mit gleichen Ausgangsvoraussetzungen eintreten. Somit müssen die Indikatoren für Integration mehrfach gemessen werden.

Abbildung 5.1 Zentrale Bereiche bei der Integration von Flüchtlingen



Darstellung des IZA auf Grundlage von Ager und Strang (2008: 170).

Neben der Beschäftigung sind Wohnen, Bildung und Gesundheit viel beachtete sichtbare Ergebnisse (und Instrumente) der Integration von Flüchtlingen. **Wohnen** ist grundsätzlich ein Schlüsselfaktor für soziale Integration (Esser 2001; Häußermann und Siebel 2000). Dies gilt auch für Personen mit Migrationshintergrund, und mutmaßlich auch für Flüchtlinge, die überdurchschnittlich häufig in prekären Wohnverhältnissen leben und vor allem in Ballungsräumen räumlicher sozialer Segregation ausgesetzt sind. Zur Erfassung dieses Bereichs kommen als Kennziffern beispielsweise die physische Größe, die Qualität und Ausstattung der Wohnung, die Mietbelastung und das Risiko des Mietausfalls zum Einsatz. Im Hinblick auf die Beschäftigungschancen spielen auch Aspekte der Qualität des Wohnumfelds,

wie etwa die Nähe zu (potenziellen) Arbeitsstätten oder die lokale öffentliche Infrastruktur, eine Rolle. Merkmale des Wohnumfelds können auch als Indikatoren für residentielle Segregation in Form der Struktur der Wohnbevölkerung dienen (Friedrichs 1980). Diese lässt sich insbesondere am Anteil der Personen mit ausländischer Staatsbürgerschaft (allgemein oder bezogen auf die Merkmale der Zielpersonen) im Quartier oder im Stadtteil festmachen (ethnische Segregation; Friedrichs und Triemer 2009). Residentielle Segregation kann sich aber auch in einer Konzentration der geflüchteten Menschen in Quartieren oder Stadtteilen ausdrücken, in denen es besonders viele sozial benachteiligte Haushalte gibt (soziale Segregation; Strohmeier 2006); Indikatoren hierfür sind etwa der Anteil der Haushalte in der Grundsicherung oder mit niedrigem Kredit-Rating. Entsprechende kleinräumige Daten zur Charakterisierung von kleinen Raumeinheiten unterhalb der Ebene der Kommune sind in Deutschland prinzipiell verfügbar.

Im Bereich **Bildung**, der sowohl für die Zugangschancen zum Arbeitsmarkt als auch für die Möglichkeiten, am Aufbau der Gesellschaft aktiv mitwirken zu können, konstitutiv ist, lässt sich Integration grundsätzlich input- oder ergebnisorientiert analysieren. Inputorientierte Maße betrachten die Intensität und Art der in Anspruch genommenen Bildungsangebote, wobei sowohl schulische oder berufliche Qualifizierung als auch die Weiterbildung in den Blick genommen werden müssen. Weiterführende inputorientierte Kennziffern bilden darüber hinaus eine Qualitätsdimension ab, beispielsweise über Klassengrößen oder Kostenkennziffern. Ergebnisorientierte Indikatoren versuchen dagegen, die individuellen Resultate von Bildungsaktivitäten zu messen, etwa an den erreichten Abschlüssen, den erzielten Noten oder standardisierten Testergebnissen.

Im Bereich der **Gesundheit** können Flüchtlinge spezifische Bedürfnisse haben. Die Ursachen von Flucht oder die während der Flucht gemachten Erfahrungen können mit physischen oder psychischen Verletzungen verbunden sein, die sich etwa auf die Leistungsfähigkeit in Beschäftigung und Bildung oder die Möglichkeiten zum Aufbau sozialer Bindungen auswirken. Hinzu kommt, dass Flüchtlinge möglicherweise kulturell angepasste oder zumindest kulturell bewusste Gesundheitsangebote benötigen. Grundsätzlich lässt sich die Gesundheitsdimension von Integration wiederum sowohl input- als auch ergebnisorientiert messen. Inputorientierte Größen wären etwa die Zahl der Arztbesuche oder die Inanspruchnahme spezifischer Hilfsangebote. Ergebnisorientierte Messungen würden etwa auf die Zahl der Krankheitstage abstellen. Eine Besonderheit in diesem Bereich ist, dass objektiv messbare Indikatoren und die subjektive Wahrnehmung der Betroffenen markant auseinanderfallen können. Darum sind auch Informationen über das subjektive gesundheitliche Wohlbefinden von Bedeutung.

Die sichtbar werdenden Ergebnisse in den Bereichen Beschäftigung, Wohnen, Bildung und Gesundheit werden durch zwei Gruppen von Faktoren stark beeinflusst: die sozialen Beziehungen im persönlichen Nahbereich und individuell zu verortendes Kapital, das Barrieren bei der Integration aus dem Weg räumen oder auch aufbauen kann.

Das zu beachtende soziale Beziehungsgeflecht lässt sich analytisch weiter unterteilen. Im Bereich **soziale Bindungen** geht es um die Nähe zur eigenen Familie, zur eigenen ethnischen Gruppe oder zu Netzwerken zur Selbsthilfe unter Flüchtlingen, die im Integrationsprozess eine bedeutsame Rolle spielen und beispielsweise das Wohnumfeld oder das Wohlbefinden stark beeinflussen können. Im Gegensatz zu den sozialen Bindungen stellt der Bereich **sozialer Austausch** auf die Intensität und Qualität der Interaktion zwischen den Flüchtlingen und der Aufnahmegesellschaft ab. Mit sozialem Austausch kann beispielsweise ein informeller Informationsfluss verbunden sein, der die Handlungsoptionen der Minderheit deutlich erweitert. Eine „Willkommenskultur“ und ehrenamtliche Hilfsangebote aus der Aufnahmegesellschaft können die Flüchtlinge ihrerseits zu mehr Bemühungen um Integration motivieren. Andererseits kann es auch einen negativen sozialen Austausch geben, etwa wenn die Flüchtlinge Diskriminierungserfahrungen machen. Ein dritter wichtiger Bereich der sozialen Beziehungen im Nahbereich sind schließlich **soziale**

**Dienste.** Hierbei geht es im Gegensatz zum Bereich des sozialen Austauschs um die Beziehungen zwischen den Flüchtlingen und lokalen öffentlichen Institutionen. Dies schließt etwa die Dienstleistungen der Jobcenter ein, ist aber nicht auf diese beschränkt, sondern umfasst unabhängig von der Trägerschaft alle Einrichtungen und Services, die für die Integration relevante Angebote machen oder Informationen und Orientierung vermitteln.

Um den Bereich der sozialen Beziehungen adäquat zu beschreiben, muss die Struktur des Beziehungsnetzwerks, das Flüchtlinge persönlich pflegen, differenziert erfasst werden. Dabei geht es sowohl um die Intensität als auch die Qualität der auf den verschiedenen angesprochenen Ebenen gepflegten Kontakte und den damit jeweils verknüpften konkreten Handlungen. Es sollte möglichst auch erfasst werden, warum bestimmte Beziehungen nicht gepflegt werden (etwa aus Unkenntnis oder fehlender Erreichbarkeit und Zugänglichkeit).

Blickt man auf die Elemente des persönlichen Kapitals, das mit Barrieren bei der Integration verbunden sein kann, sind einerseits die Verfügbarkeit ziellandspezifischen Humankapitals und andererseits Aspekte der Sicherheit und Absicherung zu nennen. Was den Bereich **Sprache und kulturelles Wissen** angeht, werden Kenntnisse der Verkehrssprache des Aufnahmelandes vielfach als Grundvoraussetzung für die Integration in Bildung und Arbeit sowie gesellschaftliche Teilhabe gesehen. Über den Erwerb von Sprachkenntnissen hinaus kann auch der Erwerb weiterer aufnahmelandsspezifischer Kenntnisse für die Integration von erheblichem Belang sein. Dazu zählen etwa Kenntnisse der Funktionsweise des Arbeitsmarkts und des Bildungssystems sowie Rechts- und Institutionenkenntnisse, die typischerweise erst nach Ankunft im Aufnahmeland erworben werden.

Die Messung von Sprachkenntnissen kann prinzipiell anhand von standardisierten Sprachstandsprüfungen erfolgen, wie sie die Flüchtlinge in Deutschland im Rahmen von Integrations- und Sprachkursen ablegen. Eine Alternative sind Selbsteinschätzungen zur Qualität der Sprachkenntnisse oder Informationen zur Intensität und Qualität der Nutzung der Landessprache im Alltagskontext. Aufwändiger zu erfassen ist die Dimension des integrationsrelevanten kulturellen Wissens. Dies kann etwa anhand der Kenntnisse spezifischer Institutionen oder des Realitätsgehalts der Vorstellungen über gängige Abläufe und Verhaltensweisen im Aufnahmeland erfolgen. Wichtig für den deutschen Kontext ist beispielsweise ein Verständnis für die Bedeutung und Funktionsweise des betrieblichen Ausbildungssystems, das die Flüchtlinge aus ihren Herkunftsländern nicht kennen.

Im Bereich **Sicherheit und Stabilität** geht es einerseits um Aspekte des Gefühls der persönlichen Sicherheit. Beeinträchtigungen dieses Sicherheitsgefühls, die etwa durch verbale, körperliche und (insbesondere bei Frauen) sexuelle Gewalterfahrungen im Aufnahmeland zustande kommen, können Integrationsprozesse erheblich stören, weil sie verunsichern und möglicherweise reziproke Handlungen auslösen. Die Dimension der Stabilität umfasst Aspekte, die die Planbarkeit und Realisierbarkeit zukunftsorientierter Investitionen in Integration betreffen. Hierzu zählen etwa die erwartete oder geplante Dauer des Aufenthalts im Aufnahmeland bis zur Rückkehr ins Herkunftsland, die wiederum vom aufenthaltsrechtlichen Status abhängen, sowie auch zeitnah beabsichtigte Wechsel des Wohnortes innerhalb des Aufnahmelandes, welche einen Aufbau sozialer Beziehungen und die Wahrnehmung lokaler Integrationsangebote erschweren. Zur Dimension der Stabilität gehören aber auch Aspekte der materiellen Sicherheit. Hierbei geht es insbesondere um die Frage der persönlichen materiellen Ressourcen – messbar etwa am laufenden Haushaltseinkommen oder an den vorhandenen Rücklagen – sowie eventueller finanzieller Verpflichtungen, etwa in Folge von fluchtbedingten Schulden. Diese Größen beeinflussen etwa die individuellen Möglichkeiten, in Bildung und Qualifizierung zur Verbesserung der Integrationschancen zu investieren.

Die genannten Dimensionen von Integration lassen sich auch als Elemente eines Konzepts sozialer Teilhabe von Geflüchteten verstehen. Dieses Konzept bezieht sich zum einen auf den Diskurs um Exklusion, zu dem Inklusion und Teilhabe die Gegenbegriffe darstellen (Leisering 2004, Kronauer 2006), zum anderen auf den Diskurs um *capabilities* bzw. Verwirklichungschancen (Sen 1992, Nussbaum 2000, Bonvin und Orton 2009). Im Sinne des Capability-Ansatzes lässt sich soziale Teilhabe definieren als die faktische Möglichkeit von Personen, ein Leben entsprechend persönlicher Lebensziele zu führen und dabei auch gesellschaftlich anerkannt zu werden (Bonvin 2006: 65). Das Konzept der Integration betont dagegen eher das Ziel der Überbrückung der Distanzen zwischen der aufgenommenen Minderheit und der aufnehmenden Mehrheit in ökonomischer, sozialer und gesellschaftlicher Hinsicht. Angesichts der anfänglichen Distanz gehen Fortschritte an sozialer Teilhabe und an Integration jedoch ein gutes Stück weit miteinander einher.

Das Konzept der sozialen Teilhabe verweist allerdings noch auf einige weitere Indikatoren, die gemäß den vorhandenen empirischen Erkenntnissen auch im Hinblick auf Integrationsziele am Arbeitsmarkt relevant sein können: die gesetzten beruflichen und sonstigen Lebensziele, sowie die Wahrnehmung der eigenen Selbstwirksamkeit, das heißt der Überzeugung über jene eigenen Fähigkeiten, die man benötigt, um eine bestimmte Handlung zu organisieren und damit die gesetzten Ziele zu erreichen (vgl. etwa Caliendo et al. 2015). Da sowohl die eigenen (Aufstiegs-)Ambitionen als auch das Vertrauen in die persönliche Fähigkeit, sie zu verwirklichen, im Integrationsprozess veränderlich sind und zugleich nicht ohne Einfluss auf die Integrationsergebnisse – nicht zuletzt am Arbeitsmarkt – sein dürften, ist es empfehlenswert, den bisher dargestellten Zielkatalog noch um diese Dimension zu erweitern.

Der oben beschriebene Zielkatalog erscheint zwar breit angelegt. Die darin aufgenommenen Bereiche erfüllen aber das eingangs dieses Kapitels benannte Priorisierungskriterium, dass sie mit hoher Wahrscheinlichkeit in einem systematischen und auch substanziellen Zusammenhang zur übergeordneten Zielgröße der längerfristigen Arbeitsmarktintegration stehen. Diese Einschätzung unterstützt eine Übersicht über einschlägige empirische Studien, die anhand von Daten für entwickelte Länder einen signifikanten Zusammenhang zwischen unterschiedlichen Dimensionen sozialer Teilhabe und der Integration von Flüchtlingen und den von ihnen am Arbeitsmarkt erzielten Ergebnissen konstatieren (Tabelle 5.1).

Nach unserem Kenntnisstand fehlen ökonometrisch angelegte Studien, die im Kontext von Flüchtlingsmigration empirische Zusammenhänge zwischen der individuellen Wohnsituation, dem kulturellen Wissen jenseits der Kenntnisse der Landessprache sowie dem Bereich der Sicherheit und Stabilität einerseits und Beschäftigungsraten oder Entgeltgrößen andererseits herstellen. Einschränkend muss zur Übersicht in Tabelle 5.1 außerdem gesagt werden, dass die Befunde vieler der genannten Studien – eine Folge der normalerweise eher kleinen Flüchtlingsbevölkerungen – nur auf relativ wenigen Beobachtungen beruhen. Ihre interne Validität ist zudem häufig dadurch eingeschränkt, dass die Wirkungsrichtung der gemessenen statistischen Korrelationen nicht eindeutig identifiziert wird. Beispielsweise ist im Bereich Gesundheit nicht ohne weiteres geklärt, ob Depression eine Ursache oder die Folge der Arbeitslosigkeit ist (Bogic 2012). Ausnahmen bilden etwa die Studien von Edin et al. (2003) und Damm (2009), die auf administrativen Zuweisungen beruhende quasi-experimentelle Variation in den Wohnorten von Flüchtlingen ausnutzen, um den Einfluss abzuschätzen, den das Wohnen in einer ethnischen Enklave auf die Arbeitsmarktchancen ausübt.

**Tabelle 5.1 Empirische Studien mit signifikantem Zusammenhang zwischen Integrationsindikator und Arbeitsmarktintegration von Flüchtlingen**

	<b>Integrations- indikator</b>	<b>Arbeitsmarkt- indikator</b>	<b>Land</b>	<b>Studie</b>
Gesundheit	Behinderung	Beschäftigung	USA	Potocky-Tripodi (2001)
	Gesundheitszustand	Beschäftigung	USA	Strand (1984)
	Gesundheitszustand	Beschäftigung	USA	Tran (1991)
	Depression	Arbeitslosigkeit	Deutschland, Italien, UK	Bogic (2012)
	Trauma	Partizipation	Norwegen	Hauff und Vaglum (1993)
Bildung	Bildungsniveau	Beschäftigung	Schweden	Bevelander et al. (2009)
	Bildungsniveau	Beschäftigung, Lohn, Berufliche Stellung	USA	Connor (2010)
	Bildungsniveau	Beschäftigung, Lohn, Jahresentgelt	USA	Siraj (2007)
	Bildungsniveau	Beschäftigung	USA	Tran (2001)
	Bildungsniveau	Haushaltseinkommen	USA	Potocky-Tripody (2001)
	Training nach Ankunft	Beschäftigung	Australien	Ibrahim et al. (2010)
	Training nach Ankunft	Berufliche Position	USA	Connor (2010)
Sprache	Sprachkenntnisse am Anfang	Beschäftigung	Australien	Ibrahim et al (2010)
	Sprachkenntnisse	Löhne	USA	Mamgain und Collins (2003)
	Sprachkenntnisse	Beschäftigung	USA	Potocky-Tripody (2003)
	Sprachkenntnisse	Beschäftigung, Lohn, Jahresentgelt	USA	Siraj (2007)
	Sprachkenntnisse	Beschäftigung	Australien	Waxman (2001)
	Sprachkurs	Beschäftigung	USA	Strand (1984)
Soziale Beziehungen	Lokale Serviceangebote	Beschäftigung	USA	Tran (1991)
	Lokales soziales (und ökonomisches) Klima	Löhne	USA	Connor (2010)
	Lokales soziales (und ökonomisches) Klima	Beschäftigung	Australien	Correa-Velez et al. (2013)
	Ethnisches Netzwerk	Beschäftigung	Australien	Correa-Velez et al. (2013)
	Integration in ethnische Gemeinschaft	Beschäftigung	Australien	Ibrahim et al. (2010)
	Beschäftigungsrate in der ethnischen Gemeinschaft	Beschäftigung	Dänemark	Damm (2012)
	Soziale Unterstützung Freundschaften	Einkommen	USA	Takeda (2000)
Wohn- umfeld	Ethnische Enklave	Löhne	Schweden	Edin et al. (2003)
	Ethnische Enklave	Beschäftigung	Dänemark	Damm (2009)
Akkul- turation	Selbstwirksamkeit	Beschäftigung	Neuseeland, Australien	Sulaiman-Hill und Thomson (2013)
	Integrations- bereitschaft	Beschäftigung	USA	Potocky-Tripodi (2003)

Darstellung des IZA.



Hinzu kommt das prinzipiell kaum befriedigend lösbare Problem der externen Validität. Es lässt sich nicht sagen, inwieweit die in einem bestimmten wirtschaftlichen und institutionellen Umfeld und für eine spezifisch zusammengesetzte Flüchtlingsbevölkerung gewonnenen empirischen Zusammenhänge auch auf einen anderen Kontext – konkret auf die in jüngster Zeit durch Deutschland in Folge von Fluchtmigration aufgenommenen Menschen – übertragen werden können. Allerdings wird die Vermutung, dass die oben diskutierten Dimensionen sozialer Teilhabe und Integration auch im vorliegenden Zusammenhang eine signifikante Vorhersagekraft entfalten könnten, auch durch die sehr umfangreiche empirische Literatur zu den individuellen und institutionellen Faktoren, welche die ökonomische Integration von Zuwanderinnen und Zuwanderern fördern oder behindern können, gestützt.<sup>7</sup>

In diesem Rahmen können einschlägige Studien nur ausschnittsweise behandelt werden. Die Übersicht in Tabelle 5.2 konzentriert sich auf eine Auswahl von Untersuchungen, welche überzeugende Lösungen zur Offenlegung von Ursache-Wirkungs-Beziehungen haben und auf Daten für Deutschland beruhen. Die hinzugenommenen internationalen Studien sollen zeigen, dass eine bestimmte empirische Wirkungsbeziehung in unterschiedlichen Kontexten nachgewiesen wird, was ein Indiz für die externe Validität oder Übertragbarkeit des Befundes sein kann.

Insgesamt sehr gut abgesichert ist die Hypothese, dass ein besseres Bildungs- und Qualifikationsniveau die Arbeitsmarktposition von Migrantinnen und Migranten verbessert. Anders als bei Einheimischen muss dabei zwischen dem Humankapital, das im Herkunftsland erworben wurde, und solchem, das im Aufnahmeland erworben wird, unterschieden werden. Letzteres wirft, wie der umfassende Survey der Befunde von Dustmann und Glitz (2011) zeigt, systematisch einen höheren Ertrag in Form verbesserter Beschäftigungschancen und höherer Löhne ab als das aus dem Herkunftsland mitgebrachte Humankapital. Bezogen auf den Kontext des Evaluationsvorhabens bedeutet dies, dass die Beteiligung der Flüchtlinge an Bildung und Qualifizierung seit ihrer Ankunft in Deutschland besondere Beachtung verdient.

Eine besondere Form des aufnahmelandsspezifischen Humankapitals sind Kenntnisse der Landessprache. Auch hier kommt die empirische Literatur zu einem qualitativ konsistenten Befund. Die Literatur zusammenfassend konstatieren Chiswick and Miller (2014), dass gute Fähigkeiten in der Sprache des Aufnahmelandes die Löhne von erwachsenen Migranten um 5 bis 35 Prozent erhöhen.

Etwas weniger eindeutig sind die empirischen Befunde, die sich auf die Dimensionen der sozialen Beziehungen – meist erfasst durch die Anzahl und Intensität der Kontakte zu Mitgliedern der Aufnahmegesellschaft sowie der familiären Bindungen im Ziel- und Herkunftsland – und das Wohnumfeld – häufig charakterisiert durch den Grad der ethnischen Durchmischung in der Nachbarschaft – beziehen. Diese Beobachtung passt allerdings dazu, dass theoretisch betrachtet die Intensität der Bindung an die eigene „Community“ sowohl positive als auch negative Wirkungen auf die ökonomischen Entwicklungschancen ausüben kann (Cutler und Glaeser 1997). Insgesamt betrachtet legen die empirischen Befunde jedoch nahe, dass Zuwanderinnen und Zuwanderer, die über mehr Kontakte zu Mitgliedern der Aufnahmegesellschaft verfügen, weniger enge Beziehungen in ihr Heimatland pflegen und außerhalb von ethnischen Enklaven wohnen, bessere Ergebnisse am Arbeitsmarkt erzielen.

