

Forschungsauftrag 5/05

**Wirksamkeit der Bekämpfung der Schwarzarbeit
durch die „Finanzkontrolle Schwarzarbeit“:**

**Empirische Untersuchung der Entwicklung der Schwarzarbeit
und des hierauf bezogenen Bewusstseins- und Wertewandels
– Machbarkeitsstudie –**

Abschlussbericht an das Bundesministerium der Finanzen
durch das

Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW)

Mannheim, im Dezember 2005

Verfasser des Berichts:

PD. Dr. Bernhard Boockmann

Dipl.-Vw. Johannes Rincke

Non-technical Summary

Im Zentrum des Berichts steht die Frage, ob und wie die kausale Wirkung der Maßnahmen der Finanzkontrolle Schwarzarbeit (FKS) auf das Verhalten der Arbeitsmarktbeteiligten ermittelt werden kann. Hierzu werden Verfahren und Vorgehensweisen vorgeschlagen, die belastbare empirische Ergebnisse über Art und Ausmaß der Wirkung versprechen. Die Maßnahmen betreffen die Arbeitsbereiche Prävention, Prüfungen und Ermittlungen sowie Ahndung der FKS. Als Indikatoren werden der Umfang der Schwarzarbeit und der illegalen Beschäftigung, der Umfang der legalen Beschäftigung und das durch Umfragen erfasste Rechtsbewusstsein verwendet.

Eine Evaluation sollte mit deskriptiven Analysen der bei der FKS vorhandenen Datenbestände beginnen, da diese bisher noch nicht systematisch ausgewertet worden sind. Insbesondere die Ergebnisse verdachtsloser Prüfungen können bei vorsichtiger Interpretation Aufschluss über das Ausmaß von Schwarzarbeit und illegaler Beschäftigung geben. Darüber hinaus erlauben die Daten eine genaue Darstellung des Einsatzes der Bekämpfungsmaßnahmen. Da in die vorhandenen Daten jedoch eine Vielzahl unbeobachteter Faktoren wie insbesondere die Risikoanalyse der beteiligten Beamten eingeht, sollte eine quantitative Evaluation der Tätigkeit der FKS auf Daten beruhen, die in ausgewählten Branchen unter besonders kontrollierten Umständen erhoben werden. Diese können zusätzlich mit vorhandenen Informationen verknüpft werden. Für diese Bereiche könnte man dann auch wesentlich präzisere Aussagen über die Verbreitung der Schwarzarbeit treffen, als dies bislang möglich ist. Daneben bieten vorhandene Daten über die legale Beschäftigung ebenfalls eine repräsentative Datenbasis.

Für eine Wirkungsanalyse werden neben der Erörterung der grundsätzlichen Anforderungen einer Evaluation detaillierte Konzepte für fünf Evaluationsprojekte vorgelegt. Die Projekte sind modular konzipiert, so dass sie einzeln oder in unterschiedlichen Kombinationen umgesetzt werden können. Ihre Auswahl richtet sich nach dem Grundsatz, dass alle Arbeitsbereiche der FKS erfasst und die möglichen Vorgehensweisen bei der Evaluation in ganzer Breite dargestellt werden. Um die Verwendbarkeit der Projektvorschläge für eine künftige Hauptstudie zu erhöhen, werden jeweils mögliche Varianten diskutiert. Leitender Gesichtspunkt ist, dass statistisch signifikante Ergebnisse immer dann gefunden werden sollen, wenn in der Realität eine Wirkung der Maßnahmen der FKS vorliegt. Nur dann sind die Ergebnisse zugleich dazu geeignet, den gesellschaftlichen Nutzen des Maßnahmeeinsatzes der FKS in aussagekräftiger Weise zu quantifizieren und den entsprechenden Kosten gegenüberzustellen.

Ein solches Evaluationshaben verursacht bei der FKS in mehrerer Hinsicht zum Teil erheblichen Aufwand. Dieser entsteht zum einen bei der Erstellung der Datenbasis. Zum anderen sehen einige Projekte vor, dass die Maßnahmen der FKS für die Zwecke der Evaluation in bestimmter Weise gesteuert werden. Hier entstehen in der Umsetzung mögliche Akzeptanzprobleme innerhalb der FKS und bei den Arbeitsmarktbeteiligten. Eine sorgfältig konzipierte Kommunikationsstrategie muss daher die Durchführung der Evaluation begleiten.

Die vorliegende Studie macht deutlich, dass eine Evaluation der Tätigkeit der FKS auf zum Teil erhebliche methodische und technische Probleme stößt. Sie reagiert auf diese Schwierigkeiten, indem sie die Evaluation von ausgewählten Teilbereichen der Tätigkeit der FKS vorsieht und hierauf zugeschnittene Konzepte entwickelt. Diese Konzepte sind nicht im Sinne von Gebrauchsanweisungen zu verstehen, die mechanisch umgesetzt werden sollten, zumal die Durchführung einer Evaluation auch davon abhängen würde, wie der Auftraggeber die Fragestellung der Evaluation präzisiert. Insgesamt erweist sich, dass trotz der zahlreichen und schwerwiegenden Probleme viel versprechende Ansatzpunkte für eine Evaluation bestehen.

Inhaltsverzeichnis

1	Einleitung und Fragestellungen der Evaluation	5
1.1	Vorbemerkung	5
1.2	Abgrenzung des Evaluationsvorhabens	6
2	Ausgangsbedingungen der Evaluation	12
2.1	Struktur, Instrumente und Einsatzgebiete der FKS	12
2.2	Zur Verfügung stehende Daten	18
3	Bestandteile einer Evaluation der verhaltenslenkenden Wirkungen der FKS	24
3.1	Deskriptive Analysen der verwendeten Indikatoren	24
3.2	Schätzung von Kausaleffekten	26
3.3	Implementationsanalysen	34
3.4	Kosten-Nutzen-Analyse	35
3.5	Vermittlungs- und Kommunikationsstrategie	36
4	Einzelne Evaluationsprojekte zur Ermittlung kausaler Wirkungen der Tätigkeit der FKS	38
4.1	Regionales Prüfklima	39
4.2	Ankündigung von Prüfungen	49
4.3	Leistungsempfänger	57
4.4	Personelle Ausstattung	63
4.5	Wiederholungstäter	67
5	Schlussfolgerungen	71
6	Projektteam	73
7	Literaturverzeichnis	75
	Anhang: Überblick über die Evaluationsmethoden	77

1 Einleitung und Fragestellungen der Evaluation

1.1 Vorbemerkung

Der vorliegende Bericht baut auf zwei forschungsbegleitenden Gesprächen beim BMF in Berlin sowie mehreren Interviews und Gesprächen mit Mitarbeitern und Mitarbeiterinnen der FKS auf:

- Interviews mit Mitarbeitern der Abteilung FKS bei der OFD Köln am 26. Juli 2005;
- einem Interview mit Mitarbeitern des Standortes Trier des HZA Koblenz sowie der Außenstelle Neustadt/Weinstraße der Abteilung FKS am 11. August 2005;
- einem Interview mit Mitarbeitern des HZA München am 24. August 2005;
- Diskussion der Evaluationsvorschläge mit Mitarbeitern der Abteilung FKS am 4. Oktober 2005.

Diese Informationsgespräche bilden zusammen mit unterschiedlichen Arbeitsmaterialien und Aufstellungen, die von der FKS zur Verfügung gestellt wurden, sowie weiteren Dokumenten und Texten die Grundlage für den vorliegenden Bericht. Neben dem ZEW war Prof. Dr. Michael Lechner (Universität St. Gallen) an der Erstellung dieses Berichts beteiligt.

Der Bericht ist folgendermaßen gegliedert: Nach einer Definition des Gegenstands und des Ziels der in diesem Bericht untersuchten möglichen Evaluationen in diesem Abschnitt wird in den Abschnitt 2 die Ausgangslage für die Evaluation beschrieben: Welche Maßnahmen der FKS zur Bekämpfung der Schwarzarbeit sind zu evaluieren und welches sind mögliche Datengrundlagen? Die Bestandteile einer Evaluation werden in Abschnitt 3 erläutert. Hierzu gehören die deskriptive Auswertung vorhandener Daten, die Durchführung quantitativer Wirkungsanalysen mit dem Ziel der Schätzung kausaler Effekte, qualitative Untersuchungsschritte im Rahmen einer Implementationsanalyse, eine Kosten-Nutzen-Analyse sowie Vorschläge für ein Kommunikationskonzept, das die Zielsetzung einer Evaluation den beteiligten Mitarbeitern der FKS intern vermitteln soll.

Während die Darstellung in den Abschnitten 2 und 3 unabhängig von der Evaluation bestimmter Maßnahmen ist, werden in Abschnitt 4 Vorschläge für die Durchführung einzelner Projekte gemacht. Sie betreffen die Wirkung bestimmter Maßnahmen der FKS und entwickeln eine darauf jeweils zugeschnittene Vorgehensweise. Dabei wird auch diskutiert, auf welche anderen Maßnahmen und Bereiche die vorgestellte Methodik angewendet werden kann. Abschnitt 5 enthält Schlussfolgerungen und Empfehlungen für das weitere Vorgehen. Eine Darstellung der Evaluationsmethoden ist im Anhang enthalten, ebenso ein Überblick über in der Literatur vorhandenen Ansätze zur Evaluation von Bekämpfungsmaßnahmen in den Bereichen Steuerhinterziehung, Schwarzarbeit sowie allgemeine Kriminalität.

1.2 Abgrenzung des Evaluationsvorhabens

1.2.1 Definition von Schwarzarbeit

Unter Schwarzarbeit verstehen wir gemäß § 1 des Schwarzarbeitsbekämpfungsgesetzes die folgenden Erscheinungsbilder illegaler Beschäftigung:

- Beschäftigung von Arbeitnehmern, ohne dass die auf Grund der Dienst- oder Werkleistungen ergebenden sozialversicherungsrechtlichen Melde-, Beitrags- oder Aufzeichnungspflichten erfüllt werden;
- Erbringung von Dienst- oder Werkleistungen ohne Erfüllung der sich daraus ergebenden steuerlichen Pflichten;
- Nichterfüllung von Mitteilungspflichten gegenüber dem Sozialleistungsträger durch Empfänger von Sozialleistungen, die Dienst- oder Werkleistungen erbringen;
- Nichtanzeige des Beginns eines selbstständigen Betriebes in einem stehenden Gewerbe bzw. Nichterwerb der Reisegewerbekarte;
- Betreiben eines zulassungspflichtigen Handwerks als stehendes Gewerbe, ohne in der Handwerksrolle eingetragen zu sein.

Die beiden letztgenannten Verhaltensweisen fallen allerdings nicht in die Zuständigkeit der FKS, sondern in die der Länderbehörden. Daneben deckt die Studie auch folgende Erscheinungsbilder illegaler Beschäftigung ab:

- illegale Arbeitnehmerüberlassung, also Arbeitnehmerüberlassung ohne Verleiherlaubnis der Bundesagentur für Arbeit;
- illegale Ausländerbeschäftigung, also Beschäftigung von Ausländern ohne Arbeitserlaubnis oder eines Aufenthaltstitels;
- Verstöße gegen das Arbeitnehmer-Entsendegesetz.

Die Begehungsarten sind sehr vielfältig, so dass wir hierauf an dieser Stelle nicht im Einzelnen eingehen können (siehe dazu Deutscher Bundestag, 2005). Wenn nicht anders vermerkt, beziehen wir im Folgenden zur Vereinfachung auch die illegale Beschäftigung in den Begriff der Schwarzarbeit mit ein.

1.2.2 Fragestellungen der Evaluation

Im Zentrum des Projekts steht die Frage, ob und wie durch begleitende Erhebungen und mit Hilfe der bei der Finanzkontrolle Schwarzarbeit (FKS) anfallenden Daten die Wirkung von Bekämpfungsmaßnahmen auf das Verhalten der Arbeitsmarkteteiligen gemessen werden kann. Hierbei handelt es sich um eine Wirkungsanalyse unterschiedlicher Instrumente der Prävention, der Prüfung und Ermittlung sowie der Ahndung. Unter dem Verhalten der Arbeitsmarkteteiligen verstehen wir das Angebot oder die Nachfrage von Schwarzarbeit oder legaler Arbeit. Unmittelbar an das beobachtete Verhalten schließen sich Fragen an, die das Rechtsbewusstsein dieser Personengruppen betreffen.

Die Bekämpfungsmaßnahmen der FKS können unterschiedliche Wirkungen auf das Verhalten haben:

- Die Maßnahmen der FKS haben allgemein eine generalpräventive Wirkung auf potentielle Anbieter und Nachfrager von Schwarzarbeit. Dies gilt im Besonderen für die explizit präventiven Maßnahmen, die im Arbeitsbereich Prävention der FKS durchgeführt werden, wie Präventionsprüfungen, Presse- und Öffentlichkeitsarbeit. Es gilt darüber hinaus aber auch für Maßnahmen, die Schwarzarbeit aufdecken, sowie für Strafen und Bußgelder, da von diesen eine Abschreckungswirkung ausgeht.
- Neben der Abschreckungswirkung beeinflusst die Tätigkeit der FKS die öffentliche Wahrnehmung und Bewertung des Phänomens Schwarzarbeit und dadurch die „Steuermoral“ und das Verhalten der Arbeitsmarktbeteiligten. Dabei kann die Steuermoral durch Bekämpfungsmaßnahmen grundsätzlich positiv, aber auch negativ beeinflusst werden. Letzterer Fall kann durch eine Verdrängung intrinsischer Motivation infolge externer Anreize entstehen. Während die präventive Wirkung auf dem Kalkül über die Erträge und möglichen negativen Folgen der Schwarzarbeit beruht und insofern dem ökonomischen Verhaltensmodell entspricht, ist dieser zweite Aspekt eher durch soziologische oder psychologische Begriffe zu erfassen.
- Zusätzlich zu einer Generalprävention ist es denkbar, dass Strafen und Bußgelder eine spezifische Prävention bei einzelnen Individuen bewirken. Personen, die einer Ordnungswidrigkeit oder Straftat überführt worden sind und insofern Kontakt mit der Prüfung, Ermittlung und Ahndung durch die FKS hatten, haben möglicherweise aufgrund dieser Erfahrung eine geringere Neigung, erneut Schwarzarbeit anzubieten oder nachzufragen. Die Tätigkeit der FKS kann überdies den Effekt haben, dass potentielle Täter an der Ausübung von Schwarzarbeit gehindert werden, im Extremfall durch die Verhängung von Gefängnisstrafen durch die Justiz.

Der Effekt der Generalprävention steht im Folgenden im Vordergrund. Die FKS kann hierbei als Sender, die einzelnen Arbeitsmarktbeteiligten als Empfänger von Botschaften verstanden werden. Die Botschaften haben einen Einfluss darauf, wie die Empfänger das Risiko der Aufdeckung von Schwarzarbeit und die Höhe der zu erwartenden Strafen und Bußgelder wahrnehmen. Wenn auch die Instrumente im Sinne einer Generalprävention wirken, heißt dies doch nicht, dass alle Individuen in gleicher Weise von diesen Botschaften erfasst werden. Einige davon können z.B. durch eine gezielte Öffentlichkeitsarbeit oder Prüfungsaktivität nur einer bestimmten Gruppe von Empfängern vermittelt werden. Diese Botschaften sind im Folgenden das, was in der Literatur der Evaluation als „Treatment“ oder „Maßnahme“ bezeichnet wird, deren Wirkung auf das Verhalten zu bewerten ist.