---

<sup>7</sup> Die einschlägige empirische Literatur ist allerdings durch ein Übergewicht an Studien gekennzeichnet, die sich mit Zuwanderern befassen. Soweit beide Geschlechter untersucht werden, finden sich häufig quantitativ oder qualitativ substantielle Unterschiede. Deswegen sollte auch im geplanten Untersuchungsvorhaben durchweg möglichen Disparitäten zwischen Frauen und Männern nachgegangen werden.

**Tabelle 5.2 Empirische Studien mit signifikantem Zusammenhang zwischen Integrationsindikator und Arbeitsmarktintegration von Zugewanderten**

	<b>Integrations- indikator</b>	<b>Arbeitsmarktindikator</b>	<b>Land</b>	<b>Studie</b>
Bildung	Bildungsniveau	Löhne	Vereinigte Staaten	Borjas (2005)
	Bildung im Herkunfts- vs. Aufnahmeland	Lohnlücke zu Einheimischen	Deutschland	Basilio und Bauer (2014)
	Humankapital	Beschäftigungslücke zu Einheimischen	Schweden	Luik et al. (2016)
	Überqualifizierung	Löhne	Spanien	Kalfa und Piracha (2013)
	Bildungsniveau	Risiko überqualifizierter Beschäftigung	Deutschland	Schwientek (2016)
	Anerkennung ausländischer Abschlüsse	Beschäftigung / Löhne	Deutschland	Brücker et al. (2014), Brücker et al. (2015b)
	Überqualifizierung	Löhne	Spanien	Kalfa und Piracha (2013)
Sprache	Sprachkenntnisse	Löhne	Deutschland	Dustmann und van Soest (2002)
	Sprachkenntnisse	Löhne	Deutschland	Dustmann (1994)
	Sprachkenntnisse	Löhne / Beschäftigung	Vereinigtes Königreich	Dustmann und Fabbri (2003)
	Gebrauch der Sprache des Aufnahmelands	Löhne / Beschäftigung	Deutschland	Aldashev et al. (2009)
	Sprachkenntnisse	Lohnlücke	Vereinigtes Königreich	Miranda und Zhu (2013)
	Sprachkenntnisse	Löhne / soziale Integration	USA	Bleakley und Chin (2004)
	Sprachkenntnisse	Beschäftigung / Löhne	Niederlande	Yao und van Ours (2015)
Soziale Beziehungen	Politische Integration Familiäre Netzwerke	Einkommen	Deutschland	Danzer und Ulku (2011)
	Kontakte zu Einheimischen	Beruflicher Status	Deutschland	Kanas et al. (2012)
	Soziales Netzwerk	Beschäftigung / Löhne	Australien	Kalfa et al. (2016)
	Soziale Kontakte	Beschäftigung / Löhne	Niederlande	Chiswick und Wang (2016)
Wohnumfeld	Ethnische Enklave	Löhne	Kanada	Warman (2007)
	Ethnische Enklave	Beschäftigung / Löhne	USA	Pedace und Kumar (2012)
	Zuwandererdichte	Beschäftigung	Italien	Boeri et al (2015)
	Ethnische Enklave	Segregation	Deutschland	Danzer und Yaman (2012)
	Zugang zu Mobilität	Beschäftigung	USA	Gautier und Zenou (2010)
Akkulturation	Ethnische Identität	Löhne	Deutschland	Zimmermann (2007)
	Ethnische Identität	Beschäftigung / Erwerbsbeteiligung	Deutschland	Constant und Zimmermann (2009)
	Äußere Erscheinung	Einstellungschance	Deutschland	Weichselbaumer (2016)
Sicherheit und Stabilität	Aufenthaltsstatus	Beschäftigungsumfang Selbstständigkeit	Deutschland	Constant und Zimmermann (2005)
	Rückkehrerwartung	Löhne / Ausbildung	Deutschland	Dustmann (1993)
	Naturalisierung	Löhne	Deutschland	Gathmann und Keller (2014)

Darstellung des IZA.

Verschiedene kausalanalytisch angelegte Studien unterstützen die Hypothese, dass mehr Sicherheit und Stabilität, was die Erwartung des Verbleibs im Aufnahmeland angeht, im Migrationskontext zu einer besseren Position im Arbeitsmarkt führt. Als Kenngrößen für diese Integrationsdimension wurden in Studien mit deutschen Daten beispielsweise die Art des Aufenthaltsstatus, die selbst geäußerte Rückkehrintention und das Erlangen der deutschen Staatsangehörigkeit verwendet. Schließlich gibt es empirische Evidenz, dass ethnische Distanz die Arbeitsmarktchancen verschlechtern kann. So beobachten Constant und Zimmermann (2008) Unterschiede zwischen Personen, die eine starke ethnische Identität in dem Sinne aufweisen, dass sie sich sowohl mit ihrer Heimat als auch mit der Aufnahmegesellschaft identifizieren, und Personen, bei denen das Zugehörigkeitsgefühl nur in einer Hinsicht ausgeprägt ist.<sup>8</sup>

Schlecht empirisch dokumentiert ist der Einfluss des Gesundheitszustands von Migrantinnen und Migranten auf die Arbeitsmarktergebnisse. Die Forschung richtet ihre Aufmerksamkeit hier mehr auf den sogenannten „Healthy Migrant-Effekt“ – Indikatoren für den Gesundheitszustand zeigen in vielen Kontexten an, dass Zuwandererbevölkerungen tendenziell gesünder sind als die Aufnahmebevölkerung – und den Einfluss der Erwerbstätigkeit auf den Gesundheitszustand, durch den sich ein schwindender „Healthy Migrant-Effekt“ mit fortschreitender Dauer des Aufenthalts erklären ließe (Jasso et al. 2004). Allgemein ist der signifikant negative Einfluss eines schlechten Gesundheitszustands auf Beschäftigungsraten und Löhne allerdings gut empirisch dokumentiert (Pelkowski und Berger 2004).

Insgesamt betrachtet lässt sich aus dem Überblick über die einschlägige empirische Forschung der Schluss ziehen, dass die Elemente, die in der Synthese von Ager und Strang (2008) die Dimensionen von sozialer Teilhabe und Integration ausmachen, auch eine prädiktive Qualität im Hinblick auf das im Kontext des geplanten Evaluationsvorhabens übergeordnete Ziel der Arbeitsmarktintegration entfalten. Somit ist eine Schätzung des kausalen Einflusses der zu untersuchenden arbeitsmarktpolitischen Integrationsmaßnahmen auf Zieldimensionen der sozialen Integration nicht nur für sich genommen informativ. Sie ermöglicht auch Tendenzaussagen zur – indirekten – Wirkung auf die Position im Arbeitsmarkt, die angesichts der relativ großen Hürden bei der ökonomischen Integration der Flüchtlinge innerhalb des vorgesehenen Untersuchungszeitraums möglicherweise nur schwach zutage tritt. Dass sich die Stärke des zu erwartenden Einflusses der Indikatoren für soziale Integration auf die ökonomische Integration nicht aus der in anderen Kontexten gewonnenen Evidenz ableiten lässt, spricht dafür, einen breit angelegten Zielkatalog zu verwenden. Die Umsetzung dieses Ansatzes setzt aber voraus, dass sich die darin enthaltenen Zielgrößen nicht nur operationalisieren, sondern auch mit wirtschaftlich vertretbarem Aufwand messen lassen.

#### 5.4 Operationalisierung und Erhebung nicht administrativ erfasster Zielgrößen

Kennziffern, die sich zur Operationalisierung der oben diskutierten Zieldimensionen sozialer Teilhabe und Integration jenseits der Arbeitsmarktintegration eignen, lassen sich aus den für die Evaluationsaufgabe nutzbaren administrativen Individualdaten praktisch nicht ableiten. Aussagekräftige Indikatoren müssten daher vielmehr über Befragungen der Zielpersonen gewonnen werden.

Hierbei können sowohl objektive als auch subjektive Informationen zur Annäherung der Zielgrößen abgefragt werden. Unter „objektiven“ Informationen verstehen wir dabei Daten, die auf ein bestimmtes, prinzipiell überprüfbares Faktum zielen. Beispiele hierfür wären etwa die Zahl der Arztbesuche in einem vorgegebenen Zeitraum oder das in einem Sprachstandstest erzielte Punktergebnis. Im Gegensatz dazu

---

<sup>8</sup> In dieser Studie wird die doppelte Identifikation von Migrantinnen und Migranten als „Integration“ bezeichnet, die Identifikation ausschließlich mit dem Aufnahmeland dagegen als „Assimilation“ und die Identifikation ausschließlich mit dem Herkunftsland als „Absonderung“ (englisch: *separation*).

geht es bei „subjektiven“ Informationen um persönliche Einschätzungen des Befragten. Beispiele hierfür wären das gesundheitliche Wohlbefinden oder eine Selbstbewertung der erreichten Fähigkeit, in Deutsch mit anderen zu kommunizieren.

Mit Selbsteinschätzungen sind jedoch besondere Messprobleme verbunden. Erstens sind sie allgemein in der Regel nur verlässlich, wenn sie sich auf einen aktuellen Zustand („Wie bewerten Sie dies heute?“) oder die Veränderungsrichtung eines Zustands („Verglichen mit der Situation vor drei Monaten, ist es heute besser, schlechter oder gleich?“) bezogen werden. Darüber hinaus kann sich das Problem ergeben, dass verwendete Skalen in Abhängigkeit vom kulturellen Hintergrund systematisch unterschiedlich interpretiert werden. Somit sollten nur Abweichungen vom Gruppendurchschnitt verwendet werden, wofür aber eine ausreichende Anzahl von Beobachtungen je Gruppe zur Verfügung stehen muss.

Diese Einschränkungen sprechen zunächst dafür, zur Operationalisierung der Zielgrößen objektive Kennziffern zu bevorzugen. Dies hat jedoch auch Grenzen: bestimmte Größen sind erfahrungsgemäß schwer messbar (etwa Einkommens- oder Vermögenstatbestände oder bestimmte Fähigkeiten und Kompetenzen); zudem kann der empirische Erklärungsgehalt von objektiven Kennziffern im Hinblick auf die damit zu operationalisierende Zielgröße (etwa die Zahl der Arztbesuche als Proxy für den Gesundheitszustand) schwach bleiben.

Darum ist es zu empfehlen, zur Operationalisierung der im Rahmen des Evaluationsvorhabens nur per Erhebung zu gewinnenden Zielgrößen sowohl subjektive als auch objektive Kenngrößen heranzuziehen. Darüber hinaus ist zu empfehlen, sich zur Konkretisierung der Erhebungsinstrumente eng an das Erhebungsprogramm der bereits laufenden IAB-BAMF-SOEP-Befragung von Geflüchteten anzulehnen.<sup>9</sup> Die Nutzung dieser Benchmark bringt verschiedene Vorteile.

Erstens liegen hiermit – in den sieben mit Blick auf die zu erwartende Zusammensetzung der Zielpersonen wichtigsten Sprachen – in der Praxis bereits erprobte Items vor. Im Entwicklungsprozess der IAB-BAMF-SOEP-Befragung wurden die darin eingesetzten Items bereits methodisch geprüft und eingeordnet. Zudem wurde ihre Verwertbarkeit sowohl durch einen rigorosen Pretest als auch durch ergänzende qualitative Begleitforschung überprüft. Wichtige Items lehnen sich dem Gehalt nach zudem an bereits länger laufende Verlaufsstudien, insbesondere an die IAB-SOEP-Migrationsstichprobe, an. Sie fügen sich somit empirisch wie methodisch in einen größeren Kontext ein. Die vorgeschlagene Übernahme dieser Items kann demnach den Aufwand zur Entwicklung des Erhebungsinstruments erheblich vermindern und ermöglicht so, im Rahmen des geplanten Evaluationsvorhabens schneller in die Feldphase überzugehen.

Zweitens ist ein starkes Argument für die Übernahme von Items aus der IAB-BAMF-SOEP-Befragung von Flüchtlingen, dass die im Rahmen des Evaluationsvorhabens von ausgewählten Zielpersonen erhobenen Daten gegen diese – repräsentativ angelegte – Benchmark abgeglichen werden können. Dieser Abgleich kann helfen, die Selektivität der untersuchten Zielgruppen zu beurteilen und spezifische Entwicklungsverläufe besser zu erkennen.

---

<sup>9</sup> Eine weitere laufende Längsschnitterhebung, mit der gewisse Verlaufsentwicklungen bei geflüchteten Menschen beschrieben werden können, ist das Panel „Arbeitsmarkt und Soziale Sicherung“ (PASS). In PASS wurden per Oversampling Haushalte von Geflüchteten aus Syrien und dem Irak aufgenommen, in denen 500 bis 600 Personen leben. Das Erhebungsinstrument des PASS ist jedoch wenig auf die besondere Zielgruppe zugeschnitten und eignet sich daher anders als die IAB-SOEP-BAMF-Befragung kaum als Benchmark für das zu erstellende Erhebungsprogramm. In Anbetracht der Fallzahlen und der Selektivität der erfassten geflüchteten Menschen ist PASS auch für die Zwecke der Evaluation offensichtlich ungeeignet.

**Tabelle 5.3**      **Vorschlag einer Item-Auswahl zur Operationalisierung von nicht administrativ erfassten Zielgrößen des Evaluationsvorhabens mit Hilfe der IAB-BAMF-SOEP-Befragung von Flüchtlingen**

<b>Zielgröße</b>	<b>Item in IAB-BAMF-SOEP-Befragung von Geflüchteten</b>
Wohnen	Art der Wohnung (privat / Gemeinschaftsunterkunft)
	Allgemeine Zufriedenheit mit der Wohnsituation
	Zufriedenheit mit verschiedenen Aspekten der Wohnung (u.a. Sicherheit der Nachbarschaft, Zugang zu öffentlichem Verkehr, Qualität des Umfelds, Privatsphäre und Störfaktoren, Zugang zu Angeboten in der Umgebung)
Bildung	Teilnahme an Sprachkursen
	Teilnahme an Schul-, Hochschul-, Aus- oder Weiterbildung in Deutschland
Gesundheit	Subjektive Einschätzung des Gesundheitszustands
	Zufriedenheit mit und Sorgen um die Gesundheit
	Leiden unter Depressionen
Sozialer Austausch	Anzahl der neuen Kontakte zu Deutschen seit Ankunft
	Aktuelle Intensität der Kontakte mit Deutschen
Soziale Bindungen	Anzahl der neuen Kontakte zu Personen aus den Herkunftsländern seit Ankunft
	Aktuelle Intensität der Kontakte mit Personen aus den Herkunftsländern
	Familienangehörige im Haushalt
	Einsamkeit
Soziale Dienste	Nutzung von Beratungsstellen und -angeboten für Geflüchtete
	Inanspruchnahme von Hilfsangeboten in verschiedenen Lebensbereichen (Asylfragen, Arbeitssuche, Wohnungssuche u.a.)
Sprache und kulturelles Wissen	Deutschkenntnisse bei Einreise
	Selbsteinschätzung der aktuellen Deutschkenntnisse
	Bekanntheit von Beratungsstellen und Unterstützungsangeboten
Sicherheit und Stabilität	Willkommensgefühl
	Absicht des langfristigen Aufenthalts in Deutschland
	Häufigkeit von Diskriminierungserfahrungen in verschiedenen Bereichen
	Finanzielle Unterstützung anderer Familienangehöriger / Rücküberweisungen
Rechtliche Basis	Kosten und Finanzierung der Flucht
	Status und Dauer des Asylverfahrens
Eigene Integrationsziele	Rechtlicher Aufenthaltsstatus
	Angestrebte Schul- und berufliche Bildungsabschlüsse
Selbstwirksamkeit	Erwerbsaspiration von Nicht-Erwerbstätigen
	Kontrollüberzeugung
	Selbstwert

Darstellung des IZA.

Die Fragebögen der laufenden IAB-BAMF-SOEP-Befragung von Geflüchteten sind bislang nicht öffentlich und konnten leider auch im Rahmen der Durchführung der Machbarkeitsstudie nicht eingesehen werden.<sup>10</sup> Sie sollen nach derzeitigem Planungsstand aber mit der geplanten Veröffentlichung des Public Use-Files im späteren Jahresverlauf 2017 allgemein zugänglich gemacht werden und wären damit auch für eine Nutzung im Rahmen des geplanten Evaluationsvorhabens rechtzeitig verfügbar.

Wegen des fehlenden Zugangs zu den Fragebögen kann derzeit kein abschließender Vorschlag für eine Auswahl aus den darin enthaltenen rund 450 Items zu individuellen Merkmalen und rund 100 Items zu Merkmalen des Haushalts unterbreitet werden. Jedoch erlaubt die umfangreiche Veröffentlichung erster Ergebnisse der IAB-BAMF-SOEP-Befragung von Geflüchteten (IAB 2016) schon jetzt eine fortgeschrittene vorläufige Aufstellung von Variablen, die sich zur Operationalisierung der vorgeschlagenen, über die Arbeitsmarktintegration hinausgehenden Zieldimensionen eignen (Tabelle 5.3). Die erklärenden Qualitäten der ausgewählten Variablen zeigen sich in der IAB-Veröffentlichung im Rahmen einiger Regressionsanalysen, in denen Maße der sozialen Teilhabe oder die Teilnahme an Integrationskursen durch Geflüchtete als abhängige Variable modelliert sind.

Mit der Entscheidung für eine Orientierung an den teils nicht einfachen Items der IAB-BAMF-SOEP-Befragung von Flüchtlingen ist die Entscheidung verbunden, die Erhebung zumindest bei der Erstsprache in Form von „computerassistierten persönlichen Interviews“ (CAPI-Interviews) durchzuführen. Bei dieser Form der Erhebung wird die zu befragende Person von einer Interviewerin oder einem Interviewer persönlich aufgesucht; der Fragebogen wird dann gemeinsam an einem Computerbildschirm durchlaufen. So lassen sich auch komplexe Sachverhalte abfragen und Verständigungshürden überwinden, mit denen wegen der besonderen Bedingungen der Zielgruppe der Erhebung in besonderem Maße zu rechnen ist. Im Rahmen der IAB-BAMF-SOEP-Befragung wurden dazu auch innovative audiounterstützte Erhebungsinstrumente eingesetzt, um etwa Personen mit eingeschränktem Verständnis von Schriftsprache zu erreichen. Dies könnte auch für die geplante Erhebung von Nutzen sein.

Wird ein CAPI-Format umgesetzt, sind in der Durchführung die Fixkosten pro Interview im Verhältnis zu den variablen Kosten sehr hoch. Hieraus ergibt sich andererseits, dass auch ein umfangreiches Fragenprogramm noch wirtschaftlich abgewickelt werden kann. Dies spricht dafür, die Item-Auswahl nicht zu sehr zu beschränken. Somit lässt sich unserer Einschätzung nach auch das in Tabelle 5.3 zusammengestellte breite Programm umsetzen, womit ex ante anhand von konzeptionellen oder empirischen Überlegungen nicht sicher angemessen zu treffende Festlegungen vermieden werden können.

Eine Engführung des Erhebungsprogramms mit dem Ziel, spätere Erhebungen auch kostengünstiger in Form von „computerassistierten Telefoninterviews“ (CATI-Interviews) durchführen zu können, sollte erst im Verlauf der Untersuchung auf Basis der im Rahmen der CAPI-Erhebung gewonnenen Daten erfolgen. So kann auf Grundlage statistischer Kriterien geklärt werden, in welchem Ausmaß Items, die einen ähnlichen inhaltlichen Sachverhalt annähern, also etwa die vier ausgewählten Kenngrößen für die Zieldimension „soziale Bindungen“, unabhängig voneinander variieren. Größen, die nicht genügend unabhängige Varianz aufweisen, sind für die empirische Analyse verzichtbar; sie können naturgemäß keinen eigenen Erklärungsbeitrag (auch im Hinblick auf die übergeordnete Zielgröße Arbeitsmarktintegration) leisten.

---

<sup>10</sup> Diesbezügliche Versuche scheiterten an den eingegangenen vertraglichen Verpflichtungen der an der Durchführung der Studie beteiligten Forschungsinstitutionen.

Die im Verlauf der Untersuchung auf Grundlage von Cluster- und Varianzanalysen vorzunehmende technische Verdichtung des Fragenprogramms braucht sich dabei nicht auf Items zu beschränken, die inhaltlich einer bestimmten Zieldimension zuzurechnen scheinen. Sie kann grundsätzlich auch über Dimensionen hinweg erfolgen und so im Ergebnis auch eine inhaltliche Neuordnung der zu untersuchenden Zielgrößen herbeiführen.

Die Konzentration und Weiterentwicklung des Fragenprogramms im Fortgang der Analysen sollte allerdings nicht nur anhand von statistischen Kriterien vorgenommen werden. Es gilt auch inhaltliche Erkenntnisse aufzunehmen, die im Verlauf des Vorhabens, etwa durch die begleitenden qualitativen Untersuchungen oder durch die Verknüpfung von Erhebungs- und administrativen Daten, gewonnen werden.

Im Übrigen sollte der Fragenkatalog der IAB-BAMF-SOEP-Befragung von Flüchtlingen nicht nur für die Erhebung der im geplanten Evaluationsvorhaben zu untersuchenden Zielgrößen herangezogen werden. Er enthält darüber hinaus umfangreiche Items zur Erhebung von Hintergrundvariablen, die relevante Einflussgrößen bei der Erklärung der individuellen Wahrscheinlichkeit der Maßnahmenteilnahme darstellen könnten, aber auch systematische verstärkende oder hemmende Einflussgrößen von Maßnahmeneffekten, die in Analysen zu Wirkungsheterogenitäten Beachtung verdienen. Entsprechende Items finden sich etwa im Zusammenhang mit den Themenkomplexen Herkunft und Weg nach Deutschland, Schul-, Hochschul- und Berufsbildung im Ausland, Erwerbstätigkeit und Einkünfte im Ausland, Einstellungen und Werte, sowie Persönlichkeit.

## 6. Methodisches Vorgehen

### 6.1 Das Fundamentalproblem der empirischen Wirkungsforschung

Dieses Kapitel befasst sich mit der Frage, mit welcher Vorgehensweise im vorliegenden Kontext aussagekräftige empirische Befunde zur Wirkung der zu evaluierenden Maßnahmen gewonnen werden können, und wie sich aus diesen Befunden möglichst praxisorientierte und konkrete Handlungsempfehlungen ableiten lassen. Um diese Ziele zu erreichen, ist eine Verschränkung von kausalanalytisch-statistischer und qualitativ orientierter Herangehensweise sinnvoll.

Allgemein gesprochen ist es das Ziel kausalanalytischer Wirkungsforschung, die Wirksamkeit einer Maßnahme bzw. eines Treatments bei den Teilnehmerinnen und Teilnehmern abzuschätzen. Die Herausforderung dabei ist, ausschließlich nur diejenigen Veränderungen bei den vorgegebenen Zielgrößen zu messen, die ursächlich auf die Teilnahme an der Maßnahme zurückgehen. Hierbei geht es um die Beantwortung einer kontrafaktischen, also hypothetischen Frage: Wie wäre die betrachtete Zielgröße für die Teilnehmerinnen und Teilnehmer einer Maßnahme ausgefallen, wenn diese nicht an der Maßnahme teilgenommen hätten?

Die Antwort auf diese Frage lässt sich unter bestimmten Voraussetzungen abschätzen. Dazu wird an die Stelle des Ergebnisses von Teilnehmerinnen und Teilnehmern im Fall der Nichtteilnahme, das sich naturgemäß nicht beobachten lässt, das beobachtbare Ergebnis von Personen gesetzt, die tatsächlich nicht an der Maßnahme teilgenommen haben. Hierfür sind jedoch nur Beobachtungen von solchen Nichtteilnehmerinnen und Nichtteilnehmern geeignet, die den Personen, die an der Maßnahme teilnehmen, hinreichend ähneln. Genauer gesagt dürfen sich die Treatment-Gruppe und die Vergleichsgruppe bis auf die Tatsache der Maßnahmenteilnahme lediglich zufällig voneinander unterscheiden. Mit anderen Worten: Die Teilnahme an der Maßnahme muss innerhalb der Gruppe von Personen, die zur Abschätzung der Maßnahmenwirkungen herangezogen wird, ein so gut wie experimentell herbeigeführtes Ereignis sein.

Unterschieden sich Teilnehmerinnen und Teilnehmer von den Nichtteilnehmerinnen und Nichtteilnehmern dagegen im Hinblick auf Merkmale, die sowohl die Wahrscheinlichkeit der Teilnahme an der Maßnahme als auch die untersuchte Zielgröße beeinflussen, kann ein Vergleich zwischen den beiden Personengruppen die Effekte der Maßnahme nur verzerrt wiedergeben. Im Kontext der Evaluation von Instrumenten der aktiven Arbeitsmarktpolitik ist mit positiven wie auch mit negativen Verzerrungen zu rechnen. Beispielsweise könnten erfolgsorientierte Vermittlerinnen und Vermittler die zur Verfügung stehenden Instrumente bevorzugt bei Kundinnen oder Kunden einsetzen, deren Arbeitsmarktchancen ohnehin besonders gut sind („cream skimming“). In dieser Konstellation würden die Teilnahmeeffekte systematisch überschätzt, weil die Treatment-Gruppe im Vergleich zur Kontrollgruppe eine Positivauswahl darstellt. Andererseits könnten Vermittlerinnen und Vermittler mit dem Ziel, einen Nachteilsausgleich zu schaffen, ihre Instrumente auf Kundinnen und Kunden konzentrieren, die am Arbeitsmarkt unabhängig von der Teilnahme an der Maßnahme besonders schlechte Ergebnisse erzielen. In dieser Konstellation würden die Teilnahmeeffekte systematisch unterschätzt, weil es sich bei der Treatment-Gruppe im Vergleich zur Kontrollgruppe um eine Negativauswahl handelt.

Solange Maßnahmen nicht, analog zur Vorgehensweise in Laborexperimenten, von vornherein unter kontrollierten Bedingungen rein zufällig innerhalb der Zielgruppe zugeteilt werden, müssen denkbare Verzerrungen in Folge von Selektivitäten im Zugangsprozess mittels geeigneter statistischer Verfahren nachträglich kontrolliert werden. Hierfür steht eine Reihe unterschiedlicher Methoden der Programmevaluation zur Verfügung (Bauer et al. 2009; Imbens und Wooldridge 2009; Angrist und Pischke



2010). Diese unterscheiden sich hinsichtlich der Voraussetzungen, die für eine Identifikation der gesuchten unverzerrten Programmeffekte erfüllt sein müssen, und stellen damit unterschiedliche Anforderungen an die Daten.

Grundsätzlich ist zu empfehlen, das Evaluationsproblem mit möglichst vielen der plausibel umsetzbaren statistischen Evaluationsmethoden anzugehen. Dies heißt, sowohl Verfahren einzusetzen, die darauf basieren, systematische Selektionsprozesse bei der Maßnahmenteilnahme anhand von beobachtbaren Größen zu beschreiben, als auch Verfahren, die so gut wie experimentelle Unterschiede beim Maßnahmenzugang ausschöpfen. Wählt man einen offenen Multi-Methoden-Ansatz, statt sich auf ein einziges statistisches Verfahren festzulegen, eröffnen sich Möglichkeiten, die Robustheit der geschätzten Maßnahmenwirkungen besser einzuschätzen. Führen nämlich verschiedene Methoden zu zahlenmäßig ähnlichen, qualitativ übereinstimmenden Befunden, ist dies ein Hinweis, dass die erzielten Resultate nicht von den hinter jeder Methode liegenden, letztlich nicht testbaren Annahmen abhängen. Darüber hinaus sollte sich eine Evaluation nicht allein auf den Einsatz statistischer Verfahren beschränken, sondern vielmehr quantitative und qualitative Evaluationsweisen miteinander kombinieren. Auch dies kann zur Validierung der statistisch gewonnenen kausalanalytischen Aussagen beitragen; dazu sind mit den Verfahren der qualitativen Sozialforschung die statistischen Befunde an den Erfahrungen, die in der Praxis von den Stakeholdern gemacht werden, zu spiegeln und plausible Erklärungsmuster dafür zu finden, wie die abgeschätzten Maßnahmenwirkungen zustande kommen.

## 6.2 Methoden zur Gewinnung kausalanalytischer Aussagen

Der folgende Abschnitt diskutiert verschiedene statistische Verfahrensweisen, die zur Schätzung der gesuchten Maßnahmeneffekte im vorliegenden Kontext grundsätzlich in Frage kommen. Inwieweit die zur Identifikation von Kausalaussagen zu erfüllenden Bedingungen in der späteren Anwendung plausibilisiert werden können, lässt sich erst nach einer Verbindung mit den Daten sagen. Berücksichtigt werden im Folgenden nur Verfahren, die ausschließlich Daten von Personen im Kontext von Fluchtmigration verwenden. Diese Beschränkung trägt den sehr besonderen Bedingungen der Zielgruppe Rechnung. Es ist davon auszugehen, dass diese gegenüber der übrigen Bevölkerung strukturell erheblich im Nachteil ist; daraus resultierende Niveauunterschiede bei den Zielgrößen lassen sich bei einem Vergleichsgruppenansatz zwar grundsätzlich durch eine Fokussierung auf Veränderungen im Zeitablauf korrigieren („Differenz-in-Differenzen“-Technik). Weil sich die Zielgruppe noch nicht lange in Deutschland aufhält, gibt es jedoch keine Möglichkeit, die für diese Methode konstituierende Annahme gemeinsamer Trends – ohne die Intervention hätten sich die Zielgrößen bei Treatment- und Vergleichsgruppe also gleich entwickelt – adäquat zu fundieren.

### 6.2.1 Selektion anhand beobachtbarer Variablen

Ein zur Evaluation aktiver arbeitsmarktpolitischer Maßnahmen häufig verwendeter Vergleichsgruppenansatz ist das statistische Matching (Heckman et al. 1997). Diese Methode vergleicht Zielindikatoren für „statistische Zwillinge“ von Teilnehmenden und Nichtteilnehmenden, die sich dadurch auszeichnen, dass sie aufgrund der beobachteten individuellen Merkmale die gleiche oder zumindest ähnliche Wahrscheinlichkeit der Teilnahme aufweisen.

Die zentrale dem Matching-Verfahren zugrunde liegende Annahme lautet: Gegeben einen Satz von beobachtbaren individuellen Merkmalen, der eine bestimmte Wahrscheinlichkeit der Maßnahmenteilnahme begründet, ist die tatsächliche Teilnahme oder Nichtteilnahme rein zufällig, so dass keine Selektionsverzerrung entsteht („conditional independence“- oder „selection on observables“-Annahme). Diese Annahme lässt sich nicht testen, sondern muss anhand der in die Schätzung

einbezogenen individuellen Merkmale plausibel begründet werden. Eine zweite Voraussetzung für die Anwendung des Verfahrens ist, dass „common support“ gegeben ist. Der Treatment-Effekt ist nur für den Bereich von Teilnahmewahrscheinlichkeiten (gegeben die beobachteten Merkmale) identifiziert, in dem es sowohl Teilnehmende als auch Nichtteilnehmende gibt, die zueinander „gematcht“ werden können.

Matching-Verfahren sind datenhungrig. Im Kontext der Evaluation aktiver arbeitsmarktpolitischer Maßnahmen werden oft umfassende retrospektive Verlaufsdaten aus administrativen Quellen eingesetzt, um die dahinter liegende Annahme der Selektion auf Basis beobachtbarer Daten zu rechtfertigen (IZA et al. 2006, ZEW et al. 2006). Die Nutzung biografischer Vorerfahrungen bis zum Eintritt in die Maßnahme kontrolliert zumindest indirekt den Einfluss mutmaßlich für Selektionsprozesse relevanter Größen, die schwierig zu beobachten sind (wie etwa Persönlichkeit, nicht-kognitive Fähigkeiten, persönliche Hemmnisse beim Arbeitsmarktzugang).

Bezogen auf die Zielgruppe der Personen im Kontext von Fluchtmigration lässt sich ein Matching-Ansatz allein auf Basis der für die Evaluation verfügbaren administrativen Daten kaum überzeugend implementieren. Weil diese Personen erst seit kurzem erfasst sind, erscheint der aus den Prozessdaten zu generierende Datenkranz derzeit noch zu klein, um damit die Annahme der Selektion anhand beobachteter Daten plausibel zu begründen. Diese Einschätzung modifiziert sich allerdings, wenn zu den administrativen Daten per Befragung der Zielgruppe zu gewinnende, umfassendere individuelle Hintergrundvariablen hinzugespielt werden, wie etwa zur Vorgeschichte bis zur Ankunft in Deutschland, zu Bildungs- und Berufserfahrungen im Herkunftsland, zur aktuellen materiellen Lage, zu Einstellungen, Werten und Ambitionen oder zu Persönlichkeitsmerkmalen. Es ist davon auszugehen, dass derartige Größen einen substantiellen Beitrag zur Aufklärung der Wahrscheinlichkeit des Zugangs zu arbeitsmarktpolitischen Integrationsmaßnahmen leisten können – diese Einschätzung unterstützen auch die Ergebnisse entsprechender Regressionsmodelle, die im Rahmen der ersten vorläufigen Auswertung der IAB-BAMF-SOEP-Befragung von Geflüchteten vorgenommen wurden (IAB 2016).

Umsetzungsprobleme beim Matching-Ansatz auf Grundlage der verbundenen administrativen und Befragungsdaten könnten allerdings dadurch entstehen, dass die geplanten Fallzahlen an Teilnehmenden und Nichtteilnehmenden bezogen auf die einzelnen zu evaluierenden Maßnahmen nicht besonders groß sind. Dies kann insbesondere dann zu Problemen führen, wenn sich die per Befragung gewonnenen Merkmale – die administrativen Merkmale werden im Rahmen der Stichprobenbildung ausbalanciert – zwischen Treatment- und Vergleichsgruppe stark voneinander unterscheiden, so dass es an „common support“ fehlt. Dann wird der durchschnittliche kausale Maßnahmeneffekt nur für einen Ausschnitt der Teilnehmenden geschätzt; zudem könnte die statistische Power der Schätzergebnisse leiden, so dass signifikante Zusammenhänge nicht als solche erkannt werden.

Neben dem Matching-Verfahren lässt sich ein durchschnittlicher Treatment-Effekt auch mit einem linearen OLS-Modell schätzen, das einen Indikator für die Maßnahmenteilnahme und zugleich Variablen, die die Teilnahmewahrscheinlichkeit beeinflussen, enthält. Auch dieser Ansatz kann grundsätzlich eine unverzerrte Schätzung liefern. Dies setzt voraus, dass entweder das lineare Modell gültig und richtig spezifiziert ist, was die Selektion anhand von beobachteten Merkmalen als identifizierende Annahme einschließt, oder dass die Gruppen der Teilnehmenden und Nichtteilnehmenden ausbalanciert sind. Das OLS-Modell kann daher normalerweise immer dann zum Einsatz kommen, wenn auch der Matching-Ansatz sinnvoll ist (Angrist und Pischke 2009: 70). Die jeweils erzielten Resultate werden sich jedoch unterscheiden, weil die beiden Verfahren – bei nicht völlig ausbalancierten Gruppen von Teilnehmenden und Nichtteilnehmenden – die Fälle außerhalb des common support unterschiedlich gewichten. Der OLS-Ansatz ist, sofern das Modell korrekt spezifiziert ist, effizienter als der Matching-Ansatz, die geschätzten Parameter weisen also eine systematisch geringere Varianz auf. Dafür ist Matching im Hinblick auf die Schätzung heterogener Maßnahmenwirkungen im Vergleich flexibler.

Wegen dieser Unterschiede sollten beide Verfahren auf dieselben Daten angewendet werden, was einen Robustheitscheck ermöglicht.

## 6.2.2 Analysen auf Basis quasi-experimenteller Variation

Die Annahme einer Selektion anhand beobachtbarer Merkmale, wie sie beim Einsatz von Matching und linearem OLS-Modell getroffen werden muss, ist ziemlich stark. Als Alternative bieten sich empirische Strategien an, die kausale Maßnahmeneffekte auf der Grundlage quasi-experimenteller Variation hinsichtlich der Teilnahme schätzen. Die Identifikation wird in diesem Fall dadurch erzielt, dass die Teilnehmerinnen und Teilnehmer aufgrund von systematischen Faktoren so gut wie zufällig in eine Maßnahme gelangen, welche keinen direkten Einfluss auf die zu untersuchende Zielgröße ausüben. Um denkbare quasi-experimentelle Variation auszuschöpfen, bieten sich im vorliegenden Kontext unserer Einschätzung nach vor allem zwei Ansätze an: Verweildauerverfahren und die Ausschöpfung institutionell bedingt regionaler Variation in den Treatment-Wahrscheinlichkeiten.

Verweildauerverfahren verwenden Längsschnittinformationen zur Entwicklung von Individuen über die Zeit. Der Grundansatz besteht darin zu untersuchen, wie lange ein bestimmter Zustand – etwa die Arbeitslosigkeit nach Ankunft als Flüchtling in Deutschland – andauert, bis es zum Übergang in einen oder mehrere andere mögliche Zielzustände – etwa den Beginn einer Ausbildung oder einer Beschäftigung – kommt. Wie lange der Ausgangszustand dauert, kann von mehreren Faktoren abhängen: der in diesem Zustand bereits verbrachten Dauer („Dauerabhängigkeit“), persönlichen Merkmalen und ggf. der Teilnahme an einer Maßnahme.

Maßnahmeneffekte werden im Kontext von Verweildauermodellen über den so genannten „Timing-of-Events“-Ansatz geschätzt (Abbring und van den Berg 2003; Heckman und Singer 1984). Allgemein gesagt modelliert dieser Ansatz simultan die zeitliche Dynamik, mit der es zur Teilnahme an einer Maßnahme kommt, und die zeitliche Dynamik, mit der ein bestimmtes Zielereignis gegeben den Eintrittszeitpunkt der Maßnahme auftritt. Um kausale Wirkungsaussagen abzuleiten, genügt es, die Zeitdauern bis zum Eintritt der beiden Ereignisse (Eintritt in die Maßnahme, Übergang in den interessierenden Endzustand) zu kennen. Dagegen ist es nicht erforderlich, individuelle Merkmale zu beobachten, die zwar den Zugang in die Maßnahme steuern, nicht aber das Ergebnis der Maßnahme. Zudem werden die Maßnahmenwirkungen selbst dann nicht verzerrt geschätzt, wenn es nicht beobachtete Größen gibt, die sowohl die Selektion in die Maßnahme als auch ihr Ergebnis beeinflussen. Implizit bildet der Zeitpunkt der Intervention die relevanten Merkmale der Teilnehmerinnen und Teilnehmer nämlich vollständig ab. Deswegen stellen Wirkungsanalysen mit Verweildauermodellen eher niedrige Datenanforderungen. Dafür sind die ökonometrischen Ansprüche bei der Umsetzung ziemlich hoch. Auch ist die Identifikationsidee für Nichtfachleute eher schwierig verständlich, was bei der Kommunikation der erzielten Resultate nachteilig sein kann.

Der „Timing-of-Events“-Ansatz eignet sich grundsätzlich für Maßnahmen, bei denen die Betroffenen den zeitlichen Beginn vorher nicht exakt antizipieren können. Dies gilt oft für Maßnahmen, die von Fallmanagerinnen und Fallmanagern gesteuert werden. Auch im Kontext der arbeitsmarktpolitischen Integrationsmaßnahmen erscheint es wahrscheinlich, dass einander ähnliche Individuen ohne eigenes Zutun zu unterschiedlichen Zeitpunkten von derselben Maßnahme erreicht werden. Dafür können beispielsweise unterschiedlich organisierte Prozesse oder Strategien auf der Ebene der lokal zuständigen Institutionen verantwortlich sein, oder auch Nachfrageüberhänge, wie sie bei den berufsbezogenen Sprachkursen für Flüchtlinge zu vermuten sind.

Die Umsetzung des Timing-of-Events-Ansatzes im vorliegenden Kontext könnte allerdings an Grenzen stoßen. Da kalendarische Daten zum Verlauf der Vorgeschichte der Personen im Kontext von Fluchtmigration in Deutschland praktisch bis zu dem Zeitpunkt fehlen, an dem sie in den Zuständigkeitsbereich der BA übergehen, müsste dieser Übergang als Nullpunkt der Untersuchung gewählt werden. Diese zeitliche Verankerung des timing-of-events-Ansatzes könnte aber zu wenig Variation liefern, falls viele Flüchtlinge bereits relativ kurz danach in eine Maßnahme einmünden. Eine andere Begrenzung ist, dass mit dem Verfahren eigentlich nur Statusübergänge untersucht werden können, deren Zeitpunkte in den administrativen Quellen abgebildet werden. Die Erhebung hinreichend genauer kalendarischer Daten per Befragung, andere Zielgrößen betreffend, erscheint dagegen kaum praktikabel. Zudem bewegen sich einige der Zielgrößen, insbesondere viele Indikatoren sozialer Teilhabe, graduell, so dass sich nicht ohne weiteres sagen lässt, was ein Übergang von einem Ausgangszustand in einen anderen Zustand überhaupt ist.

Eine Alternative zur Ausschöpfung quasi-experimenteller Variation im Maßnahmenzugang geht von regionalen Unterschieden in der Intensität des Maßnahmeneinsatzes zwischen Arbeitsmarktregionen aus, die ansonsten durch strukturell übereinstimmende Arbeitsmarktbedingungen charakterisiert sind. Vorbilder für den Einsatz einer auf räumlicher Varianz beruhenden Strategie im Kontext der Evaluation aktiver arbeitsmarktpolitischer Maßnahmen liefern etwa die Evaluation der Experimentierklausel nach § 6c SGB II durch Arntz et al. (2006) oder die auf regionalen Unterschieden in Sperrzeitenquoten beruhende Analyse der Wirkungen von Sanktionierungen durch Boockmann et al. (2009).

Voraussetzungen dafür, dass regionale Variation in der Treatment-Intensität zu Evaluationszwecken genutzt werden kann, sind erstens, dass die Ursachen für die regionalen Treatment-Unterschiede nicht in direktem Zusammenhang mit den zu untersuchenden Erfolgsgrößen stehen, und zweitens, dass Angehörige der Zielgruppe sich nicht selbst durch räumliche Mobilität in Gebiete mit höherer oder niedrigerer Treatment-Intensität selektieren können.

Auch wenn Flüchtlinge bislang nur begrenzten Residenzpflichten unterliegen, dürfte die zweite Bedingung tendenziell erfüllt sein. Ihre Verteilung in der Fläche reflektiert zum Beispiel die stark von Zufälligkeiten geprägte Wohnortzuweisung in der ersten Aufnahmephase oder ethnische Netzwerkstrukturen – Faktoren, die vermutlich wenig systematisch mit der lokalen Praxis arbeitsmarktpolitischer Integrationsmaßnahmen korrelieren.

Weniger sicher ist, ob auch die erste Voraussetzung erfüllt werden kann. Allerdings gibt es zumindest für das Feld der allgemeinen aktiven Arbeitsmarktpolitik Hinweise, dass die geschäftspolitische Ausrichtung der zuständigen lokalen Akteure und arbeitsmarktorientierte Zielgrößen nur recht lose zusammenhängen (RWI und ISG 2006). Eine Erklärung ist, dass die dezentral agierenden Akteure zentral vorgegebene Instrumente unterschiedlich interpretieren. Eine andere Erklärung verweist auf lokale Traditionen und Besonderheiten, etwa spezielle Konstellationen der lokalpolitischen Akteure, eine historisch gewachsene Trägerlandschaft oder Begrenzungen von Kapazitäten. Informationen aus Gesprächen mit Praxisakteuren legen zumindest nahe, dass solche Faktoren auch für die integrationsorientierten Maßnahmen, die für Flüchtlinge je nach Wohnort in einer Region zugänglich sind, eine wichtige Rolle spielen.