Die Untersuchungen können zugleich auch Erkenntnisse über das Ausmaß und die Entwicklung der Schwarzarbeit geben. Da im Zuge der Evaluation Datensätze angelegt werden, die eine (zumindest annähernd) repräsentative Abdeckung der Arbeitsmarktbeteiligten in den untersuchten Bereichen gewährleisten, entsteht hierbei kein zusätzlicher Aufwand. Allerdings sind die untersuchten Bereiche in branchen- und berufsmäßiger, regionaler und zeitlicher

Hinsicht begrenzt, so dass das Projekt keine globalen Zahlen über die Verbreitung von Schwarzarbeit ergeben wird.

An die Frage nach der Wirkung der Instrumente der FKS schließt sich eine Kosten-Nutzen-Analyse an, die der Überprüfung der Effizienz der Maßnahmen dient. Die Kosten-Nutzen-Analyse fragt danach, ob der Aufwand, der mit dem Einsatz bestimmter Instrumente verbunden ist, geringer oder höher ist als der Ertrag. Als Ertrag wird in dieser Studie der fiskalische Nutzen einer Maßnahme betrachtet. Dieser entsteht bei der Zurückdrängung von Schwarzarbeit durch eine Verminderung von missbräuchlich bezogenen Sozialleistungen und bei zusätzlich geschaffener legaler Beschäftigung durch zusätzliche Beiträge zu den Systemen der Sozialversicherung sowie zusätzliches Steueraufkommen. Darüber hinaus muss auch das Aufkommen berücksichtigt werden, das aufgrund von zusätzlichen Sanktionen entsteht, beispielsweise durch zusätzlich vereinnahmte Bußgelder.

Eine weitere Frage, über die eine Evaluation Aufschluss geben kann, ist die Effektivität von Maßnahmen der Prüfung hinsichtlich der Aufdeckung von Schwarzarbeit. Dabei sind einerseits die fiskalischen Wirkungen von Interesse: liegen die zusätzlich verhängten Bußgelder ohne Berücksichtigung der präventiven Effekte über den Kosten, die der zusätzliche Einsatz hervorruft? Andererseits ist die Zahl der aufgedeckten Fälle von Schwarzarbeit ein wichtiger Zwischenindikator, weil die präventive Wirkung von Prüfungen erst dann zum Tragen kommt, wenn diese das Risiko der Aufdeckung von Schwarzarbeit erhöhen. Im Zentrum der folgenden Ausführungen steht die direkte Schätzung von präventiven Effekten, jedoch wird in einigen Teilen des Berichts vorgeschlagen, den Zwischenindikator zusätzliche Aufdeckung ebenfalls zu erheben.

1.2.3 Wahl der Indikatoren

Das Verhalten wird mit Hilfe unterschiedlicher Indikatoren gemessen, die sich erstens auf das Volumen oder andere Maße des Umfangs der Schwarzarbeit beziehen, zweitens auf die legale Beschäftigung und drittens auf das Rechtsbewusstsein. Diese Arten von Indikatoren sollen sich gegenseitig ergänzen. Während die Wirkung auf die Schwarzarbeit den unmittelbaren Einfluss der Tätigkeit der FKS auf das Verhalten der Arbeitsmarktbeteiligten widerspiegelt, ist die Wirkung auf die legale Beschäftigung von weiteren Verhaltensweisen abhängig und stellt insofern einen indirekten Wirkungszusammenhang dar. Insofern als die Bekämpfung der Schwarzarbeit vor allem das Ziel hat, die Gefährdung legaler Beschäftigung durch illegale Anbieter zu unterbinden, ist die legale Beschäftigung eine wichtige wirtschaftspolitische Zielgröße für die Tätigkeit der FKS. Das Rechtsbewusstsein erfasst im Unterschied zu den Indikatoren, die das tatsächliche Verhalten abbilden, die längerfristigen Wirkungen der Tätigkeit der FKS. Hierbei geht es um die Herausbildung von Normen, die ihrerseits auf das künftige Verhalten einwirken.

Zur Erfassung des Umfangs von Schwarzarbeit und legaler Beschäftigung können Fallzahlenindikatoren, Volumenindikatoren und Wertindikatoren verwendet werden. Fallzahlenindikatoren für die Schwarzarbeit sind insbesondere die Zahl der bei Prüfungen festgestellten Beanstandungen, die Zahl der eingeleiteten Ordnungswidrigkeits- oder Strafverfahren oder

die Zahl der Verurteilungen. Diese Indikatoren könnten auf die Zahl der durchgeführten Prüfungen bezogen werden. Als Volumenindikator könnte die Dauer der in der Schwarzarbeit geleisteten Arbeitszeit ermittelt werden; bei der Erweiterung der Fallzahlen- zu Volumenindikatoren können unterschiedliche Berechnungs- oder Imputationsverfahren verwendet werden. Ein Wertindikator könnte die ermittelte Schadenssumme sein, oder es können indirekte Verfahren zur Bewertung des Volumens der Schwarzarbeit angewandt werden. Beispielsweise ist an die Ansetzung branchen- und qualifikationsspezifischer Stundenlöhne zu denken. Die Verwendung dieser Indikatoren richtet sich nach der zugrunde liegenden Politikfrage.

Ein Fallzahlenindikator in der zweiten Gruppe ist die Zahl der legal Beschäftigten, wobei unterschiedliche Abgrenzungen verwendet werden können, je nachdem ob öffentlich geförderte Beschäftigung, Minijobs oder andere Nichtstandardformen der Beschäftigung sowie selbständige Beschäftigung mit berücksichtigt werden. Ein Volumenmaß wäre die Gesamtzahl der legalen gearbeiteten Arbeitsstunden, ein Wertmaß die Bruttolohn- und Gehaltssumme. Für die fiskalische Betrachtung in der Kosten-Nutzen-Analyse geht die Bruttolohn- und Gehaltssumme in die Berechnung des durch die Tätigkeit der FKS zusätzlich generierten Steueraufkommens ein.

Die Verwendung des dritten Ergebnisindikators Rechtsbewusstsein ist aus zwei Gründen sinnvoll. Zum einen, weil es sich um einen „vorlaufenden Indikator“ für das tatsächliche Niveau der Schwarzarbeit handelt. Zum anderen ist der Vergleich zu den festgestellten Fällen von Schwarzarbeit interessant, weil eine Wirkung der Maßnahmen der FKS entweder durch einen Bewusstseinswandel oder aber durch die direkte Präventionswirkung entstehen kann. Im ersten Fall wäre von längerfristigen Wirkungen der Tätigkeit der FKS auszugehen. Dieser qualitative Indikator kann durch entsprechende Antwortkategorien in Interviews erhoben werden.

Neben Verhaltensindikatoren sind Indikatoren für die durchgeführten Maßnahmen und Indikatoren für die ökonomischen Randbedingungen zu bilden. Auch bei den Maßnahmeindikatoren sind volumenmäßige und wertmäßige Indikatoren zu unterscheiden. In den meisten im Evaluationskonzept enthaltenen Fällen sind die Maßnahmen verhältnismäßig eng definiert und können durch eine Indikatorvariable erfasst werden, die angibt, ob einzelne Individuen oder Gruppen von Individuen der Maßnahme ausgesetzt waren oder nicht. In anderen Fällen, wie z.B. der Höhe von Strafen oder Bußgeldern, ist die Maßnahmeintensität feiner zu erfassen. Die Effekte werden teilweise getrennt nach Branche, Beruf, Region, Zeitpunkt und ggf. weiteren Variablen ausgewiesen. Diese Variablen können in einem Schätzansatz als Kontrollvariablen verwendet werden, um die Varianz der Schätzergebnisse so weit wie möglich zu vermindern. Darüber hinaus können personen- oder betriebsstättenbezogene Merkmale ein Profil typischer schattenwirtschaftlicher Beschäftigung ergeben und zur zusätzlichen Differenzierung der Ergebnisse beitragen.

1.2.4 Bisher vorliegende Studien

Bisher liegen zur Frage der präventiven Wirkungen von Bekämpfungsmaßnahmen gegen Schwarzarbeit kaum empirische Studien vor. In ihrer Befragung von Arbeitsmarktbeteiligten haben Pedersen (2003) sowie Feld und Larsen (2005) versucht, den präventiven Effekt der Bekämpfungsmaßnahmen darzustellen. Dabei wird das subjektiv wahrgenommene Risiko der Aufdeckung von Schwarzarbeit in vier Kategorien (hoch, eher hoch, eher gering, gering) erfasst. Diese Variable wird dann in einen Regressionsansatz einbezogen, der erklärt, ob der oder die Befragte in den zwölf Monaten vor dem Interview Schwarzarbeit geleistet hatte. Die Studien zeigen, dass die Neigung, Schwarzarbeit zu leisten, tendenziell umso geringer ist, je größer das subjektiv wahrgenommene Risiko ist. Jedoch ist das wahrgenommene Risiko nur bei den Frauen in den Jahren 2001 und 2004 signifikant, während es bei den Männern im Jahr 2001, nicht aber im Jahr 2004 mit einem statistisch signifikanten Koeffizienten in die Schätzgleichung eingeht. Allerdings ist zu bemerken, dass außer der Altersstruktur auch keine andere Variable, die zur Erklärung der Schwarzarbeit benutzt wird, in beiden Jahren einen signifikanten Koeffizienten aufweist. Dieser Mangel an Stabilität zwischen den Schätzungen für die beiden Zeitpunkte deutet auf Probleme bei der Stichprobenziehung hin. Entsprechend zurückhaltend sollten die Ergebnisse interpretiert werden. Zwar ist es grundsätzlich möglich, dass zusätzliche Abschreckung die Schwarzarbeit nicht vermindert (Feld und Larsen 2005: 91), die Studie ergibt aber weder für noch gegen diese Aussage klare Evidenz.

Für das Jahr 2004 wird außerdem in einem Teil der Schätzungen eine Variable einbezogen, die die wahrgenommene Strafhöhe erfasst. Die Befragten sollten angeben, welche Sanktion ihnen bei der Aufdeckung drohen: nur die Nachzahlung von Steuern, Nachzahlung und eine Geldbuße, Nachzahlung und Gefängnisstrafe oder Nachzahlung, Geldbuße und Gefängnisstrafe. Die Varianz im Antwortverhalten ist relativ gering: über die Hälfte der Befragten gab zutreffend an, dass eine Nachzahlung und eine Geldbuße drohen, etwa ein Viertel verweigerte die Antwort. Dem entspricht die mangelnde Signifikanz der Koeffizienten. Die statistische Stärke dieses Tests auf das Vorhandensein einer Präventionswirkung von Strafen ist daher gering. Eine weitere Einschränkung der Studie besteht darin, dass nur private Haushalte befragt werden, nicht aber Unternehmen, bei denen man eher von einem rationalen Kalkül ausgehen kann.

Die Wirkung zusätzlicher Ermittlungsmaßnahmen oder härterer Strafen auf Steuerhinterziehung oder illegale Aktivität im Allgemeinen ist Gegenstand einer großen Vielzahl von Studien, die sich meistens auf die Vereinigten Staaten beziehen und die hier nicht einzeln aufgeführt werden können. Die meisten der dabei entstehenden methodischen Probleme würden auch auf die Evaluation der Maßnahmen der FKS zutreffen. Diese Probleme und Möglichkeiten ihrer Lösung sind im methodischen Anhang zu diesem Bericht enthalten.

Gegenüber der Verwendung von Umfragedaten hat eine Evaluation auf der Basis von Daten, die auf dem Ermittlungsprozess beruhen, den Vorzug, dass sie nicht auf geäußerten Meinungen, sondern auf tatsächlichem Verhalten beruht.

Zur Frage des Rechtsbewusstseins liegen eine Reihe von Studien über die sogenannte Steuermoral vor. Frey und Feld (2002) finden in ihrer Betrachtung auf der Basis aggregierter Daten aus Schweizer Kantonen Evidenz für eine Verdrängung intrinsischer Motivation durch ein härteres Vorgehen gegenüber Steuerhinterziehung. Auf der Basis von Daten des World Values Survey untersuchen Torgler (2004a, 2005) sowie Torgler und Schneider (2005) die Determinanten der Steuermoral in einer Reihe von Ländern. Dabei kann jedoch der Einfluss von Strafen oder Ermittlungsaktivität auf die Steuermoral nicht festgestellt werden. Dieselbe Datenquelle liegt einer Studie des IAW im Auftrag des Bundesministeriums der Finanzen zugrunde (Körner und Strotmann, 2004). Auch hier wird jedoch die Wirkung von Bekämpfungsmaßnahmen auf die Steuermoral nicht thematisiert.

Bisherige Studien über das Ausmaß und die Entwicklung der Schwarzarbeit (vgl. für einen Überblick Schneider und Enste, 2000, sowie Feld und Larsen, 2005) verlassen sich weitgehend auf indirekte Verfahren, deren implizite Annahmen kaum empirisch verifiziert werden können. Slemrod und Yitzhaki (2002) beschreiben demgegenüber das Taxpayer Compliance Measurement Program (TCMP) des Internal Revenue Service in den USA als die verlässlichste Informationsquelle. Das TCMP ist ein Programm, in dem, ähnlich wie hier vorgeschlagen, repräsentative Daten aus Prüfungen gewonnen werden. Dieses Vorgehen hat den Vorteil, dass es auf objektiven Daten beruht, die zudem in größtmöglicher Disaggregation vorliegen (zur Unterscheidung unterschiedlicher Datengrundlagen siehe auch Eide, 1999). In Deutschland ist diese Vorgehensweise nach Kenntnisstand der Verfasser bisher nicht gewählt worden.

2 Ausgangsbedingungen der Evaluation

2.1 Struktur, Instrumente und Einsatzgebiete der FKS

2.1.1 Beschreibung der Struktur der FKS

Die Finanzkontrolle Schwarzarbeit (FKS) ist der Zollabteilung des BMF untergeordnet. Die Abteilung FKS der Oberfinanzdirektion (OFD) Köln (im Folgenden stets: Abteilung FKS) ist personell und organisatorisch der OFD Köln zugehörig. Teil der Abteilung FKS sind die Außenstellen bei den acht OFD. Darunter liegt die Ebene der Zollbezirke mit 40 Hauptzollämtern (HZÄ), in denen die FKS jeweils ein Sachgebiet (Sachgebiet E) darstellt. Die Abteilung FKS hat die fachliche Weisungsbefugnis gegenüber den Sachgebieten E in den HZÄ. Die Steuerung organisatorischer und personeller Belange auf der Ebene der HZÄ obliegt dem BMF und den übrigen OFD, die personelle und organisatorische Weisungsbefugnis liegt bei dem jeweiligen HZA-Vorsteher bzw. der jeweiligen OFD. Unterhalb der HZÄ liegt die Ebene der 113 einzelnen Standorte, die Teil der HZÄ sind.