Wenn so quasi-zufällige Variation in der Wahrscheinlichkeit des Maßnahmeneinsatzes besteht, lässt sich dies für kausale Wirkungsanalysen nutzen, indem die lokale Treatment-Intensität als sogenannte Instrumentvariable (IV) verwendet wird („Intention to Treat“-Ansatz; Angrist und Krueger 2001).<sup>11</sup> Dieser

---

<sup>11</sup> Als Instrumentvariable bezeichnet man eine Größe, die mit der Wahrscheinlichkeit des Treatments – hier: des Zugangs in eine Maßnahme – möglichst hoch korreliert ist, auf die zu analysierende Zielgröße aber keinen direkten Einfluss ausübt, also diese nur indirekt über ihren Einfluss auf die Treatment-Wahrscheinlichkeit beeinflusst.

Ansatz identifiziert die Wirkungen einer Maßnahme durch Gegenüberstellung von Regionen, die ähnliche sozio-ökonomische Merkmale, aber starke Unterschiede in der Treatment-Intensität aufweisen. Ein Treatment-Effekt wird abgeschätzt, indem – für Personen mit vergleichbaren Merkmalen – die Differenz der mittleren Werte der Zielgrößen in Regionen mit hoher und niedriger Treatment-Intensität um die Differenz der Treatment-Wahrscheinlichkeiten in den gegenübergestellten Regionen korrigiert wird. Der auf diese Weise geschätzte Treatment-Effekt gibt an, welchen Effekt die zu untersuchende Maßnahme auf diejenigen Personen entfaltet, die ausschließlich deshalb von der Maßnahme erreicht werden, weil sie in einer Region mit hoher Treatment-Intensität leben und nicht in einer Region mit niedriger Treatment-Intensität. Es handelt sich damit um einen sogenannten „lokalen durchschnittlichen Treatment-Effekt“ (LATE – local average treatment effect).

Ein solcher Ansatz hat wegen seiner regionalen Ausrichtung einige Vorteile. Erstens ermöglicht die Konzentration auf ausgewählte Regionen tiefergehende qualitativ orientierte Begleituntersuchungen vor Ort. Auch lassen sich potenziell wirkungsrelevante lokale Kontextfaktoren, insbesondere das lokale Angebot von integrationsfördernden Maßnahmen außerhalb des Zuständigkeitsbereichs des BMAS, durch qualitative Begleitforschung vertieft beschreiben, erfassen und auch statistisch verwerten.

Die Robustheit der mit einem Regionen-Ansatz erzielten empirischen Befunde könnte auf zweierlei Weise geprüft werden. Erstens kann die Schätzung der Treatment-Effekte für verschiedene Teilauswahlen aus den regionalen Einheiten durchgeführt werden, die sich im Hinblick auf die Arbeitsmarktbedingungen in der Region systematisch voneinander unterscheiden. Es ist ein Indiz für externe Validität, wenn dieselben Effekte in unterschiedlichen Kontexten vorgefunden werden. Zweitens wäre die Homogenität oder Heterogenität der geschätzten Wirkungen über verschiedene, anhand von beobachteten Merkmalen abgegrenzte Teilgruppen der Personen mit Migrationshintergrund zu untersuchen.

### 6.2.3 Evaluation von Maßnahmenketten

Die Ausgangsbasis der Evaluation bilden Personen im Kontext von Fluchtmigration, die innerhalb eines vorgegebenen Zeitraums in eine zu bewertende Maßnahme zugegangen sind und an einem bestimmten Stichtag noch teilnehmen. Dieser Treatment-Gruppe zugeordnet werden von der Statistik der BA registrierte Personen im Kontext von Fluchtmigration, die bis zu diesem Stichtag nicht in die zu bewertende Maßnahme zugegangen sind und auch nicht Teilnehmerin oder Teilnehmer an einer anderen Regelmaßnahme sind. Es ist denkbar und wahrscheinlich, dass sowohl in der Treatment- als auch in der Vergleichsgruppe im weiteren Verlauf der Entwicklung Teilnahmen an anderen Maßnahmen hinzukommen, sich also Maßnahmenketten entwickeln.

Soweit die ausgewählten Personen der Treatment-Gruppe später in weitere Maßnahmen einmünden, lässt sich prinzipiell jede Maßnahmenkette als eigenes Treatment definieren. Die Basis zur Bildung einer Vergleichsgruppe bilden alle Personen in der Treatment- und Kontrollgruppe, die exakt dieses Treatment nicht durchlaufen. Soweit die Annahme der Selektion in die bestimmte Maßnahmensequenz anhand beobachtbarer Merkmale gerechtfertigt werden kann, lässt sich ein spezifischer Treatment-Effekt mit den in Abschnitt 6.2.1 dargestellten Methoden schätzen. Das Ergebnis zeigt dann den durchschnittlichen Effekt des spezifischen Treatments im Verhältnis zu allen anderen Treatments an, welche die Teilnehmerin oder der Teilnehmer anstatt dessen auch hätte durchlaufen können. Die Begrenzung dieses Ansatzes liegt, neben der möglicherweise fehlenden Plausibilität der Selektion ausschließlich anhand beobachtbarer Merkmale, in unzureichenden Fallzahlen, weil durch Kombination verschiedener Maßnahmen sehr viele unterschiedliche Maßnahmenketten – also spezifische Treatments – entstehen könnten. Ein denkbarer Weg, um Fallzahlenprobleme zu begrenzen, ist das Dimensionalitätsproblem zu verringern, indem ähnliche Maßnahmenketten zusammengefasst werden oder nur besonders häufig vorkommende

Maßnahmenketten analysiert werden. Grundsätzlich wäre auch zu prüfen, ob sich quasi-exogene Variation in der regionalen Häufigkeit des Vorkommens ganzer Maßnahmenketten findet. Soweit dies der Fall ist, könnten auch Instrumentvariablen-Ansätze für die Evaluation eingesetzt werden.

Betrachtet man den Spezialfall, dass die Personen in der Teilnahmegruppe nur an der zu evaluierenden Maßnahme teilnehmen, also nicht in eine Maßnahmenkette einmünden, während die Personen in der Vergleichsgruppe im weiteren Verlauf in eine Maßnahmenkette einmünden, ändert sich nichts an der Vorgehensweise zur Schätzung von Maßnahmeneffekten. Es ändert sich aber wiederum die Interpretation der geschätzten Treatment-Effekte. Sie zeigen nicht, wie die Maßnahme an sich wirkt, das heißt im Vergleich zur kontrafaktischen Situation, dass die Teilnehmerinnen und Teilnehmer gar keine andere Maßnahme erfahren hätten. Vielmehr messen sie die Effekte der Maßnahmenteilnahme in Relation zur durchschnittlichen Zielerreichung in den Fällen, dass sie ansonsten später die anderen in der Vergleichsgruppe vorkommenden Maßnahmen oder Maßnahmenketten durchlaufen hätten. Selbst wenn eine Maßnahme für sich genommen einen Effekt erzielen würde, kann dieser relative Vergleich das Ergebnis fundamental verändern. Dies hängt ab von der Größe und der Richtung des durchschnittlichen Maßnahmeneffekts aller möglichen Treatments, die quasi zufällig ebenso zustande hätten kommen können, wenn die Teilnehmerin oder der Teilnehmer nicht bereits in die zu evaluierende Maßnahme eingemündet wäre.

Soweit es bereits vor der Unterscheidung von Treatment- und Kontrollgruppe unterschiedliche Maßnahmenverläufe gibt, ist diese Information bei der Auswahl der für den Evaluationszweck geeigneten Nichtteilnehmerinnen und Nichtteilnehmer zu berücksichtigen. Ausreichende Fallzahlen vorausgesetzt, lassen sich dann separate Treatment-Effekte in Abhängigkeit davon schätzen, welche Maßnahmenvorgeschichte die Teilnehmerinnen und Teilnehmer der zu evaluierenden Maßnahme aufweisen.

### 6.3 *Qualitativ orientierte Begleitforschung*

Bei vielen der vom BMAS durchgeführten Evaluationsvorhaben im Bereich der aktiven Arbeitsmarktpolitik werden qualitativ und quantitativ ausgerichtete Untersuchungsmethoden miteinander kombiniert. Dieses bewährte Prinzip empfiehlt sich auch im vorliegenden Kontext. Begleitende qualitative Implementationsanalysen sollen dabei erheblich mehr leisten, als Informationen zum Stand der Umsetzung von Maßnahmen bzw. Instrumenten zu liefern und Entwicklungstrends laufend zu beobachten.

Erstens haben sie die Funktion, Wirkungsanalysen so vorzubereiten, dass die eingesetzten empirischen Modelle die relevanten Wirkungskanäle in einer möglichst geeigneten Weise abbilden. Vorzunehmen ist deshalb eine umfassende Art der Implementationsanalyse, die im Sinne einer Prozessevaluation in der Lage ist, erfolgsrelevante Prozesse der systematischen Zuweisung in Maßnahmen offen zu legen, die anhand der vorliegenden Daten nicht oder nur schwer nachvollziehbar sind. Hierbei müssen auch mögliche Wechselwirkungen und Interdependenzen zwischen Einzelinstrumenten im Kontext mehrdimensionaler Zielsetzungen aufgezeigt werden. Weil das Feld der Integrationspolitik durch unübersichtliche Strukturen sich überlagernder Unterstützungsangebote und parallel handelnder Akteure gekennzeichnet ist, ist die Schaffung von Klarheit über die Zugangsprozesse besonders wichtig.

Eine zweite zentrale Funktion von begleitenden Implementationsanalysen ist es, die Grundlagen für eine adäquate Interpretation der Ergebnisse der quantitativen Wirkungsanalyse zu schaffen. Quantitative Verfahren können zwar die Effekte einer Intervention offen legen. Da für die kausalanalytischen Untersuchungen keine strukturellen Modelle eingesetzt werden, können sie die zu dem Befund führenden

Mechanismen aber nicht aufklären. So lässt sich nicht auseinander halten, ob eine Maßnahme unwirksam bleibt, weil sie in nicht geeigneter Weise geplant wurde, oder weil sie zwar richtig geplant, aber falsch umgesetzt wurde. Darum haben Implementationsanalysen die Aufgabe, die Qualität der Umsetzung und die Rahmenbedingungen der Maßnahmendurchführung zu beschreiben und zu bewerten. Daraus können Hypothesen zu Wirkungsblockern oder -verstärkern generiert werden, anhand derer die quantitativen Befunde gespiegelt und interpretiert werden können.

Implementationsanalysen im Sinne einer umfassenden Prozessevaluation sind demnach auch für das Forschungskonzept zur Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge ein unverzichtbarer Bestandteil. Hierfür können gängige Verfahren der Sozialforschung (leitfadengestützte Experteninterviews, Dokumentenanalysen, Netzwerkanalysen) genutzt werden. Zu empfehlen ist darüber hinaus eine integrierte externe Rückkopplung von (vorläufigen) Befunden der Wirkungsanalyse an Expertinnen und Experten aus der Praxis, um Arbeitshypothesen und Interpretationsmuster im Evaluationsprozess systematisch zu verfeinern.

## 7. Konzept zur Umsetzung der Evaluation

### 7.1 Vorüberlegungen

Aus der Prüfung der Voraussetzungen für eine Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge im Zuständigkeitsbereich des BMAS, deren Ergebnisse in den vorangegangenen Kapiteln der Expertise dargelegt wurden, ergeben sich eine Reihe von Punkten, die ein Konzept für die praktische Durchführung des Evaluationsvorhabens berücksichtigen muss:

- Die Förderung bzw. Teilnahme von Personen im Kontext von Fluchtmigration an Maßnahmen der aktiven Arbeitsförderung (§ 3 SGB III Abs. 2) und den Leistungen zur Eingliederung (§ 16 SGB II) lässt sich in den administrativen Verwaltungsdaten der BA inklusive der Jobcenter gut und zeitnah beobachten. Damit gibt es eine Grundlage für differenzierte Wirkungsanalysen für einzelne oder Gruppen von arbeitsmarktpolitischen Integrationsmaßnahmen. Dagegen sind die Maßnahmen zur Sprachförderung schwieriger zu evaluieren, weil daran unterschiedliche Träger beteiligt sind und somit keine vollständigen und konsistenten Daten zur Teilnahme zur Verfügung stehen.
- Allein auf Basis der zugänglichen administrativen Daten wäre eine Evaluation der Wirkungen der integrationspolitischen Maßnahmen im Zuständigkeitsbereich des BMAS nur sehr eingeschränkt machbar. Weil die Zielpersonen in Deutschland nur über einen kurzen Zeitraum erfasst sind und Informationen zu ihrer Vorgeschichte fehlen, beruhen kausale Wirkungsaussagen auf starken Annahmen, wenn die Selektion der Teilnehmerinnen und Teilnehmer in die Maßnahme nur anhand der wenigen in den administrativen Daten beobachtbaren individuellen Hintergrundvariablen statistisch beschrieben wird. Um zusätzliche Größen zu erfassen, die sowohl die Wahrscheinlichkeit des Maßnahmenzugangs als auch die Maßnahmenwirkungen beeinflussen können, empfiehlt sich eine Befragung von Teilnehmerinnen und Teilnehmern, sowie für Vergleichszwecke auch von Nichtteilnehmerinnen und Nichtteilnehmern. Eine solche Befragung ist auch deshalb erforderlich, weil viele Indikatoren für die Zielgrößen soziale Teilhabe und Integration in den administrativen Daten nicht abgebildet sind. Die Befragung muss darum als Wiederholungsbefragung angelegt werden, um Veränderungen bei den zu untersuchenden Ergebnisvariablen im Zeitablauf beobachten zu können.
- Die Durchführung der erforderlichen Erhebungen ist wegen der besonderen Zielgruppe der Personen im Kontext von Fluchtmigration mit ganz erheblichem Aufwand verbunden. So werden mehrsprachige Erhebungsinstrumente benötigt; um eine gute Datenqualität zu sichern, sollten möglichst face-to-face Interviews (und keine Telefoninterviews) durch speziell geschultes und mehrsprachiges Personal geführt werden. Um den Erhebungsaufwand zu begrenzen, empfiehlt sich erstens eine gezielte Auswahl der Stichprobe aus der Grundgesamtheit der arbeitssuchenden Personen im Kontext von Fluchtmigration. Zum einen sollte mit einer gematchten Personen-Stichprobe gearbeitet werden. Das heißt, ausgehend von den Maßnahmenteilnehmerinnen und Maßnahmenteilnehmern sind anhand personenbezogener Merkmale, die in der Statistik der BA verfügbar sind, im Hinblick auf die Teilnahmewahrscheinlichkeit möglichst ähnliche Nichtteilnehmerinnen und Nichtteilnehmer („statistische Zwillinge“) auszuwählen, die prinzipiell als Vergleichspersonen in Frage kommen. Zum anderen empfiehlt sich eine regionale Clusterung der Stichprobe. Das heißt, die Erhebung wird auf eine Auswahl von regionalen Einheiten (Agenturbezirke der BA) beschränkt, die geeignet sind, das Gesamtbild des Geschehens wiederzugeben. Dazu muss insbesondere sichergestellt sein, dass die ausgewählten Agenturbezirke die Bandbreite der Lagen am Arbeitsmarkt sowie der arbeitsmarkt- und integrationspolitischen Rahmenbedingungen vor Ort abdecken. Eine Nebenbedingung bei der Planung der Stichprobe ist, dass in jedem der ausgewählten regionalen Einheiten hinreichend Beobachtungseinheiten zur Verfügung stehen sollten, um prinzipiell eine Schätzung von



Maßnahmenwirkungen auf Basis von quasi-experimenteller Variation in den Treatment-Wahrscheinlichkeiten vor Ort durchführen zu können.

- Die im Jahr 2016 ins Feld gebrachte IAB-BAMF-SOEP-Befragung von Geflüchteten stellt eine Benchmark dar, wie eine Panelbefragung im Rahmen des Evaluationsvorhabens gestaltet und (etwa durch innovative Lösungen zur Überwindung von Kommunikationshürden) praktisch durchgeführt werden könnte. Die laufende IAB-BAMF-SOEP-Befragung von Geflüchteten selbst eignet sich allerdings nicht für die Durchführung des geplanten Evaluationsvorhabens. Es handelt sich um eine auf Repräsentativität hin angelegte Studie. Somit bildet die Zielgruppe der Personen im Kontext von Fluchtmigration, die für die Maßnahmen der aktiven Arbeitsförderung oder die Leistungen zur Eingliederung in Frage kommen, nur eine Teilmenge der Befragten, was die zur Verfügung stehenden Fallzahlen einschränkt. Außerdem erfasst die laufende Befragung die Teilnahme an den zu evaluierenden Leistungen im Zuständigkeitsbereich des BMAS nur unvollständig. Zwar besteht die Möglichkeit der Verknüpfung mit den registrierten Daten der Statistik der BA, durch die auf Repräsentativität zielende Anlage der Erhebung ist jedoch nicht gewährleistet, dass die Fallzahlen, die zur Lösung der Evaluationsaufgabe im Hinblick auf einzelne Maßnahmentypen verfügbar sind, ausreichend groß sind. Prinzipiell denkbar wäre, für die Zwecke der Evaluation eine spezifische Ergänzungsstichprobe der IAB-BAMF-SOEP-Befragung anzulegen. Aus Gründen der Wirtschaftlichkeit müsste diese von einer gematchten Auswahl von statistischen Zwillingen mit und ohne Maßnahmenteilnahme ausgehen; würde man der repräsentativen Anlage der schon laufenden Panelerhebung folgen, wären größere Fallzahlen erforderlich, da die Untersuchungseinheiten nur teilweise in die zu untersuchenden Leistungen einmünden und zudem auch Personen ohne Maßnahmenteilnahme befragt würden, die als Vergleichsfälle nicht passen. Eine Anlehnung an die IAB-BAMF-SOEP-Befragung erscheint auch im Hinblick auf das Fragebogenmaterial vorteilhaft. Dies schafft Möglichkeiten, Besonderheiten der im Rahmen des Evaluationsvorhabens untersuchten Zielgruppen zu erkennen. Wirtschaftlich betrachtet spart der Einsatz von bereits in der Praxis getesteten Items Entwicklungsaufwand; es wären lediglich Ergänzungen des Fragenprogramms vorzunehmen, um nähere Informationen zu den auf die Zielpersonen einwirkenden Maßnahmen bzw. Treatments zu erhalten. Dem Grunde nach besteht die Aufgabe also darin, die Stichprobenauswahl, das Fragenprogramm und die Erhebungsweise der für das Evaluationsvorhaben erforderlichen Längsschnitterhebungen ausgehend vom Vorbild der IAB-BAMF-SOEP-Befragung so zu adaptieren und zu fokussieren, dass eine angemessene Informationsqualität gewährleistet und zudem der Kostenaufwand in einem vertretbaren Rahmen bleibt. In Folge der sich daraus ergebenden methodischen Differenzen lassen sich die neu zu gewinnenden Längsschnittdaten und die Daten aus der laufenden IAB-BAMF-SOEP-Befragung von Flüchtlingen zwar nicht gemeinsam analysieren. Die Ähnlichkeiten in der Anlage der beiden Erhebungen bleiben aber groß genug, um die IAB-BAMF-SOEP-Befragung als Benchmark für Selektivitätsprüfungen und die externe Validierung der Daten der im Rahmen des Evaluationsvorhabens durchgeführten Längsschnitterhebung heranzuziehen.
- Im Idealfall legen quantitativ angelegte Wirkungsanalysen die durchschnittliche Veränderung der Zielgrößen bei den Teilnehmerinnen und Teilnehmern in Folge der Maßnahmenteilnahme offen. Um zu klären, wie die geschätzten Wirkungen zustande gekommen sind und handlungsorientierte Schlussfolgerungen zu ziehen, ist jedoch eine ergänzende qualitative Begleitforschung erforderlich. Diese hat die Aufgabe, organisatorische Prozesse und institutionelle Bedingungen herauszuarbeiten, die zur Wirksamkeit von Maßnahmen beitragen oder diese hemmen. Außerdem können qualitative Beobachtungen von Prozessen und Akteuren helfen, frühzeitig Randbedingungen zu erkennen, die für ein adäquates Design der quantitativen Wirkungsanalysen oder des Erhebungsprogramms berücksichtigt werden müssen. Deswegen sollten als Bestandteil des Vorhabens auch maßnahmenspezifische Fallstudien in den für die Panelbefragung von Personen im Kontext von Fluchtmigration ausgewählten regionalen Untersuchungseinheiten vorgesehen werden, die sowohl die

Ausgangslage vor Ort als auch die im Untersuchungszeitraum zu beobachtenden Veränderungen systematisch beschreiben.

Aus diesen Vorüberlegungen ergibt sich, dass die praktische Umsetzung des geplanten Evaluationsvorhabens im Hinblick auf das Datenkonzept und die geforderten Methoden anspruchsvoll ist und ganz erhebliche Ressourcen benötigt.

Zugleich muss das Vorhaben möglichst zeitnah begonnen werden. Erstens lässt sich sonst die Zielgruppe der in letzter Zeit von Deutschland im Kontext von Fluchtmigration aufgenommenen Personen nicht mehr im Anfangsstadium des Integrationsprozesses beobachten. Dies könnte die Evaluation in Folge von Effekten der Dauerabhängigkeit weiter verkomplizieren. Es ist davon auszugehen, dass sich die Integrationschancen systematisch mit zunehmender Dauer des Aufenthalts in Deutschland verändern, und die Größenordnung dieser Dauereffekte dürfte wiederum systematisch mit den persönlichen Merkmalen der Geflüchteten und den über die Zeit einwirkenden unterschiedlichsten Erfahrungen und empfangenen Hilfen auffächern. Zweitens gibt es ein Interesse der Akteure, die Integrationsmaßnahmen gestalten, möglichst rasch belastbare Evaluationsergebnisse zu erhalten.