Die FKS bei der OFD Köln ist in fünf Bereiche gegliedert: F1 Personal/ Service/ Organisation, F2 Prävention, F3 Prüfung und Ermittlung, F4 Ahndung, F5 Querschnittsbereich Betreuung/ Dienstaufsicht/ Außenstellen. An den Standorten findet sich ebenfalls die Gliederung in die Arbeitsbereiche Prävention, Prüfungen und Ermittlungen und Ahndung. Die Sollstärke an den Standorten hängt von verschiedenen Indikatoren (z.B. Zahl der Betriebe, Zahl der sozialversicherungspflichtig Beschäftigten, etc.) ab. Festgelegt ist eine Mindestgröße von 36 Mitarbeitern je Standort. Von den Beschäftigten sollen laut Dienstvorschrift der FKS 20% im Bereich der Prävention, 65% im Bereich der Prüfungen und Ermittlungen sowie 15% im Bereich der Ahnung eingesetzt werden. Von Standort zu Standort variiert die Ist-Besetzung derzeit stark, soll aber generell bis Ende 2005 bis zur Sollstärke anwachsen.

Für die Durchführung der Evaluation ist die Verteilung der Kompetenzen zwischen der Abteilung FKS und den Standorten in der Steuerung der Maßnahmen wesentlich, weil die Evaluation teilweise durch zentrale Vorgaben gesteuert werden müsste. In den durchgeführten Interviews wurde durchgängig betont, dass der Einsatz der Instrumente stark dezentral auf der Ebene der Standorte vorgenommen und sehr oft durch die Erfahrung der einzelnen Beamten bestimmt wird. Dies gilt jedoch nicht für alle Instrumente in gleicher Weise. So unterliegen beispielsweise Schwerpunktprüfungen auch einer Reihe von zentralen Vorgaben, auf die im folgenden Abschnitt genauer eingegangen wird.

Die Abteilung FKS koordiniert die Einsätze der Standorte, wenn sich z.B. bei Ermittlungen herausstellt, dass die Aktivität sich über einen größeren geographischen Bereich verteilt. Sie ist auch für die Zusammenarbeit mit nationalen und internationalen Behörden und Institutionen zuständig. Sie koordiniert die Tätigkeit der Standorte ferner durch Standards und Leitlinien. Durch die bundesweiten Schwerpunktprüfungen und andere Aktionen gehen von ihr Vorgaben für Präventions- oder Prüfungstätigkeiten aus, die dann von den Standorten durchgeführt werden müssen. In manchen Fällen steuert die Abteilung FKS auch im Einzelfall die Tätigkeit der Dienststellen, z.B. wenn Hinweise zentral bei der Abteilung FKS eintref-

fen. Darüber hinaus werden im Bereich der Ahndung von der Abteilung FKS teilweise Rechtsauslegungen für Ordnungswidrigkeitsverfahren vorgegeben.

2.1.2 Beschreibung der Instrumente der FKS

In diesem Abschnitt wird auf die Instrumente der FKS eingegangen. Da einige Instrumente von mehreren Arbeitsbereichen benutzt werden, folgt diese Darstellung nicht in allen Fällen der Gliederung der drei Arbeitsbereiche der FKS, vielmehr wird die Darstellung nach konkreten Tätigkeiten gegliedert.

Maßnahmen mit ausschließlich oder vorwiegend präventiver Zielsetzung

- Streifenwagenpräsenz und flächendeckende Präsenz der Mitarbeiter. Ziel ist dabei, das Unrechtsbewusstsein in der Bevölkerung zu erhöhen und die gesellschaftliche Akzeptanz von Schwarzarbeit zu senken. Der Umfang, in dem gezielt flächendeckende Präsenz hergestellt wird, variiert. Es ergibt sich ein enger Zusammenhang mit der Prüfungstätigkeit: Die Wirkung der Präsenz wird stark erhöht, wenn zugleich Prüfungen vorgenommen werden. Umgekehrt haben vermutlich alle Arten von Prüfungen auch eine präventive Wirkung. Näheres zu dieser Maßnahme wird daher im nächsten Unterabschnitt dargestellt.
- Pressearbeit. Die Abteilung FKS hat ein eigenes Pressereferat, das integraler Bestandteil der Verwaltungsarbeit der FKS, allerdings personell nur schwach besetzt ist und daher nur begrenzt zu aktiver Pressearbeit eingesetzt werden kann. Schwerpunkt der Pressearbeit ist daher momentan die Reaktion auf Anfragen der Presse und die Organisation der Anwesenheit von Pressevertretern bei Einsätzen (Pressebegleitung). Vereinzelt arbeitet das Pressereferat auch aktiv. Teilweise werden Schwerpunktprüfungen oder andere verstärkte Prüfungen vorher in der Presse angekündigt, außerdem werden Pressemitteilungen herausgegeben. Neben der Zuständigkeit der Abteilung FKS ergibt sich eine grundsätzliche Zuständigkeit des BMF. Für die regionale Pressearbeit ist daneben eine Zuständigkeit der HZÄ bzw. der OFDen gegeben, dabei werden Pressemitteilungen und andere Aktivitäten mit der Abteilung FKS abgesprochen.
- Zur weiteren Öffentlichkeitsarbeit gehören die folgenden Instrumente: Produktion und Verteilen von Aufklärungsbroschüren sowie Messeauftritte (Bau, Heimwerker, Handwerk). Dieses Instrument wird jedoch noch nicht besonders stark eingesetzt.

Prüfungen

Prüfungen sind hinsichtlich der Aufgabenstellung der FKS und des Volumens des Arbeitseinsatzes die wichtigste Aufgabe der FKS. Die Prüfungsaufgaben im Allgemeinen ergeben sich aus § 2 des SchwarzArbG. Die FKS setzt grundsätzlich stark auf die Erfahrung der beteiligten Beamten. Darüber hinaus ist der Leitfaden „Allgemeine Prüf- und Ermittlungsansätze“ der Abteilung FKS zu beachten. Die Tätigkeiten umfassen die folgenden Vorgehens-

weisen, die sich gegenseitig ergänzen und oft von denselben Beamten im Rahmen ihrer Außendiensttätigkeit geleistet werden:

- Spontane Kontrollen vor Ort. Diese Prüfungen werden vorwiegend durch das Arbeitsgebiet Prävention und im enger Verzahnung mit der Aufgabe „flächendeckende Präsenz“ durchgeführt. Die Beamten arbeiten bei dieser Aufgabe stark gemäß eigener Risikoanalyse. Es gibt eine Vorgabe über den Umfang dieser Tätigkeit. Danach sind 50% der Tätigkeit im Rahmen der Prävention im Außendienst abzuleisten.
- Hinweisbezogene Prüfungen, also Prüfungen auf Anzeigen oder Hinweise aus der Bevölkerung oder von anderen Behörden oder sonstigen Stellen. Dies bindet einen großen Teil der Aktivitäten der FKS; an einem Standort wurde die Einschätzung geäußert, dass 50% aller durch den Arbeitsbereich Prävention durchgeführten Prüfungen auf Hinweise zurückgeht, an einem anderen waren es sogar praktisch alle durchgeführten Prüfungen, während spontane Prüfungen nur in geringem Umfang vorgenommen wurden. Daneben wird Hinweisen auch ihrer (fiskalischen) Bedeutung entsprechend nachgegangen.
- Bundesweite Schwerpunktprüfungen. Etwa alle 3-4 Monate findet eine Schwerpunktprüfung statt. Dabei werden durch die Abteilung FKS Branche und Zeitraum festgelegt. Zentral vorgegeben ist ferner die Zahl der Prüfungen, ggf. ein Prüfungsschema und/oder ein Fragebogen. Es werden zentral bestimmte Rahmenbedingungen beschrieben, innerhalb derer die Prüfungen stattfinden (z.B. räumliche Festlegungen). Nicht vorgegeben werden die Prüfobjekte, vielmehr suchen sich die Dienststellen bzw. die im Einsatz befindlichen Beamten selbst die Betriebe und Arbeitstätten aus, bei denen sie Prüfungen vornehmen. Insofern dominiert auch hier die Risikoanalyse der am Einsatz Beteiligten. Vorgaben an dieser Stelle sind möglich, dürften aber bei den beteiligten Beamten auf Akzeptanzprobleme stoßen, zumal an den besuchten Standorten die Einschätzung herrscht, dass Schwerpunktprüfungen nicht zu den effektivsten Instrumenten der FKS gehören. Zu strikte Vorgaben hinsichtlich der Auswahl der Prüfobjekte wären auch insofern unzweckmäßig, als z.B. viele Betriebe wiederholt aufzusuchen wären.
- Risikoorientierte Prüfungen an der Arbeitsstelle. Solchen an den Standorten durch die Teamleiter aufgrund eigener Risikoeinschätzung vorbereiteten Prüfungen geht zumeist eine Phase der Informationsgewinnung im Innendienst voraus, die zum Teil im Zuge einer Datenprüfung (siehe unten) vorgenommen werden kann.
- Prüfung von Werkverträgen. Hierbei geht es um Werkverträge zwischen inländischen Auftraggebern und ausländischen Auftragnehmern auf der Basis bilateraler Regierungsvereinbarungen zwischen der Bundesrepublik Deutschland und mittel- und osteuropäischen Staaten sowie der Türkei, bei denen der Auftragnehmer mit seinen Arbeitnehmern in Deutschland eine Leistung erstellt. Diese Werkverträge unterliegen bei der BA einem behördeninternen Zustimmungsverfahren (siehe Bundesagentur für Arbeit, 2005). Die Werkverträge liegen inklusive der Angaben, welche Arbeitnehmer

an welchen Arbeitsstätten welche Tätigkeiten verrichten, der FKS durch den Zugriff auf Datenbanken der BA vor. Es ist vorgesehen, 100% der Werkverträge vor Ort zu überprüfen, was aufgrund der vorhandenen Kapazitäten meist nicht erreicht wird; hier gibt es erhebliche Unterschiede nach Standorten. An einem Standort wurde geäußert, dass weniger als 10% aller Werkverträge einer intensiven Prüfung unterzogen werden. An anderen Standorten werden die 100% dagegen vermutlich erreicht. Auch variiert die Intensität der Prüfungen. Durch die FKS wird geprüft, ob die Arbeitnehmer tatsächlich die Tätigkeiten verrichten, die dem Werkvertrag entsprechen, oder ob sie für andere Tätigkeiten eingesetzt werden und ob im Baubereich die Mindestlöhne gezahlt werden. Abgesehen von den Werkverträgen auf der Basis bilateraler Regierungsvereinbarungen besteht bei Werkverträgen im Baubereich nach dem Arbeitnehmer-Entsendegesetz für Arbeitgeber mit Sitz im Ausland die Verpflichtung, den Beginn und die voraussichtliche Dauer der Beschäftigung bei der Abteilung FKS zu melden. Alle übrigen Werkverträge unterliegen keiner „Genehmigungspflicht“.

- Buchprüfungen. Diese können unterschiedliche Formen annehmen, z.B. einer Geschäftsunterlagenprüfung persönlich in den Geschäftsräumen des Arbeitgebers. Die Befugnisse sind hierbei ähnlich wie die Befugnisse der Finanzverwaltung im Bereich der Betriebsprüfung.
- Datenprüfungen im DALEB-Verfahren. Dies ist ein Datenabgleich zwischen Zeiten des Bezugs von Lohnersatzleistungen und gemeldeten Beschäftigungszeiten, der durch die Bundesagentur für Arbeit vorgenommen wird. Seit Hartz IV betrifft dies jedoch lediglich das ALG I, während Leistungen im Rahmen von ALG II derzeit nicht durch das DALEB-Verfahren erfasst werden. Das DALEB-Verfahren wird jedoch ab 2006 auch das ALG II beinhalten. Fälle, in denen substantielle Überschneidungen bestehen, werden von den Mitarbeitern der Agenturen für Arbeit nach Aktenlage geprüft und im Verdachtsfall ohne weitere Prüfung eines Tatvorwurfs in straf- oder bußgeldrechtlicher Hinsicht der FKS zugeleitet. Diese führt die weiteren Ermittlungen und leitet ggf. ein Ordnungswidrigkeitenverfahren bzw. bei Straftatverdacht ein Strafverfahren ein. Eingeleitete Strafverfahren werden an die Staatsanwaltschaft abgegeben. Das DALEB-Verfahren verursacht bei den Arbeitsagenturen erkennbar einen hohen Arbeitsaufwand. Seitens der FKS gibt es keine Möglichkeit, die Vollständigkeit der Zuleitungen im Rahmen des DALEB-Verfahrens zu überprüfen. Das Einhalten der Meldepflichten durch die Arbeitgeber bildet die Grundlage dafür, dass das DALEB-Verfahren funktioniert. „Schwere“ Fälle, in denen Arbeitnehmer und Arbeitgeber gemeinsam Sozialversicherungsbetrug betreiben, können nicht aufgedeckt werden. Organisatorisch ist das Verfahren im Bereich der Ahndung angesiedelt.
- Andere Datenprüfungen benutzen die Beschäftigtendatei des Verbands Deutscher Rentenversicherungsträger (VDR), die auf den Meldungen der Arbeitgeber zur Sozialversicherung beruht. Die VDR-Daten können durch die FKS im Online-Zugriff benutzt werden und erlauben die Verknüpfung von Personen und beschäftigenden Unternehmen. Benutzt werden ferner die Ausländerzentralregister, daneben werden Melderegister bei den Ordnungsämtern und Teile der Gewerberegister fallweise an-

gefordert. Hier liegen jedoch keine zentralen Datenbestände vor, auf die ein Zugriff möglich wäre.

Prüfungen an der Arbeitsstelle werden grundsätzlich unangekündigt vorgenommen; lediglich bei Buchprüfungen kann es zweckmäßig sein, vorher einen Termin zu vereinbaren. Praktisch alle Prüfungen werden aufgrund eigener Risikoanalysen oder auf der Grundlage von bereits vorhandenen Informationen vorgenommen. Das Ermessen der beteiligten Beamten bei der Kontrolle erstreckt sich nicht allein auf die Auswahl der Prüfobjekte, sondern auch auf die Intensität der Prüfung. Dabei gibt es unterschiedliche Varianten, ob z.B. auf einer Baustelle alle dort angetroffenen Personen oder nur eine Stichprobe geprüft werden. Die Anwendung dieser Varianten ist aus den vorhandenen Daten oft nicht klar zu ersehen.

Die Standorte stoßen bei den Prüfungen häufig auf Restriktionen. Insbesondere durch den hohen Aufwand der Ermittlungen sowie die Vielzahl der Hinweise können nicht alle Prüfungen, die sinnvoll wären, tatsächlich durchgeführt werden. Hierbei liegen erhebliche Unterschiede zwischen den Standorten vor. Eine „Steuerung über das Budget“ findet nicht statt, d.h. den Dienststellen, bei denen die Intensität der Schattenwirtschaft groß ist, stehen über die in Abschnitt 3.1 erwähnten Kriterien für die Festlegung von Sollstärken hinaus nicht entsprechend höhere Mittel zur Verfügung. An einigen Standorten sind die Dienststellen daher mit der Erledigung von Pflichtaufgaben und Hinweisen so belastet, dass die sonstigen Instrumente kaum zur Anwendung kommen.

Ermittlungsmaßnahmen

Ermittlungen unterscheiden sich von Prüfungen durch das Vorliegen eines Tatverdachts. Die Angehörigen der FKS sind im strafrechtlichen Ermittlungsverfahren Ermittlungspersonen der Staatsanwaltschaft und handeln auf Weisung der Staatsanwaltschaft. Daneben sind die Kompetenzen durch die Strafprozessordnung, das StGB und andere Gesetze geregelt. In bußgeldrechtlichen Ermittlungsverfahren sind sie als für die Verfolgung und Ahndung zuständige Verwaltungsbehörde hinsichtlich der Ermittlungsmaßnahmen der Staatsanwaltschaft gleichgestellt. Im Bereich der Ermittlung ist der Handlungsspielraum erheblich geringer als im Bereich der Prüfung, insbesondere ist in strafrechtlichen Ermittlungsverfahren kein eigener Ermessensspielraum gegeben. Die Instrumente wie z.B. Vernehmungen sind durch die Strafprozessordnung vorgegeben. Es besteht typischerweise eine Verpflichtung zum Einsatz und keine Ermessensfreiheit. Ermittlungsverfahren sind sehr aufwendig, sie ziehen sich häufig über einen Zeitraum von mehreren Monaten hin. An einem der besuchten Standorte nimmt die Ermittlung den größten Teil des Aufwands (ca. 75%) im Arbeitsbereich Prüfungen und Ermittlungen ein. Die Prüfung wird daher weitgehend im Bereich Prävention durchgeführt. An einem anderen Standort teilt sich der Aufwand des Arbeitsbereichs hälftig zwischen Prüfung und Ermittlung.

Ahndungsmaßnahmen

Die Ahndung von Ordnungswidrigkeiten obliegt den jeweils zuständigen Verwaltungsbehörden, im hier interessierenden Bereich also der FKS. Bußgelder werden direkt durch die

Standorte der FKS verhängt. Im Einspruchsverfahren liegt die Zuständigkeit, soweit dem Einspruch nicht abgeholfen wird, bei den Amtsgerichten. Bei Straftatbeständen führt die Staatsanwaltschaft die Verfahren. Die Ahndung wird selbst nicht initiativ tätig. Verfahren, die aus den Prüfungen kommen, sind entweder im bußgeldrechtlichen oder strafrechtlichen Ermittlungsverfahren zu führen; entsprechend erfolgt eine Ahndung durch die FKS oder die weitere Ermittlung im strafrechtlichen Verfahren.

Es gibt einen begrenzten Ermessensspielraum bei der Ahndung im Rahmen der gesetzlichen Vorschriften. So kann die Behörde entscheiden, ob sie eine Geldbuße oder eine Verfallsanordnung nach §29a OWiG verhängt oder ob ein Verfahren eingestellt und eine Vermögensabschöpfung vorgenommen wird. Für Firmen ist die Entscheidung zwischen den ersten beiden Möglichkeiten vor allem deshalb relevant, weil im ersten Fall ein Eintrag ins Gewerbezentralregister vorgenommen wird, woraus bei öffentlichen Ausschreibungen Nachteile entstehen. Vor allem deshalb geht von Bußgeldern eine erhebliche Abschreckungswirkung aus. Für die Höhe der Bußgelder gibt es einen Richtwerte-Katalog, innerhalb dessen aber gewisse Spielräume bei der Festlegung der Bußgelder bestehen. Dabei haben die Arbeitsgebietsleiter die Aufgabe, für eine einheitliche Handhabung zu sorgen. An einem Standort wurde betont, dass bei der Ahndung sehr auf den Einzelfall geachtet wird, wie er sich durch die Prüfungen darstellt: Handelt es sich um einen schweren Fall? Was sind die sonstigen Umstände der Tat? Ein Maß für die Härte der Bußgelder könnte die Zahl der Einsprüche sein, die ebenfalls im Bereich Ahndung bearbeitet werden.

Auch die Spruchpraxis der Gerichte kann regional bzw. von Gericht zu Gericht variieren. Dabei ist immer das Amtsgericht am Sitz des HZA zuständig, obwohl es meist mehrere Amtsgerichte in einem Zollamtsbezirk gibt.

2.1.3 Beschreibung der wichtigsten Einsatzgebiete

Die Branchen, in denen die FKS am stärksten tätig ist, sind die Bauwirtschaft, das Transport- und das Hotel- und Gaststättengewerbe. Daneben wurden das Reinigungsgewerbe, Gartenbau und Landschaftspflege, Schlachthöfe und allgemein bestimmte Handwerksbereiche (z.B. Friseurhandwerk und KFZ-Reparaturgewerbe) genannt. Gegenüber privaten Haushalten gibt es relativ wenig Prüfungstätigkeit, hier wird die FKS nur auf Hinweise tätig. Das hat nicht zuletzt damit zu tun, dass die FKS hier nicht die gleichen Rechte (z.B. Zutrittsrechte) hat wie im gewerblichen Bereich: Ein verdachtloses Betretensrecht für Prüfungen nach § 2 Schwarzarbeitsbekämpfungsgesetz besteht nur für Geschäftsräume, nicht für Privatwohnungen. Nur soweit Anhaltspunkte für Ordnungswidrigkeiten oder Straftaten bestehen, die in Privatwohnungen begangen werden, sind entsprechende Ermittlungen möglich.

Das Volumen der Schwarzarbeit unterscheidet sich regional, wobei vorwiegend die regionale Wirtschaftskraft ausschlaggebend ist. In Ballungsräumen gibt es entsprechend mehr Schwarzarbeit. An einem Standort wurde die Ansicht geäußert, dass das Umland bei den Einsätzen der FKS weniger berücksichtigt werde, da hier kaum Aussichten bestünden, Schwarzarbeit aufzudecken und ein vergleichsweise hoher Aufwand z.B. durch lange Fahrzeiten entstehe.

2.2 Zur Verfügung stehende Daten

Im Folgenden werden zunächst die Datenbestände der FKS dargestellt. Im Weiteren wird dann auf Daten anderer Institutionen eingegangen, die ebenfalls für eine Evaluation genutzt werden könnten. Grundsätzlich werden diese Daten hier unabhängig von der Frage diskutiert, ob sie aus rechtlichen, ermittlungstechnischen oder sonstigen Gründen einem Evaluator zur Verfügung gestellt werden könnten. Diese Fragen müssten im Falle einer Evaluation mit dem Auftraggeber geklärt werden.

2.2.1 Arbeitsstatistiken der FKS

Bei den Arbeitsstatistiken handelt es sich um monatliche Übersichten, in denen die Tätigkeit der FKS vermerkt ist. Sie enthalten unter anderem Angaben über die Zahl der durchgeführten Personenüberprüfungen, Prüfung von Arbeitgebern im Außendienst (für die Baubranche getrennt von den übrigen Branchen aufgeführt) und Werkvertragsprüfungen, ferner die Zahl der eingegangenen Hinweise, der Personalbestand, die bei Einsätzen zurückgelegten Kilometer, die eingeleiteten Ordnungswidrigkeitsverfahren, die Zahl und Höhe der ausgesprochenen Bußgelder und Strafen sowie die Zahl der Einsprüche gegen Bußgelder. Die Aggregation dieser Information auf Standortebene wird am Standort selbst durch die Teamleiter oder durch besonders damit betraute Mitarbeiter durchgeführt. Es erfolgt am Standort eine Plausibilitätsprüfung, bevor die Daten an die Abteilung FKS weitergeleitet werden. Die Daten der Arbeitsstatistik liegen seit 2002 vor. Mit Realisierung des Moduls „Controlling, Statistik, KLR“ in ProFiS (siehe Abschnitt 2.2.5) wird die manuelle Arbeitsstatistik nicht mehr geführt, die Daten werden dann aus ProFiS gewonnen.

Vorteile dieser Datenquelle sind die genaue zeitliche Abgrenzung und zeitnahe Erhältlichkeit der Angaben.

Ein Nachteil ist die aggregierte Erhebungsweise der Arbeitsstatistik. Um nach Branchen oder gar nach einzelnen Einsätzen gegliederte Angaben zu erhalten, müssten Auswertungen der Akten vorgenommen werden, die jedoch nicht elektronisch geführt werden (siehe Abschnitt 2.2.6). Wirkungen der Tätigkeit der FKS können daher, wenn überhaupt, nur für die Gesamtheit der Aktivitäten am jeweiligen FKS-Standort bestimmt werden. Es kann auch nicht nach verdachtlosen oder hinweisbezogenen Prüfungen getrennt werden. Da nur verdachtlose Prüfungen dazu geeignet sind, näherungsweise repräsentative Angaben über das Volumen der Schwarzarbeit zu erhalten, sind die Arbeitsstatistiken für diesen Zweck nicht verwendbar.

Insgesamt dürften sich die Arbeitsstatistiken daher nur für die Erfassung der Intensität der FKS-Tätigkeit, also die zu überprüfenden Maßnahmen, nicht jedoch für die Messung der verhaltenslenkenden Wirkungen eignen. Allerdings können auch die Maßnahmen nur verhältnismäßig grob dargestellt werden. Schließlich besteht das Problem, dass einige Vorgaben für die Arbeitsstatistik von Jahr zu Jahr wechseln, was die zeitliche Vergleichbarkeit beeinträchtigt.

2.2.2 Kosten- und Leistungsrechnung

Die Kosten- und Leistungsrechnung erfasst bis auf die Ebene des einzelnen Beamten herunter den zeitlichen Einsatz der Beschäftigten der FKS. Die Erfassung ist dabei in die folgenden Kategorien gegliedert: Prävention Außendienst, Prävention Innendienst, Prüfungen, Ermittlungen, Ahndung, Einsatzplanung, Ausbildungszeiten, Leitungsaufgaben und einer Reihe sonstiger Tätigkeiten. Die Kosten- und Leistungsrechnung bietet somit eine noch erheblich bessere Erfassung des Inputs in die Aktivitäten der FKS als die Arbeitsstatistik.

Ein Nachteil ist auch hier, dass nicht nach Einsätzen in bestimmten Branchen gegliedert werden kann. Ein möglicher Nachteil ist die Datenqualität, die immer dann problematisch ist, wenn die Erhebung von den Mitarbeitern als zusätzliche Belastung empfunden wird.

2.2.3 BillBAO

BillBAO (kurz für: Bekämpfung illegaler Beschäftigung unter Einsatz eines Auskunfts- und Offenbarungssystem) ist eine Datenbank für illegale Beschäftigung und Tatvorwürfe. Es handelt sich hierbei um eine bei der Abteilung FKS zentral geführte, individuell erstellte Datenbank, die aus einer Meldungs- und Verfahrensdatenbank besteht. Die erhobenen Merkmale beziehen sich auf den individuellen Fall: meldende Stelle, Tatort, Arbeitgeberadresse, gesetzlicher Vertreter des Arbeitgebers, Hinweis zum Arbeitgeber, Hinweis zum Verfahren, Tatverdächtiger, Hinweis zum Tatverdächtigen, Straftaten und Ordnungswidrigkeiten, Tatumfang, Abgabe des Verfahrens, Abschluss des Verfahrens. Die Daten aus BillBAO werden in Kürze in ProFiS integriert.

Der Vorteil dieser Datenbank ist, dass Daten über einzelne Personen und Ermittlungsverfahren zur Verfügung stehen.

Nachteile sind, dass keine statistischen Ausgaben möglich sind. Zudem ist nicht festgehalten, in welchem Kontext die Verstöße aufgedeckt worden sind (d.h. bei welchen Ermittlungen); dies wird allenfalls fallweise in Textfeldern vermerkt. Viele Eingaben sind als Volltexteingaben und nicht als Kategorien verzeichnet, so dass eine statistische Auswertung diese qualitative Information zunächst in quantitative Information umwandeln müsste.

2.2.4 Sondererhebungen im Rahmen von Schwerpunktprüfungen

Im Rahmen von Schwerpunktprüfungen werden eigene Statistiken geführt, deren Inhalt jeweils durch die Abteilung FKS vorgegeben wird. Enthalten sind meistens Personaleinsätze, Verdachtsfälle usw. Wie die anderen Datenbestände, sind auch die Rahmen von Schwerpunktprüfungen gewonnenen Informationen Teil der Datenbank ProFiS.

Wegen der Flexibilität bei der Frage, welche Merkmale erhoben werden, könnte eine Datenerhebung zum Zweck einer Evaluation in diesem Kontext vermutlich gut durchgeführt werden. Zudem könnte die Datenerhebung durch zentrale Vorgaben prinzipiell so gesteuert werden, dass systematische Verzerrungen aufgrund der eigenen Risikoeinschätzung der

beteiligten Beamten weitgehend vermieden werden und ein Datensatz entsteht, der einer repräsentativen Stichprobe des Untersuchungsgegenstands nahe kommt. Auf die besonderen Probleme dabei wird in Abschnitt 3.2.1 hingewiesen.

2.2.5 ProFiS

Für die Aufgabenwahrnehmung der FKS wurde das IT-Verfahren ProFiS erstellt, welches für den Einzelfall sämtliche Arbeitsschritte der Prävention, Prüfung und Ermittlung bis hin zu einer eventuellen Ahndung unterstützen soll. Zudem können auch nach Abschluss der Ermittlungsverfahren Entscheidungen anderer Behörden (u.a. Gericht, Staatsanwaltschaft) nachträglich erfasst werden. Die Datenerhebung erfolgt dezentral durch die Beschäftigten der FKS in unterschiedlichen Modulen. In jedem Abschnitt des Verfahrens, bei den Vorermittlungen (Tagebuch), bei den Prüfungen sowie bei den Ermittlungsverfahren können Daten erfasst werden. Zum gegenwärtigen Zeitpunkt ist das Auswertungsmodul „Controlling, Statistik, KLR“ noch nicht realisiert, es dürfte jedoch ab Anfang 2006 zur Verfügung stehen.

Auswertungen der einzelnen Module werden durch Filter- Such- und Sortierfunktionen ermöglicht, so dass gezielt Daten, einzeln oder in gebündelter Form, aufrufbar sein werden. Mit Hilfe dieser Funktionen kann z.B. bei dem Modul „Ermittlungsverfahren“ nach Branchen, Tatbeständen und/oder Erledigungsarten (Einstellungen, Verurteilungen, Buß- und Verfallsbescheide usw.) selektiert werden. ProFiS erlaubt also die Verknüpfungen zwischen Ermittlungsergebnissen, Instrumenten und Merkmalen auf der Individualebene. Gegebenfalls sind die Anforderungen des Datenschutzes zu beachten, die Leistung des einzelnen Beamten (Einsätze und ihre Ergebnisse usw.) ist deswegen nicht aus ProFiS zu ersehen.