Trotzdem darf der für die eigentliche Evaluation vorgesehene Zeitrahmen nicht zu knapp bemessen sein; dies ergibt sich nicht nur aus der Komplexität der Evaluationsaufgabe, sondern auch aus der Erwartung, dass die Integrationsprozesse in der Zielgruppe der aufgenommenen geflüchteten Menschen eher langsam verlaufen, so dass maßnahmenbedingte Fortschritte nur mit hinreichendem zeitlichem Abstand gemessen werden können. Eine zu kurzfristig angelegte Evaluation liefe also Gefahr, vorhandene positive Maßnahmeneffekte zu übersehen. Um dem weiter vorzubeugen, rücken zudem Frühindikatoren der Arbeitsmarktintegration als eigenständige Zielgrößen der Evaluation in den Blick.

Das nachfolgend dargestellte Konzept für eine systematische Evaluation der aktiven arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge trägt Wirtschaftlichkeitsüberlegungen Rechnung. Deswegen werden nicht alle in der Praxis eingesetzten Maßnahmen berücksichtigt. Insbesondere wird die heterogene Kategorie freier und sonstiger Fördermaßnahmen ausgeschieden. Zudem werden aktive Maßnahmen zur Verbesserung der Teilhabe behinderter geflüchteter Menschen nicht einbezogen. Zum einen ist hier mit relativ kleinen Fallzahlen zu rechnen, so dass die Basis für intern valide quantitative Analysen nicht ausreichen könnte. Zudem ergeben sich besondere Herausforderungen bei der Konstruktion einer für den Evaluationszweck geeigneten Kontrollgruppe. Sofern integrationspolitische Maßnahmen für behinderte Geflüchtete untersucht werden sollen, wird eine rein qualitativ orientierte Studienanlage empfohlen. Die Aufgabenstellung für eine solche komplementäre Studie wäre, Lebenslagen, Entwicklungsverläufe und Problemlagen für eine ausreichende Zahl von behinderten Zielpersonen im Kontext der Flüchtlingsmigration strukturiert zu erfassen und dabei Förderbedarfe und -wirkungen von ggf. in Anspruch genommenen Unterstützungsleistungen im Zuständigkeitsbereich des BMAS zu beschreiben.

Aus Gründen der Wirtschaftlichkeit ist das nachfolgende Konzept zur Umsetzung des Forschungsvorhabens auch darauf angelegt, Skaleneffekte zu realisieren. Dafür sollen unterschiedliche Leistungen möglichst nach denselben methodischen und Datenstandards evaluiert werden, womit Entwicklungsaufwand gespart werden kann. Dieser Ansatz führt insbesondere zu dem Vorschlag, die erforderliche Verlaufsbefragung als eigenständiges Vorhaben zu behandeln, also von der eigentlichen Wirkungsforschung abzutrennen. Dennoch birgt die eigenständige Behandlung der Datengenerierung gewisse Risiken, weil sie die Möglichkeiten einschränkt, die Informationsbasis auf spezifische Anforderungen der mit der Evaluation betrauten Forscherinnen und Forscher zuzuschneiden. Die Ergänzung des Sozio-oekonomischen Panels um das Modul „Familie in Deutschland“ im Rahmen der Gesamtevaluation ehe- und familienbezogener Leistungen zeigt jedoch, dass eine von der eigentlichen

Evaluation losgelöste Verbesserung der Datenbasis funktionieren kann, sofern bei der Gestaltung des Erhebungskonzepts von Anfang an die Belange der Evaluationsforschung im Zentrum stehen. Deswegen muss das BMAS, um die Datenqualität zu sichern, im Vergabeverfahren für die Durchführung der benötigten Längsschnitterhebungen entsprechende Vorgaben machen. Diese lassen sich aus den in dieser Machbarkeitsstudie angestellten Überlegungen ableiten.

Die weitere Darstellung eines Konzepts für die praktische Umsetzung eines umfassenden Forschungsvorhabens zur Evaluation der arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge gliedert sich wie folgt. Der folgende Abschnitt 7.2 entwickelt einen Vorschlag, wie das Gesamtvorhaben durch Aufteilung in verschiedene Module effektiv und wirtschaftlich bearbeitet werden könnte. Abschnitt 7.3 diskutiert den Zeitplan für die Umsetzung.

## 7.2 Vorschlag zur praktischen Durchführung

### 7.2.1 Modularisierung

Das geplante Evaluationsvorhaben ist sehr komplex und umfangreich. Einzelne Akteure, die für die Umsetzung fachlich grundsätzlich geeignet wären, aber alleine wahrscheinlich nicht in der Lage wären, die Gesamtheit der zu erfüllenden Aufgaben angemessen durchzuführen, müssten sich zur Durchführung einer Gesamtevaluation daher zu einem großen Konsortium zusammenschließen. Für eine solche Vorgehensweise spricht, dass damit Skaleneffekte in der Bearbeitung ausgeschöpft werden und konstruktionsbedingt eine Bearbeitung aus einer Hand erfolgen kann. Es kann damit aber auch eine Reihe von Problemen verbunden sein. Erstens könnte die Qualität der Analysen zu einzelnen Leistungen in der Tiefe leiden, wenn das zu beauftragende Konsortium nicht hinreichend groß und breit aufgestellt ist. Zweitens stellt die interne Governance der benötigten großen Struktur von wissenschaftlichen Akteuren sehr hohe Ansprüche. Diese sind innerhalb eines Konsortium aus unterschiedlichen fachlich kompetenten Akteuren bzw. Institutionen, die im Hinblick auf wissenschaftliche Reputation und auch finanzielle Ressourcen durchaus in Konkurrenz zueinander stehen, aber nicht leicht zu bewältigen.

Drittens dürfte eine große Konsortiallösung zu einem sehr eingeschränkten Bieterwettbewerb im Vergabeverfahren führen; die Anzahl der zur fachlich angemessenen Durchführung der Aufgaben grundsätzlich befähigten Wissenschaftsakteure ist nicht besonders groß, so dass die Herausbildung mehrerer hinreichend dimensionierter Anbieterkonsortien, die dann in Konkurrenz zueinander treten, eher unwahrscheinlich erscheint. Die Schaffung eines großen Konsortiums bedeutet zudem einen immensen Koordinierungsaufwand, an dem die Angebotserstellung scheitern kann. Schließlich könnte die Bereitschaft, sich an einem solchen Konsortium zu beteiligen, darunter leiden, dass der spezifische Beitrag einzelner Wissenschaftsakteure zum Ganzen nicht unbedingt klar erkennbar wird.

Eine alternative Vorgehensweise ist eine Modularisierung des Vorhabens, also die Zerlegung in eine Anzahl von Teilprojekten, die sich separat durch unterschiedliche Wissenschaftsakteure bearbeiten lassen. Dies begrenzt die mit einer großen Konsortiallösung verbundenen Risiken, erzeugt allerdings, soweit Skaleneffekte nicht realisiert werden, in der Durchführung möglicherweise auch einen Mehraufwand.<sup>12</sup> Darum ist darauf zu achten, dass die Module nicht zu klein dimensioniert werden.

Die Aufgabe der Koordination und Steuerung des Gesamtvorhabens kann bei einer Modullösung extern gelöst werden. Dies kann dem Ausgleich von Interessenkonflikten zwischen den beteiligten

---

<sup>12</sup> Soweit das Vergabeverfahren ergibt, dass gleiche Wissenschaftsakteure ihre Leistungen für verschiedene Module anbieten, könnte gegebenenfalls im Verhandlungsverfahren eine Kosten mindernde Zusammenlegung dieser Module herbeigeführt werden.

Wissenschaftsakteuren dienlich sein. Ein gewünschter Nebeneffekt der Modularisierung ist, dass damit der wissenschaftliche Entdeckungswettbewerb angeregt wird, so dass sich über die im Rahmen des Vergabeverfahrens festgelegten Vorgaben hinaus methodische Innovationen entwickeln können. Um bei einem modularen Konzept den Koordinationsaufwand zwischen den ausführenden Akteuren zu begrenzen, muss die Abgrenzung der einzelnen Evaluationsmodule von vornherein so erfolgen, dass es möglichst wenige inhaltliche Überschneidungen zwischen den einzelnen Modulen gibt.

Dazu bietet sich eine Modularisierung der kausalanalytischen Teile des Vorhabens nach dem Typus der zu bewertenden arbeitsmarktpolitischen Integrationsmaßnahmen an, weil zu deren Bearbeitung jeweils spezifische quasi-experimentelle Variation in den Daten ausgeschöpft werden muss und zudem möglichst spezifische begleitende qualitative Untersuchungen und Erhebungen vorgenommen werden müssen. Darüber hinaus wird vorgeschlagen, die Durchführung der Erhebungen zur Gewinnung der für die Wirkungsforschung als Ergänzung der administrativen Daten benötigten Verlaufsinformationen zu Personen im Kontext von Fluchtmigration in ein eigenständiges Modul zu verlagern. Der Vorschlag, ein solches Querschnittsmodul zu schaffen, trägt der Überlegung Rechnung, dass zu dessen Durchführung sehr spezifische Kompetenzen und Kapazitäten auf dem Feld der sozialwissenschaftlichen Umfrageforschung erforderlich sind. Zudem führt diese gebündelte Verfahrensweise zu größeren Fallzahlen. Damit verbundene Lerneffekte, sowie die sinkenden Setup-Kosten pro Fall, dürften gegenüber einer Struktur, bei der in die wirkungsanalytischen Module jeweils eigenständige Verlaufserhebungen integriert sind, die Gesamtkosten der Erhebung deutlich verringern. Dafür müssen eventuell gewisse Verschlechterungen bei der Datenqualität in Kauf genommen werden, die sich durch gute Vorgaben für die Erhebung und sorgfältige Steuerung mit Blick auf die Belange der Wirkungsforschung jedoch begrenzen lassen.

Die im Querschnittsmodul gebündelt gesammelten Daten sollten, sobald sie vorliegen und aufbereitet sind, stets allen mit den kausalanalytisch orientierten Modulen beauftragten Wissenschaftsakteuren zur Verfügung gestellt werden. Das Teilen von allen im Projektverlauf gewonnenen Daten führt zu besseren Möglichkeiten, Interaktionseffekte zwischen Maßnahmentypen sowie die Heterogenität von Maßnahmenwirkungen zu untersuchen.

### 7.2.2 Wirkungsanalytische Module

Eine Möglichkeit, die wirkungsanalytischen Teile des Vorhabens zu gliedern, ist, der in der BA-Statistik verwendeten Gruppierung der Maßnahmen der aktiven Arbeitsförderung und der Leistungen zur Eingliederung zu folgen.<sup>13</sup> Dies würde bedeuten, dass 1) die Leistungen zur Aktivierung und beruflichen Eingliederung, 2) die Leistungen zur Berufswahl und Berufsausbildung, 3) die Leistungen zur beruflichen Weiterbildung, sowie 4) die Leistungen zur Aufnahme einer Erwerbstätigkeit und Beschäftigung schaffende Maßnahmen separat und parallel zueinander auf ihre Wirkungen hin analysiert werden. Da die beiden letzten Maßnahmentypen auf die Aufnahme einer Erwerbstätigkeit gerichtet sind, erscheint es sinnvoll, ihre Untersuchung in einem Modul zusammen zu bearbeiten. Innerhalb dieses Moduls sollten die

---

<sup>13</sup> Eine andere Möglichkeit wäre, die Module nach der geforderten Methodik zu unterteilen. Eine solche Untergliederung scheint allerdings nicht ratsam. Erstens kommt es zu unnötigen Doppelarbeiten (etwa bei der Erarbeitung von Forschungsleitfragen oder bei der Datenaufbereitung), wenn mehrere Wissenschaftsakteure dieselbe Leistung aus mehreren statistischen Blickwinkeln beleuchten. Gravierender ist wohl noch das Argument, dass ex ante gar nicht festzulegen ist, welche der grundsätzlich denkbaren wirkungsanalytischen Ansätze, wie sie in Abschnitt 6 dargestellt wurden, überhaupt plausibel einsetzbar sind. Eine Entscheidung darüber erfordert nämlich bereits eine genaue Kenntnis der Daten. Schließlich stellt sich bei einer solchen Gliederung die Frage, wie der Abgleich und die Synthese der in unterschiedlichen Modulen zur gleichen Evaluationsfrage gewonnenen Erkenntnisse geleistet wird. Diese Aufgabe müsste auf eine übergeordnete Stelle oder in ein eigenständiges Teilprojekt verlagert werden. Wegen dieser Nachteile wird dieser Ansatz im Weiteren nicht näher verfolgt.

Wirkungen der beiden Typen von Interventionen aber getrennt voneinander analysiert werden, weil sie sich an völlig unterschiedliche Zielgruppen richten.

Eine noch feinere Zerlegung der Evaluationsaufgabe in Module zur Untersuchung für einzelne Maßnahmen- bzw. Leistungsarten wäre wegen nicht ausgeschöpfter Skaleneffekte vermutlich nicht wirtschaftlich. Dagegen wäre die Verringerung der Anzahl der wirkungsanalytischen Module durch gemeinsame Behandlung mehrerer Maßnahmen- und Leistungsgruppen denkbar, wobei innerhalb der Module, wie oben mit Blick auf das Modul zur Analyse der Leistungen zur Aufnahme einer Erwerbstätigkeit und der Beschäftigung schaffenden Maßnahmen festgestellt, den jeweils spezifischen Zielgruppen und zu erwartenden Wirkungen gerecht zu werden ist. So könnte man gegebenenfalls die Evaluation der Leistungen zur Berufswahl und Berufsausbildung und der Leistungen zur beruflichen Weiterbildung, die sachlich einigermaßen nah beieinander liegen, in einem Modul bündeln. Eine geringere Zahl an wirkungsanalytischen Modulen hätte den inhaltlichen Vorteil, dass Maßnahmenketten, welche die Teilnahmewahrscheinlichkeit und Wirkungen einzelner Maßnahmen beeinflussen können, nicht auseinander gerissen werden. Allerdings bleiben, weil die Daten für die Wirkungsforschung vollständig geteilt werden, die Untersuchungsmöglichkeiten, was Maßnahmenketten angeht, auch bei einer an einzelnen Maßnahmengruppen orientierten Modulstruktur erhalten.

Um zu gewährleisten, dass die wirkungsanalytischen Module nach einem einheitlichen Standard bearbeitet werden, sind die damit beauftragten Wissenschaftsakteure durch Vorgabe einer Reihe von Aufgabenfeldern zu steuern. Der jeweilige Untersuchungsauftrag sollte dazu insbesondere die folgenden Leitfragen formulieren:

- **Monitoring:** Wie entwickeln sich die Teilnahmezahlen, wie die Strukturen der Teilnehmerinnen und Teilnehmer? Wann nach Ankunft in Deutschland werden die Teilnehmerinnen und Teilnehmer erreicht? Gibt es typische Maßnahmenketten? Existieren regionale Unterschiede im Maßnahmeneinsatz?
- **Implementationsanalyse:** Wie werden die Maßnahmen und Instrumente für Personen im Kontext von Fluchtmigration in der Praxis umgesetzt? Wie gut wird bei der Umsetzung spezifischen Bedürfnissen der Zielgruppe Rechnung getragen? Wie laufen die Prozesse beim Zugang in die Maßnahme ab? Gibt es bei den für die Durchführung relevanten Arbeitsmarktakteuren unterschiedliche Handlungsstrategien? Wie gut funktionieren die Schnittstellen zwischen den an der Umsetzung beteiligten Akteuren?
- **Wirkungsanalyse:** Welche Effekte haben die zu untersuchenden Maßnahmen im Hinblick auf die für die Evaluation vorgegebenen Zielgrößen bei den Teilnehmerinnen und Teilnehmern, das heißt, welche Veränderungen dieser Zielgrößen lassen sich empirisch betrachtet ursächlich auf die Teilnahme zurückführen? Gibt es systematische Wirkungsunterschiede innerhalb der Zielgruppe, das heißt, unterscheiden sich die festgestellten Wirkungen qualitativ oder quantitativ in Abhängigkeit von beobachtbaren Merkmalen der Teilnehmerinnen und Teilnehmer? Gibt es systematische Einflüsse anderer Maßnahmen und Unterstützungsangebote, die zusätzlich auf die Teilnehmerinnen und Teilnehmer einwirken, auf die Wirksamkeit der zu evaluierenden Maßnahme? Lassen sich Unterschiede in der Wirksamkeit in Abhängigkeit von anderen Kontextbedingungen (insbesondere Elemente der Prozessgestaltung und Strategien bei den durchführenden Akteuren, Bedingungen am lokalen Arbeits- und Ausbildungsmarkt, Verfügbarkeit sonstiger Unterstützungsangebote zum Beispiel von Ehrenamtlichen) ausmachen?
- **Effizienzanalyse:** Wie wirtschaftlich sind die eingesetzten Instrumente und Maßnahmen, das heißt in welchem Verhältnis stehen die damit insgesamt erzielten Wirkungen zu den dafür eingesetzten

Ressourcen? Gibt es Nebenwirkungen außerhalb der Zielgruppe, insbesondere bei anderen Akteuren mit einer besonders schwachen Position im Arbeitsmarkt (Geringqualifizierte, Langzeitarbeitslose, andere Personen mit Migrationshintergrund)? Gibt es Verdrängungseffekte am Arbeitsmarkt oder bei der Verteilung von Ressourcen durch die durchführenden Akteure?

- **Gender Mainstreaming:** Inwieweit tragen die arbeitsmarktpolitischen Integrationsmaßnahmen und -instrumente den unterschiedlichen Lebenslagen und Bedürfnissen weiblicher und männlicher Flüchtlinge angemessen Rechnung? Tragen sie zum Abbau der besonders hohen Hürden beim Arbeitsmarktzugang für weibliche Flüchtlinge bei, die sich aus der Stellung der Frauen in vielen der typischen Herkunftsländer ergeben? Gibt es Hinweise auf strukturelle Diskriminierung weiblicher Flüchtlinge beim Zugang zu Maßnahmen? Wie lassen sich gegebenenfalls gemessene systematische Unterschiede der Wirksamkeit nach Geschlecht erklären?
- **Handlungsorientierte Schlussfolgerungen:** Welche konkreten Handlungsempfehlungen ergeben sich aus den Befunden? Wie ließen sich die Wirksamkeit und die Wirtschaftlichkeit der untersuchten Maßnahmen verbessern? Gibt es beim Konzept der Interventionen, bei den Prozessen ihrer Umsetzung oder im Hinblick auf die erreichte Zielgruppe Verbesserungspotenziale?

Die mit der Durchführung beauftragten Wissenschaftlerinnen und Wissenschaftler sollten darüber hinaus einen gewissen Spielraum haben, eigenständige Fragestellungen zu formulieren und zu verfolgen – vorausgesetzt, der davon zu erwartende Erkenntnisgewinn für den Auftraggeber steht in einer angemessenen Relation zum damit verbundenen Aufwand.

Um die Aufgabenstellungen der auf die Maßnahmen der aktiven Arbeitsförderung und der Leistungen zur Eingliederung der BA ausgerichteten Evaluationsmodule zu erfüllen, sind jeweils insbesondere die folgenden Arbeitsschritte durchzuführen:

- Legung eines Zugangs zu den administrativen Daten der BA, um die Teilnahme oder Nichtteilnahme von Personen im Kontext von Fluchtmigration zu erfassen;
- Beschreibung der Entwicklung von Teilnehmerstrukturen und der zugehörigen Zielgrößenentwicklungen anhand der administrativen Daten;
- Ziehung einer Stichprobe aus den administrativen Daten, die Teilnehmerinnen und Teilnehmer der Maßnahme sowie eine mehrfach übersetzte geeignete Kontrollgruppe von Nichtteilnehmerinnen und Nichtteilnehmern aus den Fokusregionen der Analyse umfasst; Übermittlung der Adressinformationen zu dieser Stichprobe an den Auftragnehmer des Datenmoduls;
- Durchführung umfassender qualitativer Analysen zur Dokumentation und Bewertung der zugangs- und wirkungsrelevanten organisatorischen und institutionellen Gegebenheiten auf regionaler Ebene,<sup>14</sup> auf Grundlage von Dokumentenanalysen, leitfadengestützten Interviews und sonstigen gängigen Verfahren qualitativer Feldforschung und Fallstudien;

---

<sup>14</sup> Die qualitative Begleitforschung sollte auf der Ebene der Agenturbezirke stattfinden, in denen auch die für das Gesamtvorhaben geplante Panelerhebung bei Personen im Kontext von Fluchtmigration durchgeführt wird. Somit wäre eine konsistente Informationsbasis gewährleistet. Um den Aufwand zu begrenzen, ist dabei eine im Hinblick auf den spezifischen Untersuchungsgegenstand des Evaluationsmoduls begründete Auswahl von Agenturbezirken vorzunehmen. Bei der Auswahl der in den Blick genommenen Regionen sollten Erkenntnisse zu Teilnehmerstrukturen und Treatment-Intensität aus den administrativen Daten der BA berücksichtigt werden; zudem ist auf eine hinreichende Streuung der lokalen Kontexte und Hintergründe zu achten.

- Implementation und Durchführung kausalanalytisch orientierter Wirkungsanalysen unter Verwendung von Befragungsdaten und administrativen Daten; Analyse der systematischen Bedeutung lokaler Kontextfaktoren für die Wirksamkeit der fokussierten Maßnahmen und Leistungen; Diagnose systematischer Veränderungen der Wirksamkeit der fokussierten Maßnahmen und Leistungen in Abhängigkeit von der Kombination mit anderen Interventionen („Evaluation von Maßnahmenketten“);
- Interpretation der quantitativen Befunde unter Einbeziehung der Erkenntnisse aus der qualitativ orientierten Begleitforschung sowie Ableitung von möglichst konkreten Handlungsempfehlungen zur Weiterentwicklung der untersuchten arbeitsmarktpolitischen Integrationsmaßnahmen für Flüchtlinge.