Die erhobenen Merkmale umfassen unter anderem:

- Schwerpunktprüfungen nach Wirtschaftszweigen bis hin zu einer Beanstandung und deren Erledigung, Streifenfahreneinsätze in Verbindung mit Spontanprüfungen bis hin zu einer Beanstandung und deren Erledigung, hinweisbezogene Prüfungen bis hin zu einer Beanstandung, verdachtsunabhängige Prüfungen bis hin zu einer Beanstandung, Personenüberprüfungen von Arbeitgebern/Arbeitnehmern
- Laufende bzw. abgeschlossene Ermittlungsverfahren sowohl im Strafverfahren als auch im Ordnungswidrigkeitenverfahren, Gerichtsentscheidungen (Höhe und Zusammensetzung der Bußgelder, Höhe und Zusammensetzung der Freiheitsstrafen, Höhe und Zusammensetzung der Geldstrafen), Höhe und Zusammensetzung der Bußgelder, Schadenssummen

ProFiS hat eine grundsätzlich andere Funktion als die Arbeitsstatistiken. ProFiS dient zur Erfassung und Bearbeitung von Vorgängen. Die Arbeitsstatistiken werden zwar auch in ProFiS übernommen, aber darüber hinaus dient ProFiS als Informationssystem bei der Wahrnehmung von Aufgaben durch die FKS (Vorgangsbearbeitung) und ist insoweit nicht in erster Linie eine Statistik, aber Grundlage für sämtliche künftigen statistischen Auswertungen.

Vorteile sind wie bei den Arbeitsstatistiken die zeitnahe Erhältlichkeit der Daten und – im Vergleich zu den Arbeitsstatistiken – das Vorliegen von Informationen auf der Ebene des einzelnen Individuums bzw. der einzelnen Überprüfung. Ein weiterer Vorteil dürfte in der hohen Datenqualität liegen. Zwar ist bei den prozessproduzierten Daten der FKS die Gefahr der Manipulation immer vorhanden (z.B. Vermerkung von Einsätzen, die nie stattgefunden haben). Jedoch werden durch die Abteilung FKS regelmäßige Prüfungen vorgenommen. Insbesondere dort, wo es um nachprüfbare Informationen über Einzelfälle geht, kann kaum manipuliert werden.

Ein weiterer Vorteil ist, dass nicht allein Verdachtsfälle festgehalten werden, sondern dass diese Fälle bis hin zu einer möglichen Ahndung dokumentiert sind. Dies bedeutet, dass die Wahrscheinlichkeit, dass die Tatvorwürfe zutreffen, erheblich höher ist, als wenn man sich auf eine Datenbank stützen würde, die allein die Verdachtsfälle verzeichnet. Überdies ist die Höhe von Strafen oder Bußgeldern festgehalten sowie die groben Umriss des Tatbestands. Damit wird nicht allein die Tatsache, sondern auch die Intensität des Verstoßes festgehalten.

Ein gewichtiger Nachteil der Datenbank als Informationsgrundlage für eine Evaluation ist, dass Datenbestände nur dort angelegt werden, wo Prüfungen durchgeführt und Verdachtsfälle angetroffen wurden. Damit ist die Datenbasis durch die Determinanten beeinflusst, die die Prüfungen auslösen, wie Anfangsverdachte und Risikoanalysen. Dies bedeutet, dass die Datenbasis statistischen Repräsentativitätskriterien in den meisten Fällen nicht genügen dürfte.

Am ehesten ist Repräsentativität bei verdachtslosen oder Spontanprüfungen (siehe Abschnitt 2.1.2) zu vermuten. Spontanprüfungen können in ProFiS im Allgemeinen nicht identifiziert werden. Zwischen Schwerpunktprüfungen, hinweisbezogenen Prüfungen und verdachtslosen Prüfungen kann hingegen unterschieden werden. Hier ist allerdings immer nur die im strafrechtlichen Sinne verdachtslose Prüfung gemeint, nicht die Abwesenheit einer Risikobewertung seitens der Angehörigen der FKS.

Wenn es bei Prüfungen keine Beanstandungen gibt, wird nur die Anzahl der Prüfungen festgehalten, aber die Prüfungen sind nicht nach Prüfobjekten aufgeschlüsselt abgelegt. Daher sind in diesen Fällen keine Informationen über Charakteristika der überprüften Personen/Unternehmen erfasst. Auch hier spielen datenschutzrechtliche Bestimmungen eine Rolle.

2.2.6 Nicht elektronisch erfasste Akten

Über jeden einzelnen Prüfungsfall werden Akten angelegt, die an den Standorten aufbewahrt werden. Diese könnten zum Zweck der Evaluation ausgewertet werden. Aktenauswertungen werden bereits jetzt gelegentlich im Auftrag der FKS bei der OFD Köln vorgenommen. Der Nachteil ist hier der erhebliche personelle Aufwand, der an den Standorten entsteht und die Akzeptanz von Evaluationsstudien unter den beteiligten Mitarbeitern der FKS gefährden könnte. Daher sollte auf diese Methode der Datengewinnung verzichtet werden.

2.2.7 Daten aus anderen Bezugsquellen

Daten aus dem DALEB-Verfahren

Das DALEB-Verfahren ist ein maschineller Datenabgleich zwischen Zeiten des Bezugs von Lohnersatzleistungen und gemeldeten Beschäftigungszeiten. Genaueres über das Prüfungsverfahren ist aus Abschnitt 2.1.2 zu entnehmen. Vorteile sind die – jedenfalls laut Vorgaben – vollständige Erfassung, die nicht einer Risikoabwägung unterliegt. Vermutlich wären zusätzliche personenbezogene Merkmale durch Verknüpfung mit den Daten der BA grundsätzlich zu erhalten (BEH, BewA, LeH, MTG; siehe unten). Die DALEB-Zuleitungen ergeben nur die Fälle einer Überschneidung von Beschäftigungszeiten mit Leistungsbezugszeiten der BA, sie geben keinen Aufschluss über Ermittlungsergebnisse. Zur Messung des Indikators Schwarzarbeit müssen diese Informationen daher mit Ermittlungsdaten (bspw. aus der Datenbank ProFiS) verknüpft werden.

Nachteile: Das Einhalten der Meldepflichten durch die Arbeitgeber bildet die Grundlage für das Funktionieren des DALEB-Verfahrens. Fälle, in denen Arbeitnehmer und Arbeitgeber gemeinsam Sozialversicherungsbetrug betrieben, können nicht aufgedeckt werden. Die Verfügbarkeit für Evaluationszwecke muss geklärt werden, ebenso die Vorgehensweise beim Abgleich der Daten. Die Nutzung setzt Zusammenarbeit mit der Bundesagentur für Arbeit voraus, die je nach Art der Datennutzung ebenfalls Daten übermitteln müsste (geprüfte Fälle). Schwer zu kontrollieren dürfte es sein, in welchem Grad die maschinell festgestellten Überschneidungszeiten in den örtlichen Arbeitsagenturen geprüft werden. Hier besteht die Gefahr von systematischer Untererfassung zu bestimmten Zeitpunkten und in bestimmten Agenturen. Die Implementationsanalyse (siehe Abschnitt 3.3) sollte hierüber Aufschluss geben. Auf der Basis von DALEB allein ist keine Aussage dazu möglich, ob Leistungen tatsächlich missbräuchlich in Anspruch genommen wurden. Die Daten müssten daher mit Hilfe von ProFiS mit entsprechenden Angaben aus dem Bereich der Ahndung verknüpft werden.

Daten der Bundesagentur für Arbeit über Beschäftigte und Leistungsempfänger

Datenquellen der BA wie Beschäftigtenhistorik (BeH), Bewerberangebotsdaten (BewA), Leistungsempfängerhistorik (LeH), oder Maßnahme-Teilnehmer-Grunddatei (MTG) (siehe Wilke und Winterhager 2004 für eine Beschreibung) könnten zur Gewinnung einer repräsentativen Datenbasis genutzt werden. Eine Verknüpfung zu den Angaben über den Einsatz präventiver und anderer Maßnahmen durch die FKS wäre über Merkmale Region, Branche, Beruf, Leistungsempfang usw. herzustellen. Im Idealfall wäre auch eine Verknüpfung auf der individuellen Ebene möglich, so dass z.B. ersichtlich würde, ob ein Leistungsempfänger eine bestimmte Informationsbroschüre erhalten hat. Für viele Zwecke und Indikatoren wäre diese Datenbasis vermutlich die einzige, die tatsächlich den Repräsentativitätskriterien genügt. Die Verfügbarkeit für Zwecke der Evaluation muss durch die BA sichergestellt werden. In jedem Fall ist bei der Nutzung dieser Daten eine enge Kooperation mit der Bundesagentur für Arbeit unabdingbar.

Daten über selbständig Erwerbstätige

Selbständig Erwerbstätige sind im Unterschied zu den sozialversicherungspflichtig Beschäftigten in den Daten der BA nicht verzeichnet. Eine Reihe von Umfragen anderer Institutionen beschäftigt sich mit Existenzgründern. Stellvertretend für mehrere sei hier auf die Daten des KfW-Gründungsmonitors verwiesen (KfW, 2005). Bei dieser handelt es sich um eine repräsentative Umfrage unter der deutschen Wohnbevölkerung, die Aufschluss über die Zahl und Struktur der neu gegründeten Unternehmen (auch Ein-Personen-Unternehmen) gibt.

Daten aus der amtlichen Statistik

Aggregierte Daten aus der amtlichen Statistik können über die Merkmale Region, Branche, Kalenderzeit und andere der Evaluation zugespielt werden. Hierbei wird es jedoch lediglich um Kontrollvariablen für die Evaluation gehen, nicht um Ergebnisvariablen.

Daten aus Umfragen, die nicht zur amtlichen Statistik gehören

Inwieweit aggregierte Informationen aus Daten aus Befragungen und wissenschaftlichen Erhebungen als ergänzende Angaben zugespielt werden können, ist jeweils im Einzelfall zu klären. Nachteil ist die üblicherweise relativ hohe zeitliche Verzögerung, mit der diese Daten wie z.B. das Sozio-ökonomische Panel (SOEP) verfügbar sind.

3 Bestandteile einer Evaluation der verhaltenslenkenden Wirkungen der FKS

In diesem Teil des Berichts werden die wesentlichen Bestandteile und Vorgehensweisen bei der Evaluation der Tätigkeit der FKS dargestellt:

- deskriptive Auswertungen des vorhandenen Datenbestands;
- Schätzung kausaler Effekte der Tätigkeit der FKS auf das Verhalten der Arbeitsmarktbeteiligten und das Rechtsbewusstsein;
- Implementationsanalyse;
- Kosten-Nutzen-Analyse;
- interne Kommunikationsstrategie.

Die Umsetzung dieser Punkte in konkreten Evaluationsvorhaben wird in Teil 4 des Berichts wieder aufgenommen. Den größten Teil der Darlegungen nehmen die quantitativen Analyse-schritte ein, da die Durchführbarkeit von qualitativen Schritten wie insbesondere der Implementationsanalyse weniger in Frage steht.

3.1 Deskriptive Analysen der verwendeten Indikatoren

In einem ersten Schritt sollten – unabhängig von den einzelnen, in Abschnitt 4 vorgestellten Evaluationsprojekten – die Indikatoren für die Schwarzarbeit deskriptiv aufbereitet werden. Die vorhandenen Datenbestände sind durchaus geeignet, die Verbreitung der Schwarzarbeit und Einsatzschwerpunkte bei der Bekämpfung von Schwarzarbeit zu beleuchten und sollten daher unabhängig von der Schätzung kausaler Effekte deskriptiv ausgewertet werden. Dies ist umso sinnvoller, als derartige Auswertungen bisher nicht im Detail durchgeführt wurden. Deskriptive Auswertungen des vorhandenen Datenmaterials würden zwar in den meisten Bereichen kein Bild der Schwarzarbeit ergeben, das den Kriterien statistischer Repräsentativität genügt (siehe Abschnitt 3.2). Jedoch geben sie einen empirischen Rahmen ab, innerhalb dessen die Plausibilität der in den Evaluationsprojekten erzielten Befunde und Ergebnisse bewertet werden kann. Aus Gründen der Begrenzung des Aufwandes sollte man sich dabei auf die Auswertung des Datenbanksystems ProFiS beschränken.

In Abschnitt 1.3 wurden Fallzahlenindikatoren, Volumenindikatoren und Wertindikatoren voneinander unterschieden. Die in ProFiS enthalten Indikatoren sind in der Regel Fallzahlenindikatoren, wie die Zahl der bei Prüfungen festgestellten Beanstandungen, die Zahl der eingeleiteten Ordnungswidrigkeits- oder Strafverfahren oder die Zahl der Verurteilungen. Ein für die Erfassung der Schwarzarbeit besonders wertvolle Eigenschaft von ProFiS ist, dass die jeweiligen Verfahren bis zu ihrem Ende nachverfolgt werden können, so dass nicht nur Fallzahlen von Verdachtsfällen, sondern auch rechtlich erhärtete Fälle von Schwarzarbeit zur Verfügung stehen. Damit stehen in der Datenbasis Angaben bereit, die einen hohen Grad der Objektivität aufweisen.

Volumenindikatoren für die Schwarzarbeit sind in ProFiS nicht enthalten. Als Wertindikatoren sind die errechneten Schadenssummen verfügbar. Während bei den einzelnen Evaluationsprojekten diese Schadenssummen in der Kosten-Nutzen-Analyse selbst gebildet werden, sollen in diesem Schritt die in ProFiS erfassten Schadenssummen direkt übernommen werden.

Die Indikatoren müssen auf eine Vergleichsgröße bezogen werden, um hinsichtlich der Verbreitung von Schwarzarbeit interpretierbar zu werden. Für Fallzahlenindikator ist die Zahl der durchgeführten Prüfungen ein geeigneter Indikator, der in ProFiS verfügbar ist. Für den Indikator Schadenssummen müsste der Indikator selbst gebildet werden, z.B. könnten für die geprüften Personen mit Hilfe von externer Information die innerhalb eines bestimmten Zeitraums gezahlten Arbeitsentgelte oder Sozialversicherungsbeiträge abgeschätzt werden. Die entsprechenden Schätzungen würden dann eine Vergleichsgröße darstellen.

Die deskriptiven Darstellungen sollten nicht nur global erfolgen, sondern lassen sich getrennt nach

- Branchen und Berufen
- Regionen und Zeitpunkten
- Begehungsarten bzw. Tatbeständen
- Art der Prüfungen (Schwerpunktprüfungen, verdachtlose Prüfungen, hinweisbezogene Prüfungen, Werkvertragprüfungen)

bzw. Kombinationen dieser Merkmale gliedern. Alle genannten Gliederungsmerkmale können aus ProFiS übernommen werden.

Besonders interessant sind dabei die Ergebnisse verdachtsloser Prüfungen, weil diese am ehesten repräsentativ für die Gesamtheit der Prüfobjekt sein dürften. Hier könnten Kennzahlen gebildet werden, die die Zahl der Verstöße relativ zur Zahl der Prüfobjekte wiedergeben. Auch die verdachtlosen Prüfungen hängen von der Risikoeinschätzung der Beamten der FKS ab. Dennoch ergibt sich ein Unterschied zu den hinweisbezogenen Prüfungen, bei denen konkrete Anhaltspunkte vorliegen. Wie groß dieser Unterschied ist, muss allerdings die deskriptive Analyse erst erweisen. Sie vermittelt insofern auch ein Bild über die Arbeitsweise der FKS.

Neben der Darstellung der Indikatoren für Schwarzarbeit sollte die deskriptive Analyse auch eine Auswertung des Einsatzes der FKS vornehmen. Die meisten Maßnahmeindikatoren sind hierbei wiederum Fallzahlenindikatoren. So sollte die Zahl der Prüfungen nach den oben dargestellten Gliederungsmerkmalen dargestellt werden. Auch hier sind relative Zahlen aussagekräftiger als absolute. Als Bezugsgröße kann man – je nach Gliederungsmerkmal – beispielsweise die Wertschöpfung der Branche oder der Region heranziehen. Neben den Fallzahlenindikatoren findet man im Bereich der Ahndung auch Volumenindikatoren (Summe der verhängten Freiheitsstrafen) und Wertindikatoren (Höhe der verhängten Bußgelder, Verfallsbescheide oder Geldstrafen).

Die Evaluation sollte diese deskriptiven Darstellungen unabhängig von den Kausalanalysen durchführen. Zugleich sind den einzelnen Kausalanalysen deskriptive Auswertungen der jeweils teilweise speziell erhobenen Datenbasis voranzustellen. Aufgrund der kontrollierten Art der Datenerhebung kommen diese Auswertungen einer repräsentativen Erhebung der Schwarzarbeit wesentlich näher. Zudem werden teilweise noch andere als die in diesem Abschnitt dargestellten Indikatoren erhoben. Dies betrifft insbesondere die Indikatoren legale Beschäftigung und Rechtsbewusstsein, aber auch Indikatoren, die aus ProFiS entnommen werden können wie z.B. Beanstandungen bei Personen und/oder Unternehmen, die in der Vergangenheit wegen Schwarzarbeit bereits belangt worden sind. Auf diese zusätzlichen deskriptiven Darstellungen wird jeweils bei der Darstellung der einzelnen Evaluationsprojekte in Abschnitt 4 hingewiesen.

3.2 Schätzung von Kausaleffekten

Für die Güte einer quantitativen Wirkungsanalyse sind zwei Faktoren ausschlaggebend: die Eignung der Datenbasis und die Art der Variation in der zu bewertenden Maßnahme, die als Grundlage der Evaluation verwendet wird. Diese beiden Probleme bilden den Schwerpunkt dieses Abschnitts.

3.2.1 Eignung der Datenbasis

Ergebnisvariable Schwarzarbeit

Bei der Erfassung von Schwarzarbeit stellt sich das grundlegende Problem, dass die interessierende Aktivität von den sie ausübenden Individuen verheimlicht wird. Der Auftrag des Projekts sieht als Lösung vor, Daten aus dem Ermittlungsprozess zu gewinnen, durch den die Aktivitäten aufgedeckt werden. Dies bewirkt jedoch eine Reihe eigener Probleme, auf die in diesem Abschnitt eingegangen wird.

Bei der Nutzung der FKS-Daten ist zu berücksichtigen, dass die Datenbasis zugleich mit den Prüfungsaktivitäten geschaffen wird und daher von diesen Maßnahmen abhängt. Dies führt zu einer nicht-repräsentativen Stichprobe, wenn die Prüfungen von einem Anfangsverdacht oder einer Risikoanalyse geleitet sind. Das Ausmaß, in dem Erfahrungswerte, Vorermittlungen und ähnliches in die Zahl der Prüfungen und in die Auswahl der Prüfobjekte eingehen, kann in der Evaluation praktisch nicht erfasst werden. Da sich die einzelnen Standorte dabei vermutlich unterscheiden, besteht die Gefahr, dass festgestellte Unterschiede in den Daten nicht auf tatsächliche Unterschiede in den Ergebnisindikatoren zurückgehen, sondern auf unterschiedlichen Vorgehensweisen bei der Prüfung. Die Evaluation muss sich daher auf Datenquellen stützen, die diese Probleme möglichst nicht aufweisen, weil ihr Zustandekommen in geringem Maße ermessensgesteuert ist.

Aus den Darlegungen aus Abschnitt 2.2 folgt, dass im Wesentlichen vier Datenquellen der FKS als (näherungsweise) repräsentative Datenquellen als Ausgangspunkte für die Messung von Schwarzarbeit verwendet werden können: (1) Schwerpunktprüfungen, (2) Daten aus

dem DALEB-Verfahren, (3) Daten über Werkverträge und (4) Ergebnisse verdachtsloser Prüfungen.

(1) Schwerpunktprüfungen bieten die besten Möglichkeiten, auf kontrolliertem Wege zu einer repräsentativen Datenbasis über Schwarzarbeit zu gelangen. Die Nutzung des Dateninformationssystems ProFiS bietet dabei die Möglichkeit, die in den Schwerpunktprüfungen festgestellten Verdachtsfälle bis zu einer eventuellen Ahndung weiter zu verfolgen. Die Repräsentativität der bei Schwerpunktprüfungen gewonnenen Daten hängt jedoch in hohem Maße von der Möglichkeit der FKS ab, durch Vorgaben zu steuern und den Einfluss des Ermessens der befassten Beamten auf das Zustandekommen der Datenbasis zu begrenzen. Dabei stellt sich das Problem, dass die Kooperationsbereitschaft der die Prüfung durchführenden Beamten durch zu strikten Vorgaben beeinträchtigt werden könnte, was die Datenqualität negativ beeinflussen dürfte.

Idealerweise müssten die Prüfobjekte so definiert werden, dass die Risikoanalyse der Mitarbeiter vor Ort bei der Auswahl der konkreten Objekte und bei der tatsächlichen Prüfintensität keine bedeutsame Rolle mehr spielt. Dies wäre beispielsweise durch die Vorgabe einer Liste von Prüfobjekten durch die Abteilung FKS zu erreichen. Die Realisierbarkeit dieses Vorgehens wird von der Abteilung FKS als problematisch beurteilt, wobei sich je nach Vorgehensweise deutliche Unterschiede ergeben. Relevante Gesichtspunkte sind dabei:

- Bei einer genauen Vorgabe der Prüfobjekte ist eine gute und intensive Kommunikation mit den Standorten erforderlich. Diese ist kaum mit allen 113 Standorten durchzuführen, denn bei einer so großen Zahl von Standorten wird man auf schriftliche Anweisungen angewiesen sein. Tendenziell gilt daher: Je kleiner die Zahl der an der Evaluation teilnehmenden Standorte, desto genauer können die Vorgaben sein und desto eher ist davon auszugehen, dass die Vorgaben in hinreichendem Maße umgesetzt werden.
- Wenn die Schwerpunktprüfung an allen 113 Standorten durchgeführt werden soll, kann man die Prüfobjekte nicht vorgeben. Vielmehr muss dann das Segment der Schwerpunktprüfung so eng gewählt werden, dass der Risikoanalyse der FKS-Mitarbeiter nicht mehr viel Raum bleibt. Ein Beispiel hierfür: Es wird vorgegeben, dass ausschließlich Gaststätten mit einer bestimmten Kapazität zu prüfen sind.
- Wenn nur ein Teil der Standorte an der Evaluation teilnehmen muss, ist es hinsichtlich der Umsetzung der Prüfungen am besten, wenn die Standorte freiwillig an den Schwerpunktprüfungen teilnehmen. Dazu sind ihnen bestimmte Anreize zu geben, z.B. durch eine Lockerung der Zielvorgaben hinsichtlich der Aufdeckung der Schwarzarbeit oder die Entpflichtung des Standortes bei der Durchführung der nächsten Schwerpunktprüfung. Ferner kann die Akzeptanz von Schwerpunktprüfungen nach genauen Vorgaben dadurch erhöht werden, dass diese den normalen Schwerpunktprüfungen nach eigener Risikoanalyse „beigemischt“ werden. Insgesamt ist zu bedenken, dass die Teilnahme an Schwerpunktprüfungen bereits jetzt bei den Standorten nicht zu den beliebtesten Aufgaben gehört.

- Die Wiederholung von Schwerpunktprüfungen in derselben Branche, wie sie in einigen der in Abschnitt 4 dieses Berichts vorgeschlagenen Projekten vorgesehen sind, scheint dagegen kein besonderes Problem darzustellen, vielmehr gibt es bereits jetzt Schwerpunktprüfungen in enger zeitlicher Folge (z.B. einmal jährlich in der Bauwirtschaft).

(2) Der Datenabgleich aus dem DALEB-Verfahren könnte insofern Grundlage für eine geeignete Datenquelle sein, als prinzipiell die Gesamtheit der Überschneidungen von Beschäftigungszeiten und Leistungsbezugszeiten in der Grundgesamtheit aller Leistungsempfänger erfasst wird. Informationen über den tatsächlich mit solchen Überschneidungen verbundenen Missbrauch von Leistungen der BA sind erhältlich, wenn die Informationen aus dem DALEB-Verfahren mit Daten der FKS über Ahndungen verknüpft werden. Daher könnte die Zahl der Ahndungen als Indikator für die Verbreitung bestimmter Formen des Leistungsmissbrauchs verwendet werden. Nachteil ist hier die mangelnde Datenqualität, die auf die Restriktionen bei den Agenturen für Arbeit zurückzuführen ist. Diesen Faktor sollte man dadurch eingrenzen, dass man sich bei der Analyse auf Agenturen beschränkt, die das DALEB-Verfahren mit hinreichender Genauigkeit und Regelmäßigkeit anwenden. Ein zweiter Nachteil besteht in der Tatsache, dass es hier nur um einen verhältnismäßig eng umrissenen Kreis von Tatbeständen geht.

(3) Werkverträge: Eine weitere Möglichkeit, zu einer repräsentativen Datenbasis zu gelangen, sind die Daten, die auf die Prüfung von Werkverträgen (aufgrund bilateraler Regierungsvereinbarungen) zurückgehen. Der Grund ist, dass hier im Prinzip eine Vollerhebung vorgenommen wird, da die zu entrichtenden Gebühren einen Anspruch auf Prüfung begründen. Allerdings wird in der Praxis nicht an allen FKS-Standorten in gleicher Weise vorgegangen. Insbesondere dort, wo besonders viele Werkverträge vorkommen, wird jeder einzelne in geringerem Maße geprüft. An einigen Standorten spielen auch Risikoanalysen bei der Auswahl der tatsächlich geprüften Werkverträge eine Rolle. Sinnvoll wäre es deshalb, nur Daten aus Standorten zu nutzen, die bei der Prüfung von Werkverträgen ähnlich verfahren. Eine Auswertung der Datenbasis Werkverträge in Verbindung mit Daten aus dem Ermittlungs- und Ahndungsprozess lässt Schlüsse darüber zu, in welchem Umfang die bilateralen Regierungsvereinbarungen missbräuchlich in Anspruch genommen wurden. Die Kontingente für derartige Beschäftigungsverhältnisse betragen etwa 50.000 Jahresarbeitskräfte. Aussagen zu sonstigen Formen von Schwarzarbeit oder illegaler Beschäftigung können damit nicht getroffen werden.

(4) Die verdachtlosen Prüfungen lassen hinsichtlich der Zielsetzung einer annähernd repräsentativen Darstellung des Umfangs der Schwarzarbeit aufschlussreiche Ergebnisse erwarten. Verdachtslosigkeit ist allerdings nicht in dem Sinne zu verstehen, dass nach einem Zufallsprinzip geprüft wird. Vielmehr ist auch hier eine Risikoanalyse der Mitarbeiter der FKS vorgeschaltet. Infolgedessen werden hauptsächlich solche Objekte zu Prüfungen herangezogen, bei denen die Ausgangsvermutung, dass dort Schwarzarbeit anzutreffen ist, verhältnismäßig hoch ist. Zudem hängt die Risikoanalyse von schwer kontrollierbaren Faktoren ab, so dass auch Vergleiche zwischen Standorten oder zwischen Branchen in der Regel kein exaktes Abbild des relativen Vorkommens von Schwarzarbeit ermöglichen. Der Grad, in dem

diese Daten nichtrepräsentativ sind, könnte sich zwischen Branchen unterscheiden, so dass die Repräsentativität im Einzelfall zu diskutieren wäre.

Ergebnisvariable legale Beschäftigung

Legale Beschäftigung wird nicht durch die Arbeitsstatistiken der FKS erfasst. Daher ist es bei dieser Ergebnisvariablen notwendig, auf Daten anderer Datenproduzenten zurückzugreifen. Hierbei kommen vor allem die Daten der Bundesagentur für Arbeit in Betracht. Diese Daten basieren, soweit sie Beschäftigungsverhältnisse umfassen, auf den Meldungen bei den Rentenversicherungsträgern und enthalten alle neu aufgenommenen legalen Beschäftigungsverhältnisse einschließlich der geringfügigen Beschäftigung. Nicht von der BA erhältlich sind Neugründungen von selbständig Erwerbstätigen. Auch in dieser Variable könnte sich jedoch die Tätigkeit der FKS niederschlagen. In diesen Fällen ist zu prüfen, ob Daten aus Gründungsdatenbanken wie dem KfW- Gründungsmonitor oder dem ZEW-Gründungspanel verwendet werden können oder ob eine eigene Erhebung zumindest auf diesen Datenbasen aufbauen kann. Daneben könnte die BA über Daten zu Neugründungen von selbständig Erwerbstätigen verfügen, soweit diese Neugründungen durch die BA gefördert wurden (z.B. Ich-AG).

Ergebnisvariable Rechtsbewusstsein

Information über das Rechtsbewusstsein ist in erster Linie direkt über Umfragen zu erhalten, während indirekte Verfahren (d.h., der Rückschluss von tatsächlichem Verhalten auf das Rechtsbewusstsein) mit großen Unsicherheiten behaftet sein dürfte. Die Sekundäranalyse vorhandener Daten zum Rechtsbewusstsein (wie z.B. die World Value Surveys oder des Eurobarometer) dürfte ausscheiden, weil die Daten räumlich nicht hinreichend disaggregiert sind. In selbst durchgeführten Umfragen kann die Auswahl der Befragten durch die Evaluation selbst vorgenommen werden, so dass das Problem der Repräsentativität beherrscht werden kann.

Die Durchführung von Umfragen ist verhältnismäßig aufwendig. Darüber hinaus ist die Gestaltung des Fragebogens für die Qualität der Ergebnisse im Bereich der Schwarzarbeit kritisch, da die Tendenz bestehen könnte, sich im Fragebogen als „gesetzestreu“ darzustellen als es dem tatsächlichen Rechts- bzw. Unrechtsbewusstsein entspricht. Dies dürfte für Unternehmen sogar noch stärker gelten als für Privatpersonen. Dieser Gesichtspunkt wird in Abschnitt 4.1 aufgenommen.