Wegen der vorgesehenen Verzahnung von qualitativ und quantitativ orientierten Analysen wird für eine angemessene Durchführung der Evaluationsmodule normalerweise eine multi-disziplinäre Herangehensweise erforderlich sein. Wünschenswert ist insbesondere, dass volkswirtschaftliche bzw. arbeitsmarktökonomische und sozialwissenschaftliche Kompetenzen miteinander kombiniert werden.<sup>15</sup>

Potenzielle Anbieter für die wirkungsanalytischen Teile des Vorhabens müssen über hinreichend Erfahrung in der Durchführung quantitativ und qualitativ orientierter Studien zur Wirksamkeit aktiver arbeitsmarktpolitischer Maßnahmen verfügen. Unverzichtbar sind Erfahrungen in der Analyse von administrativen Sozialdaten der Bundesagentur für Arbeit. Kenntnisse in der Erforschung von Integrationsprozessen von Personen mit Migrationshintergrund sind darüber hinaus wünschenswert. Als Auftragnehmer in Frage kommen insbesondere wirtschafts- und sozialwissenschaftliche Forschungsinstitute, aber auch einzelne Forscher oder Forschergruppen mit einschlägiger Expertise. Ob sich potenzielle Anbieter zur vollumfänglichen Erfüllung der Anforderungen zu Konsortien zusammenschließen oder sich Nachunternehmern bedienen, bleibt den in Frage kommenden Akteuren selbst überlassen.

Es wird angeregt, parallel zu den vier den Maßnahmen der aktiven Arbeitsförderung und der Leistungen zur Eingliederung gewidmeten Modulen ein fünftes wirkungsanalytisches Evaluationsmodul durchzuführen, um die Wirkungen von Maßnahmen der Sprachförderung zu fokussieren. Die für diese Analyse erforderlichen Teilnahmeinformationen zu verschiedenen Typen der Sprachförderung sind im Rahmen der geplanten Panelerhebung bei den Zielpersonen im Kontext von Fluchtmigration zu erfassen, da sie in den administrativen Daten nicht vollständig und konsistent abgebildet sind. Entsprechend entfallen in diesem Modul die vorbereitenden Arbeitsschritte der Legung eines Zugangs zu den administrativen Daten der BA, um die Teilnahme oder Nichtteilnahme von Personen im Kontext von Fluchtmigration zu erfassen, sowie der Ziehung einer spezifischen Stichprobe, die in die vorgesehene Längsschnittbefragung von Personen im Kontext von Fluchtmigration einfließt.

---

<sup>15</sup> Eine denkbare Alternative wäre, die qualitative Evaluation und die quantitativ orientierte Wirkungsforschung in voneinander getrennten Modulen zu organisieren. Es kann Synergieeffekte schaffen, wenn die Fokussierung auf einzelne Leistungs- und Maßnahmentypen in den qualitativen Untersuchungen aufgegeben wird. Die in der Praxis sehr bedeutsame Koppelung von Maßnahmen und Leistungen unterschiedlichen Typs kann gemeinsam behandelt werden, was eine Verringerung der erforderlichen Zahl an Fallstudien mit sich bringen dürfte. Die Trennung der beiden unterschiedlichen Herangehensweisen an eine Evaluation birgt aber auch das Risiko, dass die Qualität der quantitativ orientierten Wirkungsforschung leidet. Die Gewinnung von Arbeitshypothesen, Auswahl und Design geeigneter statistischer Evaluationsmodelle und die Interpretation der aus der statistischen „Black Box“ kommenden Ergebnisse erfordern eine intensive Befassung mit den Erkenntnissen zur Implementation der Leistungen vor Ort und darum einen engen Austausch mit den die qualitative Begleitforschung durchführenden Wissenschaftsakteuren. Bei einer Trennung beider Sphären wächst damit der Aufwand für die Sicherung der zwischen ihnen notwendigen Austauschprozesse. Dies lässt sich durch vertragliche Vorgaben allein nur bedingt regeln; vielmehr muss die Einhaltung der Kooperationsvorgaben auch gut überwacht werden. Nach Abwägung der Vor- und Nachteile scheint eine Integration von quantitativen und qualitativen Evaluationsansätzen in einem Untersuchungsmodul vorteilhafter. Diese Lösung schließt nicht aus, dass ein dafür geeigneter Anbieter die qualitative Begleitforschung in mehreren Evaluationsmodulen übernimmt und damit die oben skizzierten Skalenvorteile realisiert.

Da fast alle Flüchtlinge im Verlauf die eine oder andere Form der Sprachförderung erhalten, lassen sich deren Wirkungen vermutlich nicht isolieren. Vielmehr sollte das Evaluationsmodul zur Sprachförderung systematisch der Interaktion der Effekte der Sprachfördermaßnahmen in Kombination mit anderen arbeitsmarktpolitischen Integrationsmaßnahmen nachgehen. Die Wirkungsanalysen auf dieser Basis können mittels linearen Regressionen und Matching-Techniken erfolgen. Darüber hinaus soll der Auftragnehmer prüfen, ob sich in den Daten kausalanalytisch verwertbare quasi-experimentelle Variation findet, mit der Selektivitäten beim Zugang zu Maßnahmen der Sprachförderung kontrolliert werden können. Analog zu den übrigen kausalanalytischen Modulen soll auch hier eine die quantitativen Analysen unterstützende und ergänzende qualitative Begleitforschung durchgeführt werden, welche die Bedingungen und Ursachen für den Erfolg (und Misserfolg) der Sprachförderleistungen durch eine strukturierte Erfassung von Maßnahmenumsetzung und Rahmenbedingungen in dafür gezielt ausgewählten lokalen Kontexten systematisch herausarbeitet.

Die vorgeschlagene Behandlung der Maßnahmen zur Sprachförderung zeigt im Übrigen wie eine Blaupause, nach welchem Standard prinzipiell auch weitere Maßnahmen außerhalb der Zuständigkeit des BMAS im hier entwickelten Evaluationsrahmen untersucht und bewertet werden können. Die Voraussetzung dafür ist, dass die Teilnahme an solchen Maßnahmen, die in den administrativen Daten der BA normalerweise nicht zu beobachten ist, in der ergänzenden Befragung von Personen im Kontext von Fluchtmigration erfasst wird. Eine Erfassung von Leistungen und Hilfen, die auf die Zielgruppe einwirken, ist im Erhebungskonzept grundsätzlich vorgesehen. Allerdings kann diese nur summarisch erfolgen; es geht um den Typus (Art und Intensität) der Interventionen, nicht um die Teilnahme an ganz spezifischen Maßnahmen oder Programmen. Diese Beschränkung hat nicht nur Kostengründe; es ist damit zu rechnen, dass die Teilnehmerinnen und Teilnehmer vielfach nicht genau wissen, wie eine sie erreichende Unterstützungsleistung genau zu bezeichnen ist.

Eine weitere gewichtige Beschränkung, Wirkungsaussagen zu anderen einzelnen potenziell integrationsfördernden Hilfen für Flüchtlinge abzuleiten, ergibt sich aus den für die Befragung geplanten Fallzahlen. Diese sind aus wirtschaftlichen Gründen so dimensioniert, dass sich nur für Interventionen, die häufiger auftreten, Fallzahlen realisieren lassen, die für belastbare Analysen ausreichen. Um im abgesteckten Rahmen Leistungen mit einer geringen Verbreitung in der Grundgesamtheit zu evaluieren, müsste jeweils eine spezifische Ergänzung der Stichprobe vorgenommen werden. Dazu müssen geeignete Informationen vorliegen, um ein Oversampling gezielt betreiben zu können. Ginge es beispielsweise um eine Maßnahme, die nur in bestimmten Agenturbezirken zum Einsatz kommt, müsste eine eventuelle Ergänzungsstichprobe Untersuchungseinheiten aus diesen Bezirken gezielt auswählen. Solche Erweiterungen sind im Rahmen dieses Evaluationsvorhabens nicht vorgesehen; das hier entwickelte Konzept könnte aber als Benchmark für andere geplante Evaluationsvorhaben im Bereich der integrationsfördernden Maßnahmen für Flüchtlinge und Schutzberechtigte herangezogen werden.

### 7.2.3 Befragungsmodul

Die verbindende Klammer zwischen den wirkungsanalytischen Modulen stellt die Panelbefragung von Teilnehmerinnen und Teilnehmern sowie Nichtteilnehmerinnen und Nichtteilnehmern dar. Die Durchführung dieser aufwändigen Erhebung ist Gegenstand eines eigenständigen Befragungsmoduls, das nur von äußerst erfahrenen und leistungsfähigen Befragungsinstituten erfolgreich bewältigt werden kann. Der Auftragnehmer muss insbesondere mit mehrsprachigen Befragungen vertraut sein und sollte idealerweise bereits Erhebungen bei der speziellen Zielgruppe der geflüchteten Menschen durchgeführt haben.



Die Aufgabenstellung dieses Befragungsmoduls ist es, mit einem einheitlichen Instrument, einem einheitlichen Feldzugang und zu einheitlichen Erhebungszeitpunkten bei den – schwer anzusprechenden und mutmaßlich hochmobilen – Zielpersonen im Kontext der Fluchtmigration erhobene Daten zu liefern, die als Kontrollvariablen oder Zielgrößen in den wirkungsanalytischen Modulen gebraucht werden. Die Arbeiten im Befragungsmodul erfüllen innerhalb des Gesamtvorhabens eine rein unterstützende Dienstleistungsfunktion. Es ist daher bis auf eine umfassende Dokumentation des Feldverlaufs und der Feldergebnisse kein inhaltlich eigenständiger Endbericht gefordert.

Die durch die Auslagerung der Datenerhebung in ein spezielles Modul zu erzielenden finanziellen Einsparungen sind angesichts des mit der Zielgruppe verbundenen besonderen Erhebungsaufwands notwendig, damit das Vorhaben überhaupt wirtschaftlich durchführbar wird. Dies rechtfertigt auch den Nachteil der vorgeschlagenen Vorgehensweise, dass besondere Datenanforderungen der einzelnen Evaluationsmodule nur sehr beschränkt aufgenommen werden können. In der modulübergreifenden Erhebung lässt sich lediglich ein Platzhalter für eine strikt begrenzte Zahl von modulspezifischen Items vorsehen.

Bei der Zeitplanung des Gesamtvorhabens ist darauf zu achten, Risiken durch eine unzureichende Verzahnung zwischen den Evaluationsmodulen und dem Erhebungsmodul zu vermeiden. Die wirkungsanalytischen Module sind im zeitlichen Vorlauf vor dem Befragungsmodul zu vergeben, so dass besondere Datenanforderungen, die in den diesbezüglichen Vergabeverfahren erkennbar werden, gegebenenfalls noch in die Leistungsbeschreibung für die Vergabe des Befragungsmoduls aufgenommen werden können. Eine entsprechende Zeitplanung lässt sich realisieren, da die Auftragnehmer der kausalanalytischen Module ohnehin einen mehrmonatigen Vorlauf für die Bildung einer für ihre Untersuchung geeigneten Bruttostichprobe benötigen. Nach deren Fertigstellung erhält der Auftragnehmer des Befragungsmoduls das zugehörige Adressmaterial und kann erst dann mit der Feldarbeit beginnen.

Die Finalisierung der im Rahmen der Panelerhebung einzusetzenden Fragebögen sollte in Abstimmung mit dem für die Durchführung der Befragung beauftragten Befragungsunternehmen erfolgen. Die Fragebögen sind durch das eingesetzte Befragungsunternehmen zur Durchführung der Erhebung zu programmieren. Des Weiteren sind im Rahmen des Erhebungsmoduls folgende Erhebungsleistungen zu erbringen:

- **1. Welle:** Die Erhebung ist mehrsprachig als computerunterstütztes persönliches Interview (CAPI) mit einer durchschnittlichen Befragungszeit von 60 Minuten anzulegen. Die umfassende Ersterhebung dient dazu, eine näherungsweise Nullmessung der Zielgrößen zu bekommen, die in den administrativen Daten nicht erfasst sind. Zudem geht es darum, eine größere Zahl von individuellen Hintergrundfaktoren zu erfassen, die sowohl die Zugangswahrscheinlichkeit in die Maßnahme als auch die Zielgrößen der Evaluation systematisch beeinflussen könnten. Die Ersterhebung sollte, sich am Vorbild der IAB-BAMF-SOEP-Befragung von Flüchtlingen orientierend, face-to-face durchgeführt werden. Für jeden der fünf zu evaluierenden Maßnahmentypen (zwei Maßnahmentypen sind im wirkungsanalytischen Modul 4) zusammengefasst) sind je 900 Interviews mit Teilnehmerinnen und Teilnehmern und je 1.800 Interviews mit Nichtteilnehmerinnen und Nichtteilnehmern zu realisieren. Die Interviews sind in einer strukturiert geplanten Auswahl von 70 bis 80 Agenturbezirken, die das Spektrum der wirtschaftlichen und sonstigen Kontextfaktoren gut abdeckt, zu realisieren.
- **2. Welle:** Diese Welle verfolgt hauptsächlich den Zweck das Panel zu stabilisieren. Es geht in erster Linie darum, den Kontakt zu den Panelteilnehmerinnen und Panelteilnehmern zu erhalten, die telefonischen oder digitalen Kommunikationswege zu prüfen und ggf. Kontaktdaten zu aktualisieren. Für den eigentlichen Zweck der Evaluation benötigte Zielgrößen sollen nur sehr knapp abgefragt

werden. Dennoch halten wir den mit einer Zwischenwelle verbundenen Aufwand für wirtschaftlich: Da sich damit die Rate der Panelausfälle verringert, kann die Erhebung mit einer kleineren Eingangsstichprobe auskommen. Die Erhebung ist mehrsprachig als computerunterstütztes telefonisches Interview (CATI) mit einer durchschnittlichen Befragungszeit von 5 Minuten anzulegen. Für jeden der fünf zu evaluierenden Maßnahmentypen sollen insgesamt 630 Interviews mit Teilnehmerinnen und Teilnehmern und 900 Interviews mit Nichtteilernehmerinnen und Nichtteilnehmern der ersten Welle realisiert werden. Dieser Ansatz reflektiert eine unterstellte Ausfallrate von 30 Prozent bei Teilnehmerinnen und Teilnehmern und von 50 Prozent bei Nichtteilernehmerinnen und Nichtteilnehmern.

- **3. Welle:** Die abschließende Erhebung soll eine erneute umfassende Messung der Zielgrößen der Evaluation vornehmen. Sie ist mehrsprachig als computerunterstütztes telefonisches Interview (CATI) mit einer durchschnittlichen Befragungszeit von 20 Minuten anzulegen. Für jeden der fünf zu evaluierenden Maßnahmentypen sollen 450 Interviews mit Teilnehmerinnen und Teilnehmern und 450 Interviews mit Nichtteilernehmerinnen und Nichtteilnehmern der ersten Welle realisiert werden. Dieser Ansatz unterstellt noch einmal ähnlich hohe Ausfallraten wie von der ersten zur zweiten Welle.

Die Planung für dieses Befragungsgerüst berücksichtigt die im Rahmen der Durchführung der IAB-BAMF-SOEP-Befragung von Flüchtlingen gemachten Erfahrungen. Danach ist trotz des damit verbundenen hohen Aufwands eine persönliche Ansprache der Zielgruppen der Untersuchung vorzusehen. Nur auf diese Weise können – soweit erforderlich auch unter Einsatz innovativer Instrumente wie computergestützter Übersetzungshilfen – angesichts der besonderen Zielgruppe zu erwartende gravierende Verständigungshürden angemessen überwunden werden. Zudem ist das für die Eingangserhebung vorzusehende Befragungsprogramm so umfangreich, dass bei einer telefonischen Abwicklung mit hohen Abbruchquoten zu rechnen wäre.

Ein intensiver persönlicher Erstkontakt scheint auch erforderlich, um die Chancen auf eine Beteiligung an der zweiten und dritten Panelwelle hinreichend zu erhöhen. Trotzdem ist einzuplanen, dass für ein gutes Tracking Adressrecherchen bei den Einwohnermeldeämtern durchgeführt werden müssen.

Die Durchführung der zweiten und dritten Welle per CATI statt als CAPI ist durchaus mit Risiken verbunden. Erstens dürfte dies zu einer höheren Ausfallrate im Panel führen, weil Telefonnummern veralten. Dem lässt sich in gewissem Rahmen vorbeugen, indem auch digitale Kontaktinformationen (User-Name in sozialen Netzwerken, E-Mail) abgelegt werden. Zweitens lassen sich bei telefonischer Ansprache Verständnishaürden schwerer überwinden als im persönlichen Gespräch. Die zweite und dritte Welle des Panels ebenfalls per CAPI durchzuführen erscheint jedoch aus wirtschaftlichen Gründen nicht vertretbar. Um die Teilnahmebereitschaft in den späteren Wellen zu erhöhen, sollten deshalb auch Mittel für eine monetäre Incentivierung eingeplant werden.

Die Übersetzung von Teilnehmerinnen und Teilnehmern zu Nichtteilernehmerinnen und Nichtteilnehmern im Verhältnis 1:2 trägt der Erwartung Rechnung, dass sich anhand der zur Stichprobenbildung verwendeten administrativen Daten eine zur Ableitung kausalanalytischer Wirkungsaussagen geeignete Vergleichsgruppe von Nichtteilernehmerinnen und Nichtteilnehmern nur mit großer Unsicherheit bilden lässt. Die Übersetzung verbessert die Möglichkeiten, „statistische Zwillinge“ anhand der in der ersten Welle erhobenen individuellen Hintergrundvariablen zu bilden. Eine höhere Anzahl von Nichtteilernehmerinnen und Nichtteilnehmern ist auch erforderlich, weil bei dieser Zielgruppe erfahrungsgemäß mit höheren Ausfallraten zu rechnen ist. Außerdem könnte ein Teil davon noch zu einem späteren Zeitpunkt in die zu analysierende Maßnahme einmünden und wäre dann als Mitglied der Kontrollgruppe nur noch bedingt geeignet.

Technisch gesehen ist es zwar kein fundamentales Problem, falls Nichtteilnehmerinnen oder Nichtteilnehmer zu einem späteren Zeitpunkt in Maßnahmen einmünden. Diese Tatsache ist aber bei der Interpretation der geschätzten Behandlungseffekte zu beachten. Die diagnostizierten Effekte geben immer an, wie eine Maßnahmenteilnahme im Verhältnis zu allen anderen Maßnahmenverläufen wirkt, die zu einem späteren Zeitpunkt der Entwicklung eintreten können. Würden etwa alle Personen, die zum Startzeitpunkt nicht teilnehmen, später in die zu untersuchende Maßnahme einmünden, würde der geschätzte Effekt die Wirkungen einer frühen Teilnahme im Verhältnis zu einer späteren Teilnahme messen; somit wäre der eigentlich gesuchte Effekt der Maßnahme an sich nicht zu identifizieren. Bei späterem Einmünden in andere Maßnahmen misst der geschätzte Effekt die Wirkung der Maßnahme im Verhältnis zur kontrafaktischen Situation, dass die Teilnehmerinnen und Teilnehmer ansonsten mit bestimmten Wahrscheinlichkeiten in andere Maßnahmen eingemündet wären.

Ob ein reiner Teilnahmeeffekt bezogen auf die Maßnahme im Fokus ermittelt werden kann, hängt davon ab, wie viele Fälle in der „reinen Kontrollgruppe“ verbleiben, die nicht in weitere Maßnahmen einläuft. Neben der Frage der Fallzahlen ist aber auch bedeutsam, inwiefern die Gruppe derjenigen, die nie in eine Leistung einmünden, in der Grundgesamtheit der Personen im Kontext von Fluchtmigration eine reine Zufallsauswahl ist. Soweit es sich um eine positiv oder negativ selektierte Auswahl handelt, wird durch Bezugnahme auf eine reine Kontrollgruppe der durchschnittliche Effekt der Maßnahme an sich unterschätzt oder überschätzt.

Die für die dritte Welle eingeplanten Fallzahlen für jeden der zu untersuchenden Maßnahmentypen erscheinen gerade noch ausreichend, um gewisse Unterschiede bei den Wirkungen von Maßnahmen zwischen verschiedenen sozialen Gruppen analysieren zu können. Zusätzlich könnte bei der Ziehung der Bruttostichprobe eine Schichtung nach wesentlichen beobachtbaren Merkmalen (Herkunftsland, Alter, Geschlecht) vorgenommen werden, um abzusichern, dass für belastbare differenzierte Analysen am Ende noch hinreichend Fälle verfügbar sind. Da die Auswahlgrundgesamtheit bekannt ist, lassen sich Gewichtungsfaktoren bestimmen, um die anhand der geschichteten Stichprobe ermittelten Ergebnisse hochzurechnen.

#### 7.2.4 Koordinierung durch eine Geschäftsstelle

Trotz Zerlegung des Forschungsvorhabens in unabhängig voneinander zu bearbeitende Einzelmodule muss die Konsistenz der Evaluationsergebnisse gewährleistet bleiben. Unterschiede in den geschätzten Wirkungen sollen im Idealfall nur auf Unterschiede in der Wirksamkeit der untersuchten Maßnahmen zurückzuführen sein, und nicht etwa auf Unterschiede in den eingesetzten Methoden oder Daten. Andernfalls ließen sich aus einer vergleichenden Zusammenschau der Befunde nur schwer handlungsorientierte Schlussfolgerungen zur Weiterentwicklung des Tableaus der arbeitsmarktpolitischen Integrationsleistungen für Personen im Kontext von Fluchtmigration ableiten. Präzise Leistungsbeschreibungen allein dürften zur Sicherung der Konsistenz eines auf verschiedene Auftragnehmer verteilten komplexen Forschungsvorhabens nicht ausreichen. Vielmehr empfiehlt sich, den Beispielen der umfassenden Evaluation der aktiven arbeitsmarktpolitischen Leistungen im Zuge der Hartz-Reformen und der Gesamtevaluation der ehe- und familienbezogenen Leistungen in Deutschland folgend, zur Erfüllung der übergeordneten Koordinationsaufgaben die Gründung einer Geschäftsstelle. Diese kann zugleich die Mittlerrolle zwischen den Akteuren, die die Evaluationsmodule bearbeiten, und den Adressaten der Evaluationsergebnisse auf der politischen und administrativen Ebene einnehmen.