Fazit

Im Rahmen der Tätigkeit der FKS werden bisher keine Daten generiert, die das Phänomen Schwarzarbeit gemäß den Kriterien der statistischen Repräsentativität erfassen. Allerdings können in einigen Bereichen (DALEB-Verfahren und in den Arbeitsstatistiken festgehaltene Werkverträge) prinzipiell Datenbasen geschaffen werden, die durch eine statistische Vollabdeckung der betreffenden Grundgesamtheiten gekennzeichnet sind, wenn auch die Ausschöpfung teilweise erheblich unter dieser Marke bleibt. Hier hängt es von der Handhabung an den einzelnen Standorten der FKS ab, inwiefern der Umfang dieser Datenbestände und

die sich darin wiederfindenden Fälle das Ausmaß der Schwarzarbeit wirklichkeitsgetreu wiedergeben.

Die verdachtslosen Prüfungen können in der Datenbasis ProFiS getrennt von hinweisbezogenen Prüfungen dargestellt werden. Hier bietet sich eine weitere Möglichkeit, Daten aus den vorhandenen Beständen für eine Evaluation nutzbar zu machen. Auch wenn die statistische Repräsentativität dieser Daten in der Regel nicht gewährleistet ist, wäre es unzweckmäßig, sie als Datenbasis für eine Evaluation von vornherein auszuschließen. Ihre Verwendbarkeit für Evaluationszwecke muss jedoch im Einzelfall begründet werden.

Für die Wirkungsanalyse mit dem Indikator Schwarzarbeit als Ergebnisvariable wird man in den meisten Fällen zugleich mit der Evaluation eine geeignete Datenbasis schaffen müssen, etwa im Rahmen von Schwerpunktprüfungen. Diese Datenbasis sollte mit der Datenbank ProFiS verknüpft werden, um die Verdachtsfälle (als direktes Ergebnis der Schwerpunktprüfung) bis zu einer eventuellen Ahndung weiterverfolgen zu können. Dies ist bei den Schwerpunktprüfungen möglich, da das Auswertungsmodul „Controlling, Statistik, KLR“ eine „Nachverfolgung“ der Ergebnisse aus Schwerpunktprüfungen vorsieht. Zusammen mit den Daten aus ProFiS entstünde eine nicht nur national, sondern auch international einzigartige Datenbasis über das Vorkommen von Schwarzarbeit im geprüften Bereich. Diese würde zwar keine Zahlen über das Ausmaß der Schwarzarbeit auf dem Aggregationsniveau der deutschen Volkswirtschaft liefern. Sie würde jedoch Aussagen erlauben, die um ein Vielfaches präziser wären als die Ergebnisse der bislang zur Abschätzung von Schwarzarbeit verwendeten Verfahren.

Was den Indikator Beschäftigung anbelangt, stehen repräsentative Datenbestände der BA prinzipiell zur Verfügung. Für die Evaluation käme es darauf an, mit der BA zu einer Vereinbarung über die Datennutzung zu gelangen und die Daten für den jeweiligen Evaluationszweck mit Daten über den Einsatz von Maßnahmen durch die FKS zu verknüpfen. Der Indikator Rechtsbewusstsein ist ausschließlich mit Hilfe eigener Erhebungen zu erfassen, wobei die Gestaltungsmöglichkeiten wiederum sehr groß sind.

3.2.2 Exogene Variation des Instrumenteneinsatzes

Eine Kernfrage jedes Evaluationsvorhaben ist, worauf die Variation auf der Maßnahmenseite zurückgeht. Entscheidend ist, dass diese Variation nicht ihrerseits von Ergebnisvariablen abhängig ist. Letzteres könnte z.B. dann der Fall sein, wenn Prüfungen oder andere Maßnahmen besonders in den Bereichen vorgenommen werden, in denen ein besonders hohes Niveau der schattenwirtschaftlichen Aktivität zu verzeichnen ist. In diesem Fall besteht nicht nur eine Kausalbeziehung von der Maßnahme zum Ausmaß der Schattenwirtschaft, sondern auch in der anderen Richtung. Die Simultanität zwischen der Maßnahme und dem Ausmaß der Schwarzarbeit führt dazu, dass Schätzungen über die Wirksamkeit der Maßnahmen systematisch verzerrt sind. Auf der Basis solcher verzerrter Schätzungen sind keine verlässlichen Aussagen über den interessierenden Wirkungszusammenhang möglich.

In diesem Abschnitt wird wir auf zwei Vorgehensweisen eingegangen, mit Hilfe derer das beschriebene Problem vermieden werden kann. Die erste ist die Nutzung experimenteller Variation in den Maßnahmen, die durch die Evaluation vollständig kontrolliert wird. Die zweite ist die Ausnutzung natürlicher Variation im Maßnahmeeinsatz, die nicht ihrerseits vom Ergebnisindikator abhängig ist. Die Darstellung geht unmittelbar auf die Anwendung dieser Vorgehensweisen bei der Evaluation der FKS-Maßnahmen ein. Eine allgemeinere Darstellung mit Beispielen aus der Literatur ist im Anhang enthalten.

Experimentelle Variation

Experimentelle Variation bedeutet, dass bestimmte, im Rahmen der Evaluation durch einen Zufallsprozess ausgewählte Individuen von der Maßnahme betroffen sind, während andere nicht davon betroffen sind. Der Vorteil dieser Vorgehensweise besteht darin, dass ein Vergleich der Mittelwerte der Ergebnisindikatoren zwischen den beiden durch den Zufallsprozess geschaffenen Gruppen direkt den gesuchten Kausaleffekt wiedergeben kann.

Die Tatsache, dass die Individuen durch einen Zufallsprozess ausgewählt werden, setzt der Anwendung der experimentellen Vorgehensweise klare Grenzen. In Bereiche, wo Rechtsansprüche bestehen oder bei der Anwendung der Maßnahme andere verbindliche rechtliche Vorgaben bestehen, kann diese Vorgehensweise nicht angewendet werden. Das betrifft im Fall der FKS insbesondere den Bereich der Ermittlung und den Bereich der Ahndung. Daneben kann es sich aus praktischen Gründen verbieten, die Anwendung der Maßnahme durch die Evaluation zu steuern, wenn dies einen unververtretbaren Aufwand darstellen würde. Gut geeignet sind daher nur Maßnahmen, die einen verhältnismäßig geringen Aufwand verursachen. Dabei ist beispielsweise an Maßnahmen der Prävention wie Öffentlichkeitsarbeit oder Pressearbeit, möglicherweise auch an Präventionsprüfungen zu denken.

Eine wichtige Frage bei der Umsetzung eines experimentellen Evaluationsprojekts ist, ob die Maßnahme nur über Standorte der FKS variiert werden kann oder ob die Variation auch innerhalb der Zuständigkeit der Standorte, im Extremfall also auf der Ebene des Individuums erzeugt werden kann. Bei der Pressearbeit ist der Ausstrahlungsbereich der Maßnahme unbestimmt, der Kreis der Individuen, die „Empfänger“ der durch die Maßnahme ausgesandten „Botschaft“ sind, ist nicht klar begrenzt. Hier können nur Regionen für die Maßnahme zufällig ausgewählt werden und die Durchschnitte in der Ergebnisvariablen zwischen den Regionen verglichen werden. Hierbei stellt sich speziell das Problem, dass Presseberichte in lokalen Medien unter Umständen auch eine weitere Verbreitung in der Zielgruppe haben, als es dem unmittelbaren Einzugsbereich des Mediums entspricht. Ein anderes Instrument, bei dem über den Regionenvergleich vorgegangen werden könnte, wäre die Einsatzintensität von Prüfung und Ermittlung in einer bestimmten Branche. Hier wäre es möglich, innerhalb einer Branche ein „regionales Prüfklima“ zu erzeugen, das sich von Region zu Region unterscheidet. Von diesem Prüfklima sind jedoch alle Arbeitsmarktbeteiligten innerhalb der Region betroffen, so dass keine Variation auf der Ebene des Individuums entsteht.

Andere Maßnahmen zielen genauer auf bestimmte Individuen als Empfänger von Botschaften. Ein Beispiel hierfür sind angekündigte Prüfungen bei bestimmten Arbeitgebern oder Ar-

beitnehmern, wie sie im Bereich der Buchprüfungen auch derzeit schon vorgenommen werden. So könnte einer Gruppe von Unternehmen ein Schreiben zugesandt werden, das Prüfungen in einer bestimmten Intensität ankündigt. Geprüft würden diese Unternehmen und eine Kontrollgruppe von Unternehmen, denen keine zusätzlichen Prüfungen angekündigt wurden. Im Fall von Arbeitnehmern könnte beispielsweise bestimmten, aus einer Stichprobe zufällig ausgewählten Personen oder Leistungsempfängern Informationen vermittelt werden, die deren subjektive Einschätzung des Risikos einer Aufdeckung von Schwarzarbeit beeinflussen. Der Effekt dieser Maßnahme kann dann berechnet werden, indem man die Häufigkeit bestimmter Vorkommnisse (ob das Individuum später in Schwarzarbeit angetroffen bzw. verurteilt wird, ob das Individuum einer legalen Tätigkeit nachgeht usw.) zwischen dieser Auswahl und den anderen Angehörigen der Stichprobe vergleicht.

Der Vorteil der Maßnahmevariation innerhalb des Standortes besteht darin, dass die Evaluation nur an wenigen Standorten durchgeführt werden muss. Dies würde helfen, die Kosten der Evaluation zu begrenzen. Beruht die Evaluation dagegen nur auf der Variation zwischen den Standorten, verfügt die Evaluation über maximal 113 unabhängige Datenpunkte, also eine verhältnismäßig kleine Datenbasis. In diesem Fall ist es essentiell, dass möglichst alle Standorte in die Evaluation einbezogen werden.

Die Varianz der Ergebnisse lässt sich mindern, indem die Daten zweimal erhoben werden, einmal vor und nach der Durchführung der Maßnahme. Auch hier sind Kostengesichtspunkte gegen den Gewinn an Genauigkeit abzuwägen. Da experimentelle Vorgehensweisen bei der FKS bisher nicht angewendet wurden, wäre eine besondere interne Kommunikationsstrategie nötig, um die Kooperation auf der lokalen Ebene sicherzustellen. Dies ist insbesondere bei einem Evaluationsdesign erforderlich, das den Maßnahmeinsatz innerhalb des Standorts variiert.

Natürliche Variation

Natürliche Variation bedeutet, dass der Maßnahmeinsatz von einem exogenen Ereignis gesteuert wird, das seinerseits nicht durch die Schwarzarbeit beeinflusst wird. Auch bei der exogenen Variation kann man zwischen Variation auf der Ebene des Standorts und Variation auf den darunter liegenden Ebenen, beispielsweise der Ebene des einzelnen Arbeitgebers, unterscheiden. Natürliche Variation auf der Ebene des Standortes könnte z.B. darauf beruhen, dass die einzelnen Standorte aus bestimmten Gründen eine unterschiedliche personelle Iststärke aufweisen, wenn diese Unterschiede nicht ihrerseits auf der Dringlichkeit der Bekämpfung der Schwarzarbeit an den einzelnen Standorten beruhen. Generell ist es außerordentlich schwierig, eine plausible Quelle natürlicher Variation für den Einsatz bestimmter staatlicher Maßnahmen zu finden, da diese Maßnahmen jeweils hinsichtlich ihres Erfolgsgrades gesteuert werden.

Ausschließlich natürliche Variation kann verwendet werden, wenn untersucht werden soll, wie durch die Ahndung bewirkte Unterschiede in der Strafhöhe auf das Verhalten der Arbeitsmarktteilnehmer wirken. Dies liegt daran, dass die Strafhöhe nicht den Maßgaben der Evaluation folgen kann. Regionale Variation in der Strafzumessung scheint durchaus zu be-

stehen. Ein Evaluationsprojekt zu dieser Frage ist jedoch mit dem Problem belastet, dass bei der Erhebung der Strafhöhe für den Tatbestand ausreichend kontrolliert werden muss (Feld und Larsen, 2005: 35f). Es erscheint fraglich, ob dies angesichts der verfügbaren Daten gelingen kann. Zum anderen ist die Endogenität der Strafhöhe hinsichtlich des Volumens der Schwarzarbeit problematisch: Wenn die Strafzumessung mit der Häufigkeit und Schwere der festgestellten Fälle variiert, kann der Effekt der Strafhöhe auf das Verhalten nicht mehr sauber identifiziert werden.

In vorhandenen Studien zur Evaluation anderer staatlicher Maßnahmen werden häufig gesetzliche Änderungen als Quelle natürlicher Variation verwendet. Für die Evaluation der generalpräventiven Wirkungen der Ahndung ließen sich Änderungen im gesetzlichen Rahmen bzw. Änderungen des Bußgeldkatalogs, sofern diese im Zeitraum der Evaluation durchgeführt werden, als Ausgangspunkt nutzen. Dabei ist es problematisch, allein den gewählten Indikator vor und nach der Änderung im gesetzlichen Rahmen zu vergleichen, weil diese Änderungen auch durch die zeitliche Veränderung anderer Determinanten verursacht sein können. Erheblich aussagekräftiger werden die Resultate, wenn die gesetzlichen Änderungen nur einen Teil der Arbeitsmarktteilnehmer betreffen. Häufig ist es plausibel anzunehmen, dass zeitliche Veränderungen anderer Determinanten in gleicher Weise auf die von der Gesetzesänderung Betroffenen als auch die Nichtbetroffenen wirken, so dass der Effekt des Gesetzes isoliert werden kann (sogenannter Differenzen-von-Differenzen-Ansatz, vgl. den Anhang).

Insgesamt wurde die Frage nach dem Vorhandensein natürlicher Variation in den durchgeführten Interviews sehr skeptisch beurteilt. So wurde geäußert, dass der Instrumenteneinsatz stets durch eine eigene Risikoanalyse gesteuert werde. Fragen nach Determinanten des Instrumenteneinsatzes, die nicht vom Standort zu steuern sind, ergaben insgesamt nur wenig Aufschluss über geeignete Variation.

Sofern natürlich Variation zwischen Regionen benutzt werden soll, ergibt sich bei 113 Standorten überdies ein Fallzahlenproblem. Bei der Ausnutzung natürlicher Variation ist dieses erheblich gravierender als bei experimenteller Variation, da durch den weniger als vollständigen Zusammenhang zwischen dem natürlichen Ereignis und der Maßnahme zusätzliche stochastische Variation in die Analyse eingeht. Diese Grenzen für die Ausnutzung natürlicher Variation lassen es geraten scheinen, ein experimentelles Vorgehen zu wählen, wo immer dies vom Aufwand her vertretbar ist.

3.2.3 Schätzung kausaler Effekte

Auf der Grundlage der Daten und einer exogenen Variation im Einsatz der Instrumente lassen sich die kausalen Effekte der Maßnahmen bestimmen. Auf die dabei verwendeten Methoden wird im Anhang zu diesem Bericht eingegangen, so dass an dieser Stelle auf die Darstellung der Schätzmethode verzichtet wird.