Die vorgeschlagene Geschäftsstelle hätte ein Bündel sowohl inhaltlicher als auch organisatorischer Aufgaben zu erledigen. Auf der inhaltlichen Ebene hätte die Geschäftsstelle insbesondere die folgenden Dinge zu erledigen:

- **Bereitstellung der Dateninfrastruktur:** Es gilt, die als Querschnittsdienstleistung, das heißt unabhängig von den wirkungsanalytischen Modulen organisierte Panelerhebung bei den Zielpersonen im Kontext der Fluchtmigration zu steuern; hierfür muss die Geschäftsstelle insbesondere zu Projektbeginn das Erhebungsinstrument in Abstimmung mit dem BMAS, den mit der Durchführung der wirkungsanalytischen Module betrauten wissenschaftlichen Akteuren sowie den Verantwortlichen für die als Benchmark dienende IAB-BAMF-SOEP-Befragung von Flüchtlingen finalisieren; im weiteren Verlauf sind ggf. projektbegleitend Details von Erhebungsverfahren und -inhalten (neu) festzulegen;
- **Aufbereitung von Ergebnissen:** Die Geschäftsstelle hat die Aufgabe, die in den verschiedenen Modulen des Evaluationsvorhabens erzielten Erkenntnisfortschritte in regelmäßigen, adressatengerecht formulierten Berichten zusammenzufassen und daraus handlungsorientierte Schlussfolgerungen abzuleiten. Im Projektverlauf umfasst diese Aufgabe die Entwicklung von Verfahrensvorschlägen bei einer gegebenenfalls erforderlichen Nachsteuerung des Evaluationsvorhabens; am Projektende obliegt die Erstellung des zusammenfassenden Abschlussberichts zum Gesamtvorhaben der Geschäftsstelle. Der Abschlussbericht soll die erzielten Erkenntnisfortschritte in der Zusammenschau bewerten und anhand eines Gesamtbilds möglichst konkrete Handlungsvorschläge zur Weiterentwicklung des Gesamtableaus arbeitsmarktpolitischer Integrationsmaßnahmen für die Flüchtlinge unterbreiten.

Auf der organisatorischen Ebene hat die Geschäftsstelle insbesondere die folgenden Aufgaben zu erfüllen:

- **Regelmäßiger Informationsaustausch mit dem zuständigen Referat des BMAS:** Einerseits sollen eventuelle Veränderungen der Erwartungen an die Evaluation beim Auftraggeber frühzeitig aufgenommen und an die Auftragnehmer weitergegeben werden; andererseits soll das Ressort über den aktuellen Stand der Umsetzung der Evaluation in Kenntnis gesetzt werden;
- **Regelmäßiger Informationsaustausch mit den beauftragten wissenschaftlichen Akteuren:** Damit soll die Einhaltung der grundlegenden Leitlinien der Evaluation sichergestellt und bei den parallel wie sequentiell bearbeiteten Evaluationsmodulen Konsistenz und Vergleichbarkeit der Ergebnisse so weit wie möglich gewahrt bleiben;
- **Förderung des Informationsaustauschs zwischen den beauftragten wissenschaftlichen Akteuren:** Dies dient dem Ziel, ein wechselseitigen Lernen im Projektverlauf zu unterstützen;
- **Interne und externe Qualitätssicherung zur Gewährleistung der fachlichen Güte der Analyse:** Die Geschäftsstelle soll die Qualität der im Rahmen der Durchführung der Module erstellten Berichte inhaltlich und formal sorgfältig prüfen und darüber hinaus für eine unabhängige Qualitätskontrolle von außen sorgen – Methoden und Resultate des Forschungsvorhabens sollten in wissenschaftlichen Foren, die von der Geschäftsstelle evaluationsbegleitend organisiert werden, präsentiert und zur Diskussion gestellt werden;
- **Koordination und sachliche Unterstützung der Akteure beim Zugang zu den administrativen Sozialdaten:** Dies dient dem Ziel, unnötige Doppelarbeit und Kosten bei der Interaktion mit der Daten bereitstellenden Stelle zu vermeiden;
- **Zusammenführung der für die Evaluationsmodule 1-4 erstellten Datensätze und Bereitstellung des Gesamtdatensatzes für das übergreifende Evaluationsmodul 5.**

Das skizzierte Aufgabenspektrum an eine Geschäftsstelle zu übertragen, würde das BMAS von kleinteiligen und teilweise ein hohes Maß an wissenschaftlichem Spezialwissen voraussetzenden Vorgängen entlasten. Dagegen verblieben Planung und Vorgabe der Leitlinien des Evaluationsvorhabens vollständig im Zuständigkeitsbereich des BMAS. Dies schließt auch die Durchführung des Vergabeverfahrens und die Entscheidung über die Auftragsvergabe für das Forschungsvorhaben, oder genauer der verschiedenen zu seiner Durchführung vorgesehenen Module, ein.

Eine Herausforderung bei der Umsetzung dieses Arbeitsmodells könnte darin bestehen, einen für die Besetzung der Geschäftsstelle geeigneten Akteur zu finden. Dieser muss einerseits mit der Koordination und Durchführung umfangreicher Projekte der wissenschaftlichen Beratung auf dem Feld der Arbeitsmarktpolitik sehr gut vertraut sein. Andererseits ist hinreichende wissenschaftliche Kompetenz notwendig, um das Gesamtvorhaben fachlich adäquat zu steuern und dabei von den mit der Durchführung der Module zu beauftragenden Akteuren als inhaltlicher Rat- und Impulsgeber voll respektiert zu werden.

Die Geschäftsstelle muss über die gesamte Laufzeit des Evaluationsvorhabens hinweg bestehen. Damit sie gleich zum Start des Evaluationsvorhabens ihre Arbeit aufnehmen kann, muss das diesbezügliche Vergabeverfahren zeitgleich zu den Vergabeverfahren zur Durchführung der Evaluationsmodule 1-4 ablaufen.

### 7.3 Zeitplan

Der nachfolgend beschriebene Zeitplan zur Durchführung des Vorhabens geht davon aus, dass das geplante Forschungsvorhaben noch in der zweiten Jahreshälfte 2017 auf den Weg gebracht werden kann – unterstellt wird ein Starttermin am 1. Oktober 2017. Weiterhin wird eine Gesamtlaufzeit des Evaluationsvorhabens von 36 Monaten angesetzt, die sich an den kommunizierten Planungen des BMAS orientiert.

Vorab muss darauf hingewiesen werden, dass die geplante Laufzeit eng bemessen erscheint, was die analytischen Möglichkeiten beschränkt und Fallzahlenrisiken birgt. Erstens lassen sich im gesetzten Zeitrahmen Maßnahmeneffekte nur über einen Zeitraum von maximal zwei Jahren abschätzen, weil die Eingangserhebung zur Erfassung von Zielgrößen, die nicht in den administrativen Daten vorhanden sind, zeitlichen Vorlauf benötigt und zugleich mit Blick auf die abschließende Berichtslegung die letzten empirischen Analysen nicht zu spät erfolgen dürfen. Angesichts der Erwartung, dass der sozio-ökonomische Integrationsprozess bei den Zielpersonen langsam verlaufen dürfte und auch die zu analysierenden Maßnahmen ihre Wirkungen erst allmählich entfalten, birgt dies die Gefahr, dass eigentlich vorhandene Wirkungen im Rahmen der Untersuchung (noch) nicht erkannt werden. Bei einer Verlängerung der Projektlaufzeit könnte die dritte Welle der Verlaufsbefragung entsprechend später terminiert werden, so dass eine bessere Abschätzung sich langsam entfaltender Maßnahmen zu erreichen wäre. Zugleich würde damit die zweite Welle der Verlaufserhebung, die bei einer kurzen Projektlaufzeit vor allem der Panelpflege dient, inhaltlich an Bedeutung gewinnen. Mittig zwischen erster und dritter Welle platziert, läge sie später als in einem Projektrahmen von 36 Monaten und würde so bessere Möglichkeiten bieten, die Dynamik von Integrationsprozessen zu erfassen.

Allerdings darf die dritte Erhebungswelle auch nicht zu spät gelegt werden, denn mit zunehmender zeitlicher Spanne der Verlaufsbefragung steigt die Wahrscheinlichkeit von Panelausfällen. Um diesen zu begegnen, müsste man entweder die Eingangsstichprobe vergrößern oder eine weitere Erhebungswelle zur Panelpflege einplanen; beides würde jedoch den Kostenaufwand für die Erhebung noch einmal deutlich erhöhen. Darum scheint eine Ausweitung der Projektlaufzeit über 48 Monate hinaus nicht sinnvoll.

Langfristige Maßnahmeneffekte, zumindest im Hinblick auf die Arbeitsmarktintegration, können in Zukunft im Rahmen von eventuellen späteren Folgeprojekten ohne gesonderte Erhebungen (die ohnehin wahrscheinlich nicht auf ausreichende Fallzahlen kämen) durchgeführt werden, indem die im geplanten Evaluationsvorhaben gewonnenen Individualdaten mit den Daten der künftigen Beschäftigtenstatistik verknüpft werden. Um diese Möglichkeit vorzubereiten, ist im Rahmen der Verlaufsbefragung die Zustimmung zu einer entsprechenden Verknüpfung bei den Befragten einzuholen. Darüber hinaus müssen Schlüsseldaten, die eine solche Verknüpfung ermöglichen, unter Einhaltung der datenschutzrechtlichen Bestimmungen ebenso wie die eigentlichen Befragungsdaten aufbewahrt und gesichert werden. Die im Rahmen des Evaluationsvorhabens gewonnenen Daten sollten dazu nach Abschluss des Projekts an ein geeignetes Forschungsdatenzentrum überstellt werden.

Eine solche Verknüpfung bietet sich auch an, um sich für einen späteren Zeitpunkt auch die Option zu erhalten, mögliche Spillover-Effekte von Maßnahmen zur Integration von Geflüchteten auf dem Arbeitsmarkt zu analysieren, also den Fragen nachzugehen, inwieweit Fortschritte bei der Integration dieser Zielgruppe möglicherweise zu Verdrängungseffekten bei anderen Zielgruppen, insbesondere Geringqualifizierten und Langzeitarbeitslosen, führen und andererseits die Beschäftigungschancen von komplementären Arbeitskräften verbessern. Solche Effekte könnten grundsätzlich anhand von Unterschieden in der Entwicklung der Arbeitsmarktposition von Personen außerhalb des Kontexts von Fluchtmigration ermittelt werden, die in vergleichbaren Arbeitsmärkten aktiv sind, die sich jedoch hinsichtlich der maßnahmenbedingten Integrationsfortschritte von den Personen im Kontext von Fluchtmigration unterscheiden. Die Identifikation von solchen Effekten auf der aggregierten Ebene ist allerdings nicht einfach und setzt voraus, dass die Maßnahmen zur Integration der Flüchtlinge erst einmal eine substantielle Beschäftigungswirkung entfalten. Da stark zu vermuten ist, dass sich diese Wirkungen innerhalb der geplanten Laufzeit des Evaluationsvorhabens noch nicht voll entfalten werden, scheint es sinnvoll, entsprechende Analysen erst in eventuellen Folgeprojekten vorzunehmen.

Die Projektlaufzeit über die vorgesehenen 36 Monate hinaus um bis zu weitere 12 Monate zu verlängern, würde noch weitere Vorteile mit sich bringen. Erstens ließen sich heterogene Maßnahmenwirkungen für Gruppen von Zielpersonen unterscheiden, die zu unterschiedlichen Zeitpunkten in die Maßnahme neu eintreten, etwa weil Lerneffekte bei den Anbietern von Hilfsangeboten oder bei den Personen im Kontext von Fluchtmigration auftreten. Dazu müsste die erste Welle der Panelbefragung über ein längeres Zeitfenster hinweg geführt werden, sodass die Entwicklung von mehreren Kohorten von Maßnahmenteilnehmerinnen und Maßnahmenteilnehmern (und der ihnen zugeordneten Vergleichspersonen) beobachtet werden kann.<sup>16</sup>

Zweitens kann die Möglichkeit, Eintritte über ein längeres Zeitfenster hinweg zu beobachten und anschließend zu verfolgen, sehr wichtig sein, um Risiken in der Durchführung des Vorhabens zu verringern. Es erscheint nämlich keineswegs gesichert, dass die Anzahl der Maßnahmeneintritte in einem engen Zeitfenster genügt, um – insbesondere bei den seltener eingesetzten Instrumententypen – überhaupt auf die Mindestfallzahlen zu kommen, die für belastbare Wirkungsaussagen benötigt werden. Bei einem engen Zeitfenster entsteht ein erhöhter Druck, die zur Verfügung stehende Bruttostichprobe weit auszuschöpfen, was die Kosten der Durchführung der Erhebung erhöht. Stattdessen Personen einzubeziehen, die bereits vor längerem in die Maßnahme eingetreten sind, ist keine überzeugende Alternative. Es fehlen dann belastbare Ausgangswerte für die nur per Befragung und nicht retrospektiv zu erfassenden Zielgrößen; vor allem aber ergibt sich ein Selektivitätsproblem: Teilnehmerinnen und Teilnehmer, die länger in der Maßnahme verblieben sind, könnten sich systematisch von

---

<sup>16</sup> Hierbei muss bei der Auswahl der regionalen Untersuchungseinheiten zum Zweck der Kausalanalyse eventuell nachgesteuert werden, falls sich die lokalen Abläufe bei der Zuweisung von Maßnahmen über die Zeit verändern.

Teilnehmerinnen und Teilnehmern unterscheiden, die neu in die Maßnahme eintreten. Diese Probleme lassen sich nur vermeiden, wenn mehrere Eintrittskohorten in ihrer Entwicklung verfolgt werden. Wird ein längeres Zeitfenster geöffnet, um Eintritte zu beobachten, muss der Untersuchungszeitraum nach hinten entsprechend verlängert werden, um Veränderungen der Zielgröße in einem angemessenen Abstand zum Maßnahmenbeginn beobachten und bewerten zu können.

Der in Abbildung 7.1 dargestellte Zeitplan basiert auf der Vorgabe einer dreijährigen Projektlaufzeit. In dieser Konstellation können die Wirkungsanalysen in den vier zentralen wirkungsanalytischen Modulen sinnvoll nur für eine Kohorte von Teilnehmerinnen und Teilnehmern durchgeführt werden, die im dritten oder vierten Quartal 2017 in die zu evaluierenden Maßnahmen eintreten. Würde man eine zweite Kohorte von Teilnehmerinnen und Teilnehmern einbeziehen, die erst im ersten und zweiten Quartal 2018 mit der Maßnahme beginnen, rückt der Abschluss der Panelerhebung für diese Gruppe zu nah an das Ende der Projektlaufzeit von 36 Monaten heran. Soll eine zweite Eingangskohorte untersucht werden, wofür die oben ausgeführten Argumente sprechen, wäre die Projektlaufzeit auf mindestens 42 Monate zu verlängern.

Abbildung 7.1 Zeitplan und Meilensteine

2017	10	Regionenauswahl Stichprobenziehung	Modul 1) Aktivierung und berufliche Eingliederung	Modul 2) Berufswahl und Berufsausbildung	Modul 3) Berufliche Weiterbildung	Modul 4) Aufnahme einer Erwerbstätigkeit Beschäftigung schaffende Maßnahmen	Modul 5) Sprachförderung	
	11							
	12							
2018	1							
	2							
	3							
	4							
	5							
	6							
	7	Zwischenberichte						
	8	Selektivitätsanalyse						
	9							
	10							
	11							
	12							
2019	1							
	2							
	3							
	4	Zwischenberichte						
	5	Wirkungsanalyse						
	6							
	7							
	8							
	9							
	10							
	11							
	12							
2020	1							
	2							
	3							
	4							
	5							
	6							
	7	Endberichte Module						
	8							
	9							
	10							
Abschlussbericht								

Darstellung des IZA.

Der Zeitplan für die Panelerhebung im Befragungsmodul geht davon aus, dass die erste Welle der Befragung möglichst früh nach dem Eintrittszeitpunkt der Teilnehmerinnen und Teilnehmer erfolgt, um zumindest näherungsweise noch eine Nullmessung der Ausgangsbedingungen und Zielgrößen zu

realisieren. Die Gefahr von Verzerrungen, sofern sich die Zielgrößen in Folge der Maßnahmenteilnahme unmittelbar stark verändern, lässt sich mit vertretbarem Ressourcenaufwand kaum vermeiden.<sup>17</sup>

Die zweite, im Wesentlichen der Panelpflege dienende Welle der Erhebung soll zehn bis zwölf Monate nach der Ersterhebung stattfinden. Die dritte, umfassender angelegte Erhebungswelle zur abschließenden Erfassung der nicht in den administrativen Daten vorhandenen Zielgrößen findet in etwa zwei Jahre nach der Erstbefragung gegen Ende der Projektlaufzeit statt.

Der Zeitplan sieht vor, dass die parallele Bearbeitung der kausalanalytischen Evaluationsmodule sofort mit Beginn der Laufzeit des Gesamtvorhabens beginnt. Die Auswahl der Stichproben für die Panelerhebung und die Übermittlung des zugehörigen Adressmaterials an den Auftragnehmer des Befragungsmoduls muss bis spätestens Anfang Januar erfolgen, damit die geplante Panel-Befragung rechtzeitig ins Feld kommt. Die Bearbeitung des kausalanalytischen Moduls zur Sprachförderung kann etwas später beginnen, weil darin keine eigenständige Stichprobenauswahl vorgesehen ist. Im Anschluss laufen alle kausalanalytischen Module nach einem inhaltlich und zeitlich einheitlichen Anforderungsschema, das bereits in den Leistungsbeschreibungen der Vorhaben spezifiziert werden muss.

Um die Projektfortschritte in den parallel laufenden Evaluationsmodulen nachvollziehbar zu dokumentieren und die Informationsbasis für gegebenenfalls erforderliche Abstimmungen zwischen den verschiedenen Modulen zu verbreitern, ist eine regelmäßige Berichterstattung über den Verlauf des Vorhabens hinweg eingeplant. Ein erster Zwischenbericht ist zehn Monate nach Projektbeginn vorgesehen. Dieser soll die sich bis dahin abzeichnenden Prozesse beim Maßnahmenzugang fokussieren. Zu dokumentieren ist insbesondere der Grad der quasi-experimentellen regionalen Variation, sowie die Selektivität der Teilnehmerinnen und Teilnehmer anhand beobachtbarer Merkmale. Zur Erstellung des ersten Zwischenberichts können neben den administrativen Daten auch die Ergebnisse der ersten Welle der Panelerhebung sowie der qualitativen Untersuchungen vor Ort herangezogen werden.

Ein zweiter Zwischenbericht soll nach weiteren zehn Monaten erstellt werden. Dieser soll bereits erste Hinweise auf potenzielle Maßnahmenwirkungen enthalten. Hierfür können die – beschränkten – Informationen zu Zielgrößen aus der zweiten Welle der Panelerhebung herangezogen werden. Zudem soll dieser Zwischenbericht ökonomische Analysen auf Basis der administrativen Daten enthalten.

Die Endberichte zu den Evaluationsmodulen 1-4, die umfassende Wirkungsanalysen auf Grundlage der in der dritten Welle der Panelbefragung erhobenen Zielgrößen enthalten und auch mögliche Erklärungen für die gemessenen Treatment-Effekte geben sollen, sind für das Frühjahr 2020 geplant. Die frühe Terminierung sorgt dafür, dass anschließend bis zum Abschluss des Gesamtvorhabens noch ausreichend Zeit für die Erstellung eines übergreifenden Endberichts durch die Geschäftsstelle bleibt. Zudem trägt diese konservative Zeitplanung möglichen Verzögerungen wegen unvorhergesehen auftretender besonderer Schwierigkeiten im Projektverlauf Rechnung.

Das Gesamtvorhaben soll bei einer Laufzeit von 36 Monaten im September 2020 mit der Vorlage eines zusammenfassenden und handlungsorientierten Abschlussberichts durch die Geschäftsstelle enden.

---

<sup>17</sup> Eine echte Nullmessung ließe sich nur mit einer Eingangskohorte von Personen im Kontext von Fluchtmigration erreichen, die zum Zeitpunkt der Erstmessung ausschließlich aus Nichtteilnehmerinnen und Nichtteilnehmern besteht und eventuell erst später in eine Maßnahme eintritt (oder auch nicht). Um ausreichende Fallzahlen an Teilnehmerinnen und Teilnehmern zu realisieren, müsste die Eingangsstichprobe jedoch sehr groß dimensioniert werden. Die Umsetzung dieses Ansatzes scheidet angesichts der zielgruppenbedingt stark erhöhten Erhebungskosten jedoch von vornherein aus.



## Literaturverzeichnis

- Abbring, J. und G. J. van den Berg (2003), The nonparametric identification of treatment effects in duration Models, *Econometrica*, 71, S. 1491-1517.
- Ahonen, E.Q., Benavides, F.G und J. Benach (2007), Immigrant populations, work and health—a systematic literature review, *Scandinavian Journal of Work, Environment & Health*, 33(2), S. 96-104.
- Aldashev, A., J. Gernandt und S. L. Thomsen (2009), Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection, *Labour Economics*, 16(3), S. 330-341.
- Andersson Joona, P. und Nekby, L. (2012): Intensive Coaching of New Immigrants: An Evaluation Based on Random Program Assignment, *The Scandinavian Journal of Economics*, 114, S. 575–600.
- Angrist, J.D. und A.B. Krueger (2001), Instrumental variables and the search for identification: From supply and demand to natural experiments, *Journal of Economic Perspectives*, 15(4), S. 69-85.
- Angrist, J. D. und J.-S. Pischke (2009), *Mostly harmless econometrics*, Princeton: University Press.
- Angrist, J. D. und J.-S. Pischke (2010), The credibility revolution in empirical economics: How better research design is taking the con out of econometrics, *Journal of Economic Perspectives*, 24(2), S. 3-30.
- Arbel, Y., Tobol, Y. und E. Siniver, Social Involvement and Level of Household Income among Immigrants: New Evidence from the Israeli Experience, IZA Diskussionspapier Nr. 6419, Bonn.
- Arntz, M., R. Wilke und H. Winterhager (2006), Regionenmatching im Rahmen der Evaluation der Experimentierklausel des § 6c SGB III: Methodische Vorgehensweise und Ergebnisse, ZEW Diskussionspapier Nr. 06-061, Mannheim.
- Ager, A. und A. Strang (2008), Understanding integration: a conceptual framework, *Journal of Refugee Studies*, 21(2), S. 166-191.
- Aslund, O. und D.-O. Rooth (2007), Do When and Where Matter? Initial Labor Market Conditions and Immigrant Earnings, *Economic Journal*, 117, 422-448.
- Aslund, O. und P. Johansson (2011), Virtues of SIN: Can Increased Public Efforts Help Disadvantaged Immigrants?, *Evaluation Review*, 35(4), 399–427.
- BA – Bundesagentur für Arbeit (2016), Hintergrundinformation Geflüchtete Menschen in den Arbeitsmarktstatistiken, Erste Ergebnisse, Statistik der Bundesagentur für Arbeit: Nürnberg.
- BAMF (2016), Aktuelle Zahlen zu Asyl, Ausgabe: November 2016, Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Basilio, L. und T. K. Bauer (2014), Transferability of human capital and immigrant assimilation: An analysis for Germany, SOEP Papers on Multidisciplinary Panel Research Nr. 671, Berlin: Deutsches Institut für Wirtschaftsforschung.
- Bauer, T. K., M. Fertig und C. M. Schmidt (2009), *Empirische Wirtschaftsforschung: Eine Einführung*, Berlin: Springer.
- Bevelander, P., M. Hagström und S. Rönqvist (2009), Resettled and included? The employment integration of resettled refugees in Sweden. Malmö Institute for Studies of Migration, Diversity and Welfare (MIM), Malmö University.
- Bleakley, H. und A. Chin (2010), Age at arrival, English proficiency, and social assimilation among U.S. immigrants, *American Economic Review: Applied Economics*, 2(1), S. 165-192.
- Boeri, T., M. de Phillipis, E. Patacchini und M. Pellizzari (2015), Immigration, housing discrimination and employment, *Economic Journal*, 125(August), S. F82-F114.

- Bogic, M. (2012). Factors associated with mental disorders in long-settled war refugees: Refugees from the former Yugoslavia in Germany, Italy and the UK. *British Journal of Psychiatry*, 200, S. 216-223.
- Bonin, H. (2016), Gewinne der Integration, Berufliche Qualifikation und Integrationstempo entscheiden über die langfristigen fiskalischen Kosten der Aufnahme Geflüchteter, böll.brief 1, Heinrich Böll-Stiftung: Berlin.
- Boockmann, B., S.L. Thomsen und T. Walter (2014), Intensifying the use of benefit sanctions: An effective tool to increase employment? *IZA Journal of Labor Policy*, 3:21.
- Borjas, G. J. (1995), Assimilation and changes in cohort quality revisited: What happened to immigrant earnings in the 1980s? *Journal of Labor Economics*, 13(2), S. 201-245.
- Brücker, H., E. Liebau, A. Romiti und E. Vallizadeh (2014), Anerkannte Abschlüsse und Deutschkenntnisse lohnen sich, *DIW-Wochenbericht*, 81(43), S. 1141-1151.
- Brücker, H., A. Hauptmann und E. Vallizadeh (2015a), Flüchtlinge und andere Migranten am deutschen Arbeitsmarkt: Der Stand im September 2015, *Aktuelle Berichte* 14/2015, Nürnberg.
- Brücker, H., A. Glitz, A. Lerche und A. Romiti (2015b), Occupational Recognition and Immigrant Labor Market Outcomes, Papier präsentiert auf der SOLE-Tagung 2015, Download: <http://www.sole-jole.org/16456.pdf>
- Bundesministerium des Innern und Ramboll-Management (2006), Evaluation der Integrationskurse nach dem Zuwanderungsgesetz. Abschlussbericht und Gutachten über Verbesserungspotenziale bei der Umsetzung der Integrationskurse. Berlin: Ramboll Management.
- Butschek, S. und T. Walter (2013), What Active Labour Market Programmes Work for Immigrants in Europe? A Meta-Analysis of the Evaluation Literature, *ZEW Discussion Paper No. 13-056*.
- Caliendo, M., D.A. Cobb-Clark und A. Uhlendorff (2015), Locus of Control and Job Search Strategies, *Review of Economics and Statistics*, 97(1), S. 88-103.
- Cangiano A. (2012), Immigration policy and migrant labour market outcomes in the European Union: New evidence from the EU Labour Force Survey, *FIERI Working Papers*.
- Chiswick, B.R. und P.W. Miller (2007), *The Economics of Language: International Analyses*. London: Routledge.
- Chiswick, B.R. und P.W. Miller (2009), The economics of language: An introduction and overview, in: T. Wiley, J. Sook Lee and R. Rumberger (Hrsg.), *The education of language minority immigrants in the United States*, Multilingual Matters Ltd., Bristol, UK, S. 72-91.
- Chiswick, B. R. und P. W. Miller (2014). International migration and the economics of language, in: B. R. Chiswick und P. W. Miller (Hrsg.), *Handbook of the Economics of Immigration*, Bd. 1A, Amsterdam: Elsevier, S. 211-270.
- Clausen, J., E. Heinesen, H. Hummelgaard, L. Husted, und M. Rosholm (2009), The Effect of Integration Policies on the Time Until Regular Employment of Newly Arrived Immigrants: Evidence from Denmark. *Labour Economics*, 16(4), S. 409-417.
- Connor, P. (2010), Explaining the refugee gap: Economic outcomes of refugees versus other immigrants. *Journal of Refugee Studies*, 23(3), S. 377-397.
- Constant, A. F. und K. F. Zimmermann (2005), Legal status at entry, economic performance, and self-employment proclivity: A bi-national study of immigrants, *IZA Diskussionspapier Nr. 1910*, Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Constant, A. F. und K. F. Zimmermann (2008), Measuring ethnic identity and its impact on economic behavior, *Journal of the European Economic Association*, 6(2-3), S. 424-433.

- Correa-Velez, I., A. G. Barnett und S. Gifford (2013), Working for a better life: Longitudinal evidence on the predictors of employment among recently arrived refugee migrant men living in Australia. *International Migration*, 53(2), S. 321–337.
- Cutler, D. A. und E. L. Glaeser (1997), Are ghettos good or bad? *The Quarterly Journal of Economics*, 112(3), S. 827–872.
- Damm, A. P. (2009), Ethnic enclaves and immigrant labor market outcomes: Quasi-Experimental evidence. *Journal of Labor Economics*, 27(2), S. 281–314.
- Danzer, A. M. und H. Ulku (2011), Integration, social networks and economic success of immigrants, *Kyklos*, 64(3), S. 342–365.
- Danzer, A. M. und F. Yaman (2012), Do ethnic enclaves impede immigrants' integration? Evidence from a quasi-experimental social-interaction approach, IZA Diskussionspapier Nr. 6939, Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Daumann, V., M. Dietz, B. Knapp und K. Strien (2015), Early Intervention - Modellprojekt zur frühzeitigen Arbeitsmarktintegration von Asylbewerberinnen und Asylbewerbern – Ergebnisse der qualitativen Begleitforschung, IAB-Forschungsbericht 3/2015, Nürnberg.
- De Paola, M. und G. Brunello (2016), Education as a Tool for the Economic Integration of Migrants, IZA Diskussionspapier Nr. 9836, Bonn.
- Dustmann, C. (1993), Earnings adjustment of temporary migrants, *Journal of Population Economics*, 6(2), S. 153–168.
- Dustmann, C. (1994), Speaking Fluency, Writing Fluency and Earnings of Migrants, *Journal of Population Economics*, 7(2), S. 133–156.
- Dustmann, C. und F. Fabbri (2003), Language Proficiency and Labour Market Performance of Immigrants in the UK. *Economic Journal*, 113(489), S. 695–717.
- Dustmann, C. und A. Glitz (2011), Migration and education, in: E. Hanushek (Hrsg.), *Handbook of Economics of Education*, Bd. 4, North Holland: Elsevier, S. 327–439.
- Dustmann, C. und A. van Soest (2002), Language and the earnings of immigrants, *Industrial and Labor Relations Review*, 55(3), S. 473–492.
- Edin, P.-A., P. Fredriksson und O. Aslund (2003), Ethnic enclaves and the economic success of immigrants - evidence from a natural experiment, *Quarterly Journal of Economics*, 118(1), S. 329–357.
- Europäische Kommission (2004): *Gemeinsamer Bericht über die soziale Eingliederung. Soziale Sicherheit und soziale Integration*. Europäische Kommission. Luxemburg.
- Esser, Hartmut (2001), *Integration und ethnische Schichtung*. Arbeitspapiere des Mannheimer Zentrums für Europäische Studien 40, Mannheim.
- Evans, O. und R. Murray, R. (2009), *The gateway protection programme: An evaluation*. Home Office: London.
- Filsinger, D. (2008), *Bedingungen erfolgreicher Integration – Integrationsmonitoring und Evaluation*, Expertisen und Dokumentationen zur Wirtschafts- und Sozialpolitik, WISO-Diskurs, Friedrich-Ebert-Stiftung: Bonn.
- Foged, M. und G. Peri (2016), Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data. *American Economic Journal: Applied Economics*, 8(2), S. 1–34.
- Gathmann, C. und N. Keller (2014), Returns to Citizenship? Evidence from Germany's Recent Immigration Reform, IZA Diskussionspapier Nr. 8064, Bonn: Forschungsinstitut zur Zukunft der Arbeit.

- Gautier, T. und Y. Zenou (2010), Car ownership and the labor market of ethnic minorities, *Journal of Urban Economics*, 67(3), S. 392-403.
- Grin, F. (1996), The Economics of Language: Survey, Assessment, and Prospects, *International Journal of the Sociology of Language*, 121(1), 17–44.
- Hauff, E. und P. Vaglum (1993). Integration of Vietnamese refugees into the Norwegian labor market: The impact of war trauma. *International Migration Review*, 388-405.
- Häußermann, H. und W. Siebel (2000), Wohnverhältnisse und Ungleichheit. In A. Harth, G. Scheller und W. Tessin (Hrsg.), *Stadt und soziale Ungleichheit*, VS Verlag für Sozialwissenschaften: Opladen, S. 120-140.
- Heckman, J. J. und B. Singer (1984), Econometric duration analysis, *Journal of Econometrics*, 24(1), S. 63-132.
- Heckman, J. J., H. Ichimura und P. E. Todd (1997), Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic Studies*, 64(4), S. 605-654.
- Heinesen, E., L. Husted und M. Rosholm (2011), The Effects of Active Labour Market Policies for Immigrants Receiving Social Assistance in Denmark, *IZA Discussion Paper No. 5632*.
- IAB (2015). Flüchtlinge und andere Migranten am deutschen Arbeitsmarkt: Der Stand im September 2015, Aktuelle Berichte des IAB 14/2015, Nürnberg.
- IAB (2016). IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse, IAB Forschungsbericht 14/2016, Nürnberg.
- Ibrahim, M. H., P. Sgro, F. Mansouri und C. Jubb (2010). Refugees' labour market access in Australia: A case study of Eritrea African immigrants. *Citizenship and Globalisation Research Papers* 1, 84-116.
- Imbens, G. W. und J. M. Wooldridge (2009), Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, 47(1), S. 5-86.
- IZA, DIW und infas (2006), Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission. Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen, Forschungsbericht F356b. Bundesministerium für Arbeit und Soziales.
- Jasso, G., D. Massey, M. R. Rosenzweig und J. P. Smith (2004), Immigrant health: selectivity and acculturation, *IFS Working Paper WP04/23*, London: The Institute for Fiscal Studies.
- Kalfa, E., M. Tani und M. Piracha (2016), Social capital and immigrants' labour market performance, *Papers in Regional Science*, 95(Supplement), S. S107-S126.
- Kalfa, E. und M. Piracha (2013), Immigrants' educational mismatch and the penalty of over-education, *IZA Diskussionspapier Nr. 7721*, Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Kanas, A., B. R. Chiswick, T. Van der Lippe und F. Van Tubergen (2012), Social contacts and the economic performance of immigrants: A panel study of immigrants in Germany, *International Migration Review*, 46(3), S. 680-709.
- Kronauer, M. (2006). Teilhabekonzept, Ausgrenzung, gesellschaftliche Zugehörigkeit. In: P. Bartelheimer (Hrsg.): *Berichterstattung zur sozioökonomischen Entwicklung Deutschlands – Zweiter Bericht. Zwischenbericht, Teil 1*. Göttingen, S. 59–63.
- Lemaître, G. (2007). The integration of immigrants into the labour market: The case of Sweden, *OECD Social, Employment And Migration Working Papers No 48*.
- Luik, M.-A., H. Emilsson und P. Bevelander (2016), Explaining the male native-immigrant employment gap in Sweden: The role of human capital and migrant categories, *IZA Diskussionspapier Nr. 9943*, Bonn: Forschungsinstitut zur Zukunft der Arbeit.

- Mamgain, V. und K. Collins (2003). Off the boat, now off to work: Refugees in the labour market in Portland, Maine. *Journal of Refugee Studies*, 16(2), S. 113-146.
- Miranda, A. und Y. Zhu (2013), English deficiency and the native immigrant wage gap. *Economics Letters*, 118(1), S. 38-41.
- Mirbach, T., K. Triebel und S. Bartsch (2014). Programmevaluation – ESF-Bundesprogramm zur arbeitsmarktlichen Unterstützung für Bleibeberechtigte und Flüchtlinge mit Zugang zum Arbeitsmarkt, Abschlussbericht: Lawaetz-Stiftung und Univation, Hamburg.
- OECD (2014). *International Migration Outlook 2014*, OECD Publishing. Paris.
- OECD (2015). *Is this humanitarian migration crisis different?*, Migration Policy Debates No. 7, OECD Publishing. Paris.
- OECD (2016). *Erfolgreiche Integration: Flüchtlinge und sonstige Schutzbedürftige*, OECD Publishing. Paris.
- Ohliger, R. (2007). Zuwandererintegration und -monitoring in Deutschland, in: *Der Beauftragte des Senats von Berlin für Integration und Migration* (Hrsg.): *Indikatoren zur Messung von Integrationserfolgen*: Berlin, S. 18–28.
- Ott, E. (2011). *Get up and go: Refugee resettlement and secondary migration in the USA*, UNHCR New Issues in Refugee Research No. 219.
- Ott, E. (2013). *The Labour Market Integration of Resettled Refugees*. United Nations High Commissioner for Refugees, Policy Development and Evaluation Service, Geneva.
- Pedace, R. und S. R. Kumar (2012), A warm embrace or a cold shoulder? Wage and employment outcomes in ethnic enclaves, *Contemporary Economic Policy*, 31(1), S. 93-110.
- Pelkowski, J. M. und M. C. Berger (2004), The impact of health on employment, wages, and hours worked over the life cycle, *The Quarterly Review of Economics and Finance*, 44(1), S. 102-121.
- Platts-Fowler, D. und D. Robinson (2011). *An Evaluation of the Gateway Protection Programme*. Sheffield Hallam University: Sheffield.
- Potocky Tripodi, M. (2001). Micro and macro determinants of refugee economic status. *Journal of Social Service Research*, 27(4), S. 33-60.
- Potocky-Tripodi, M. (2003). Refugee economic adaptation: Theory, evidence, and implications for policy and practice. *Journal of Social Service Research*, 30(1), S. 63-91.
- Potocky-Tripodi, M. (2004). The role of social capital in immigrant and refugee economic adaptation. *Journal of Social Service Research*, 31(1), S. 59-91.
- Rinne, U. (2013), *The Evaluation of Immigration Policies*, in: Constant, Amelie F. und Klaus F. Zimmermann (Eds.), *International Handbook on the Economics of Migration*, Cheltenham, UK, and Northampton, USA: Edward Elgar Publishing, Kapitel 28, S. 530-551.
- RWI und ISG (2006), *Evaluation der Umsetzung der Vorschläge der Hartz-Kommission – Arbeitspaket 1, Modul 1f, Verbesserung der beschäftigungspolitischen Rahmenbedingungen und Makrowirkungen der aktiven Arbeitsmarktpolitik*, Gutachten im Auftrag des Bundesministeriums für Arbeit und Soziales. Berlin.
- Sarvimäki, M. and K. Hämäläinen (2010), *Assimilating Immigrants: The Impact of an Integration Program*. HECER Discussion Paper 306, Helsinki Center of Economic Research (HECER), Helsinki.
- Schwientek, C. (2016), *Are immigrants overeducated in Germany? Determinants and wage effects of educational mismatch*, FAU Discussion Paper in Economics Nr. 7/2016, Nürnberg: Friedrich-Alexander Universität Erlangen-Nürnberg, Download: <https://www.iwf.rw.fau.de/files/2015/12/07-2016-1.pdf>.

- Shields, M. A., and S. Wheatley Price (2001), Language Fluency and Employment Prospects: Evidence from Britain's Ethnic Minorities. *Applied Economic Letters*, 8, S. 741-745.
- Shields, M. A., and S. Wheatley Price (2002), The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics*, 15, S. 137-160.
- Siegert, M. (2006), Integrationsmonitoring – State of the Art in internationaler Perspektive, Studie im Auftrag des Bundesamts für Migration und Flüchtlinge (BAMF), Nürnberg.
- Siraj, T. (2007). The effect of refugees' homeland education on their earnings in the United States. Thesis.
- Strand, P. J. (1984). Employment predictors among Indochinese refugees. *International Migration Review*, 18(1), S. 50-64.
- Sulaiman-Hill, C. M. und S. C. Thompson (2013). Learning to fit in: An exploratory study of general perceived self efficacy in selected refugee groups. *Journal of Immigrant and Minority Health*, 15(1), S. 125-131.
- Svantesson, E. und T. N. Aranki (2006), Do Introduction Programs Affect the Probability of Immigrants getting Work? Working Paper 2006:3, Swedish Business School, Örebro University, Örebro.
- SVR (2016). Was wir über Flüchtlinge (nicht) wissen. Berlin: Forschungsbereich beim Sachverständigenrat deutscher Stiftungen für Integration und Migration (SVR).
- Takeda, J. (2000). Psychological and economic adaptation of Iraqi adult male refugees. *Journal of Social Service Research*, 26(3), S. 1-21.
- Walter, T., H. Bonin, S. Butschek, H. Schütz, H. Schröder, P. Knerr, J. Steinwede und S. L. Thomsen (2014): Evaluation „Programm zur berufsbezogenen Sprachförderung für Personen mit Migrationshintergrund (ESF-BAMF-Programm)“. Bonn: Abschlussbericht im Auftrag des Bundesministeriums für Arbeit und Soziales.
- Thomsen, S.L., T. Walter und A. Aldashev (2013), Short-Term Training Programs for Immigrants in the German Welfare System: Do Effects Differ from Natives and Why? *IZA Journal of Migration* 2013, 2:24.
- Thränhardt, D. (2015): Die Arbeitsmarktintegration von Flüchtlingen in Deutschland: Humanität, Effektivität, Selbstbestimmung, Bertelsmann Stiftung: Gütersloh.
- Tran, T. V. (1991). Sponsorship and employment status among Indochinese refugees in the United States. *International Migration Review*, 25(3), S. 536-550.
- Yao, Y. und J. van Ours (2015), Language skills and labor market performance of immigrants in the Netherlands, *Labour Economics*, 34(1), S. 76-85.
- Warman, C. (2007), Ethnic enclaves and immigrant earnings growth, *Canadian Journal of Economics*, 40(2), S. 401-422.
- Waxman, P. (2001). The economic adjustment of recently arrived Bosnian, Afghan and Iraqi refugees in Sydney, Australia. *International Migration Review*, 35(2), 472-505.
- Weichselbaumer, D. (2016), Discrimination against Female Migrants Wearing Headscarves, IZA Diskussionspapier Nr. 10217, Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Werth, M., S. Delfs und W. Stevens (1997), Using Indicators of Integration for the Implementation of Integration policies – Summary and Conclusions, in: Council of Europe (Hrsg.), *Measurement and Indicators of Integration, Community Relations*, Council of Europe Publishing: Strasbourg, S. 187-194.
- Worbs, S. und E. Bund (2016), Asylberechtigte und anerkannte Flüchtlinge in Deutschland – Qualifikationsstruktur, Arbeitsmarkteteiligung und Zukunftsorientierungen, BAMF-Kurzanalyse 1/2016, Nürnberg.

ZEW, IAB und IAT (2006), Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission.  
Modul 1d: Eingliederungszuschüsse und Entgeltsicherung. Forschungsbericht F356c,  
Bundesministerium für Arbeit und Soziales, Berlin.

ZEW, Ramboll und infas (2010), Vorstudie zur Evaluation von Fördermaßnahmen für Jugendliche im SGB II und  
SGB III, Endbericht, Mannheim.

Zimmermann, K. F. (2007), Migrant ethnic identity: concept and policy implications. *Ekonomia*, 10(1), S. 1-17.

Diese Publikation wird im Rahmen der Öffentlichkeitsarbeit des Bundesministeriums für Arbeit und Soziales kostenlos herausgegeben. Sie darf weder von Parteien noch von Wahlbewerbern oder Wahlhelfern während des Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für Europa-, Bundestags-, Landtags- und Kommunalwahlen. Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen, an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Publikation dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Bundesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte. Außerdem ist diese kostenlose Publikation - gleichgültig wann, auf welchem Weg und in welcher Anzahl diese Publikation dem Empfänger zugegangen ist - nicht zum Weiterverkauf bestimmt.

Erstellt im Auftrag des Bundesministeriums für Arbeit und Soziales.

Die Durchführung der Untersuchungen sowie die Schlussfolgerungen aus den Untersuchungen sind von den Auftragnehmern in eigener wissenschaftlicher Verantwortung vorgenommen worden. Das Bundesministerium für Arbeit und Soziales übernimmt insbesondere keine Gewähr für die Richtigkeit, Genauigkeit und Vollständigkeit der Untersuchungen.

Alle Rechte einschließlich der fotomechanischen Wiedergabe und des auszugsweisen Nachdrucks vorbehalten.