Beispiele für mögliche Ergebnisse der Kausalanalyse sind die folgenden:

- Um wie viel Prozentpunkte reduziert sich infolge der untersuchten Maßnahme das Verhältnis der Verdachtsfälle oder Ahndungen zur Zahl der vorgenommenen Prüfungen?
- Um wie viel Prozentpunkte erhöht sich durch die Maßnahme der Anteil der Personen, die in einer Befragung Schwarzarbeit ablehnen?
- Um wie viel Prozent erhöht sich aufgrund der Maßnahmen die legale Beschäftigung?

Der Kausaleffekt richtet sich auf die Effekte, die bei denjenigen Individuen entstehen, welche von den Maßnahmen erfasst werden. Dahinter steht die Politikfrage, welcher zusätzliche präventive Effekt von der evaluierten Maßnahme auf die Teilnehmergruppe ausgegangen ist. Eine andere Politikfrage wäre, welche Effekte die Maßnahmen gehabt hätten, wenn sie auch die Kontrollgruppe erfasst hätte. Insofern als sich die kausalen Effekte zwischen Teilnehmer- und Kontrollgruppe unterscheiden, können die Ergebnisse für beide Frage unterschiedlich ausfallen.

In den meisten Fällen wird aufgrund der Größe der Datenbasis nur ein einziger Kausaleffekt für die gesamte Stichprobe von Prüfungsfällen geschätzt. In Wirklichkeit können die Effekte durchaus heterogen sein, z.B. könnte die präventive Wirkung vom Bildungsgrad der Betroffenen oder von der Einträglichkeit der Schwarzarbeit abhängen. Solche Differenzierungen der Ergebnisse sind nur bei einer sehr umfassenden Datenbasis möglich. Wenn Variation im Maßnahmeneinsatz auf der Ebene des einzelnen Arbeitsmarkteteiligen vorliegt, können bei entsprechender Anpassung der Fallzahl unterschiedliche Effekte geschätzt werden. Bei Variation auf der Ebene des einzelnen FKS-Standes ist dies jedoch nicht möglich, da die Zahl der Standorte mit 113 zu niedrig ist.

Im Einzelfall zu analysieren, ob es sich um langfristige Effekte auf das Verhalten handelt oder lediglich um kurzfristige Strategien zur Verminderung des Aufdeckungsrisikos (vgl. dazu insbesondere den Abschnitt 4.2). Ebenfalls ist zu thematisieren, ob die Effekte im Sinne einer tatsächlichen Verminderung von Schwarzarbeit interpretiert werden können, oder ob sie in Wirklichkeit nur das Ausweichen von Schwarzarbeit vom Bereich der durchgeführten Maßnahme in andere Bereiche messen. Die Schätzung von kausalen Effekten auf der Ebene der einzelnen Person kann beispielsweise Effekte auf andere Gruppen der Bevölkerung unberücksichtigt lassen. So stellte sich in der Studie von Di Tella und Schargrodsky (2004) heraus, dass der zusätzliche Einsatz von Polizeikräften zwar einen präventiven Effekt in der unmittelbaren Umgebung der Einsätze hatte, dass sich jedoch die kriminelle Aktivität in anderen räumlichen Bereichen verstärkte, so dass per saldo kein Effekt auf die Kriminalität nachzuweisen war. Ein solcher Effekt wäre insbesondere unter Kosten-Nutzen-Aspekten problematisch.

3.3 Implementationsanalysen

Begleitend zur quantitativen Evaluation sollten qualitative Untersuchungsschritte vorgenommen werden. Hierbei geht es um Befragungen bei den Arbeitsmarkteteiligen sowie bei der FKS und ggf. weiteren Akteuren, insbesondere in der Arbeitsverwaltung. Es sollten dabei die

Wirksamkeit der Instrumente und ihre verhaltenslenkenden Wirkungen aus der Sicht der Beteiligten erfragt werden, um einen Bezugspunkt für die quantitativen Ergebnisse zu gewinnen. Gleichzeitig sollte die Implementationsanalyse die Bedeutsamkeit bestimmter für die Evaluation kritischer Punkte aufdecken. Beispiele hierfür sind die Spielräume der einzelnen Beamten oder der Standorte der FKS, die Varianz im Verhalten über Regionen, die Intensität des Einsatzes bestimmter Strategien sowie die budgetären Spielräume. Insbesondere sollte die Implementationsanalyse klären, inwieweit die Vorgaben der Abteilung FKS bei der Steuerung der Evaluation umgesetzt werden, also inwieweit z.B. tatsächlich dem Konzept einer Zufallsauswahl gefolgt wird, wo die Evaluation eine solche vorsieht. Zudem kann die Implementationsanalyse dabei helfen, die Qualität der anfallenden Daten zu bewerten.

3.4 Kosten-Nutzen-Analyse

Eine über die Wirkungsanalyse hinausgehende Frage betrifft die Durchführbarkeit einer Kosten-Nutzen-Analyse. Im Rahmen einer Kosten-Nutzen-Analyse wird der Aufwand, der mit dem Einsatz bestimmter Instrumente verbunden ist, mit dem Ertrag des Instrumenteneinsatzes verglichen. Es stellt sich dabei die grundlegende Frage, was genau unter dem Ertrag oder Nutzen des Instrumenteneinsatzes zu verstehen ist. Die in Abschnitt 4 bei den jeweiligen Evaluationsprojekten diskutierten Abwägungen von Kosten und Nutzen beschränken sich auf die unmittelbaren fiskalischen Effekte der Tätigkeit der FKS. Weitere gesamtwirtschaftliche Wirkungen der Bekämpfung der Schwarzarbeit sind schwierig zu quantifizieren und werden daher außer Acht gelassen. Hierbei ginge es um das gesamtwirtschaftliche Einkommen, das infolge der Bekämpfung der Schwarzarbeit zusätzlich durch legale Arbeit geschaffen würde. Hiervon abziehen müsste man das Einkommen, das in Schwarzarbeit erzielt wird.

In aller Regel werden die gesamtwirtschaftlichen Erträge der Tätigkeit der FKS höher sein als die unmittelbar fiskalischen. Daher sind die Ergebnisse der vorgeschlagenen Kosten-Nutzen-Analysen so zu verstehen, dass sie auf einer zurückhaltenden Schätzung der wahren gesamtwirtschaftlichen Effekte der Bekämpfung der Schwarzarbeit beruhen.

Die Kosten-Nutzen-Analyse sollte auf den verhaltenslenkenden Effekten aufbauen, also den Ergebnissen der Evaluation hinsichtlich der festgestellten Änderungen im Verhalten und des Bewusstseinswandels. Bevor eine sinnvolle Kosten-Nutzen-Analyse durchgeführt werden kann, muss also die Frage beantwortet werden, welchen Effekt der Einsatz der Instrumente der FKS hat. Dabei ist im Wesentlichen an die Verminderung der Schwarzarbeit an sich und der Ausdehnung von legaler Beschäftigung zu denken.

Eine Verminderung von Schwarzarbeit hat unmittelbare positive fiskalische Effekte, falls sie mit dem Missbrauch von Sozialleistungen einhergeht. In dem Maße, in dem der Leistungsmissbrauch vermindert wird, werden missbräuchlich bezogene Sozialleistungen eingespart. Die unmittelbaren fiskalischen Effekte einer Ausdehnung der legalen Beschäftigung liegen ebenfalls auf der Hand. Sie speisen sich im Wesentlichen aus zwei Quellen: Einerseits steigen bei einer Ausdehnung der legalen Beschäftigung die Beitragszahlungen in die Kassen

der Sozialversicherungssysteme. Darüber hinaus steigt tendenziell auch das Aufkommen an Einkommensteuer. Im Vergleich mit diesen fiskalischen Effekten dürften die Einnahmen, die mit der Aufdeckung von Schwarzarbeit durch den Instrumenteneinsatz selbst verbunden sind, von untergeordneter Bedeutung sein.

Zur Darstellung der Kosten des Instrumenteneinsatzes der FKS können Angaben aus der Kosten- und Leistungsrechnung verwendet werden. Die Kosten sind weit überwiegend Personalkosten und können daher durch die Feststellung, welcher Personaleinsatz für eine bestimmte Maßnahme nötig ist, verhältnismäßig exakt berechnet werden. Zu beachten ist, dass einige Maßnahmen, insbesondere im Bereich der Prävention, unmittelbar nur geringe Kosten verursachen. Beispielsweise ist eine Pressemitteilung unter Umständen unter minimalen Kosten zu erstellen. Die Wirkung solcher Maßnahmen beruht aber oft darauf, dass Unternehmen und Arbeitnehmer durch die Maßnahme auf die Existenz bzw. die Tätigkeit der FKS hingewiesen werden. Damit sind die eigentlichen Kosten einer solchen Maßnahme weit höher als unmittelbar sichtbar, weil eine Verhaltensänderung bei den Adressaten dauerhaft nur dann zu erzielen ist, wenn die „Drohkulisse“ der Prüfungs- und Ermittlungskapazitäten der FKS glaubwürdig ist. Konsequenterweise müssten daher die Kosten, die dadurch entstehen, dass Prüfungs- und Ermittlungskapazitäten vorgehalten werden, teilweise präventiven Maßnahmen zugerechnet werden. Derartige Zurechnungen von Kostenbestandteilen dürften sich in der Praxis als schwierig erweisen. Daher sollte das Instrument der Kosten-Nutzen-Analyse nur sehr zurückhaltend eingesetzt werden, wenn Maßnahmen der Presse- und Öffentlichkeitsarbeit evaluiert werden.

Die im Abschnitt 4 vorgestellten Evaluationsprojekte sollten jeweils durch eine speziell auf das Vorgehen und die Datenverfügbarkeit zugeschnittene Kosten-Nutzen-Analyse begleitet werden. Allerdings sind bei der Interpretation der Ergebnisse gewisse Einschränkungen zu berücksichtigen. Es ist vor allem zu beachten, dass die vorgestellten Projekte jeweils nur einen Teilbereich der Tätigkeit der FKS beleuchten. Die daraus jeweils abgeleiteten Abwägungen von Kosten und Nutzen haben daher auch nur eingeschränkte Bedeutung. Insbesondere sollten die Ergebnisse der projektbezogenen Kosten-Nutzen-Analysen nicht ohne weiteres auf andere Bereiche der Tätigkeit der FKS übertragen werden.

3.5 Vermittlungs- und Kommunikationsstrategie

Es ist davon auszugehen, dass eine Evaluation in den meisten Fällen aufgrund der Erhebung von Daten Einfluss auf die Arbeitsweise der Beamten an den einzelnen Standorten hätte. In den durchgeführten Interviews wurde hier vor allem seitens der Standorte unterstrichen, dass die Beamten der FKS das Anliegen der Evaluation verstehen und aus eigenem Antrieb heraus unterstützen müssen, wenn eine Evaluation erfolgreich sein soll. Dies gilt umso mehr, je stärker die Evaluation in die Abläufe an den Standorten und in die Entscheidungsspielräume der Beamten eingreift. Auch in anderen Bereichen als in der Bekämpfung illegalen Verhaltens ist aus der Literatur bekannt, dass Widerstand bei der Umsetzung der von der Evaluation vorgesehenen Maßnahmen dazu führte, dass die Zielsetzungen der Evaluation nicht erreicht wurden (siehe dazu z.B. Smith, 2000).

Manche Vorgehensweisen bei der Evaluation verlangen dagegen keine Kommunikationsstrategie oder nur geringfügige Aktivitäten in dieser Hinsicht. So brauchen in einigen Fällen die Standorte vermutlich gar nicht Kenntnis davon zu erhalten, dass sie an einer Evaluation teilnehmen. Da sich die Kommunikationsstrategie stark zwischen den einzelnen Evaluationsprojekten unterscheidet, wird darauf im folgenden Kapitel näher eingegangen.

Die Durchführung der Kommunikationsstrategie obliegt dem Pressereferat der FKS. Es können sehr unterschiedliche Vorgehensweisen gewählt werden. Ausschlaggebend ist dabei vor allem die Zahl der beteiligten Standorte. Sofern alle 113 Standorte der FKS beteiligt sind, wird die Kommunikationsstrategie aus Gründen des Aufwands überwiegend auf dem Schriftverkehr mit den Standorten beruhen müssen. Bei einer geringeren Zahl von Standorten ist es dagegen möglich, Zielsetzung und Durchführung von Evaluationsprojekten im persönlichen Kontakt zu kommunizieren. Dabei bietet es sich an, Seminare oder ähnliche Veranstaltungen durchzuführen, auf denen das Vorgehen der Evaluation mit einer größeren Zahl unterschiedlicher Teilnehmer besprochen werden kann. Diese Veranstaltungen wären frühzeitig durchzuführen, so dass bei der Planung der Evaluation Anregungen der Teilnehmer aufgenommen werden könnten.

Ein weiterer Aspekt ist, dass neben der internen Kommunikationsstrategie vielfach auch eine Kommunikationsstrategie in die Öffentlichkeit erforderlich ist. Damit ist nicht die Kommunikation der Evaluationsergebnisse an Medien und die weitere Öffentlichkeit gemeint, sondern Unterstützung während der Durchführung der Evaluation. Die einzelnen Schritte der Evaluation werden je nach Umfang der damit verbundenen Maßnahmen in der Öffentlichkeit mehr oder weniger bekannt werden. Hier sollte frühzeitig bedacht werden, inwieweit das Pressereferat den beteiligten Standorten Hilfestellung bei der Kommunikation mit den Arbeitsmarktteilnehmern und der weiteren Öffentlichkeit leisten kann, damit beispielsweise nicht durch eine unbeeinflusste Berichterstattung in den Medien das Ziel der Evaluation konterkariert wird.

4 Einzelne Evaluationsprojekte zur Ermittlung kausaler Wirkungen der Tätigkeit der FKS

In diesem Teil des Berichts werden konkrete Vorschläge für Evaluationsvorhaben vorgestellt. Diese decken nicht systematisch das gesamte Spektrum der Tätigkeit der FKS ab, sondern versuchen, dort in die Tiefe zu gehen, wo Evaluationen realisierbar und mit vertretbarem Aufwand durchzuführen sind. Die Auswahl der Vorhaben folgt den Gesichtspunkten, die im Abschnitt 3 entwickelt worden sind. Ausgehend von diesen Projekten werden dann Varianten diskutiert, um zu zeigen, wie das Evaluationskonzept auf andere Bereiche übertragen werden könnte, wo mehrere Optionen und Alternativen für die Durchführung der Hauptstudie vorhanden sind und wie der Umfang der Evaluationsstudie variiert werden kann.

Im Kern bestehen diese Evaluationsprojekte aus der Kombination von untersuchten Maßnahmen, Datenbasis und Evaluationsmethodik. Ziel ist es, die kausalen Wirkungen der Maßnahmen der FKS auf die Nutzung von Schwarzarbeit, den Umfang der legalen Beschäftigung und das Rechtsbewusstsein darzustellen. Bei der Ausarbeitung der Vorschläge werden die einzelnen Schritte der Evaluation detailliert beschrieben. Die Darstellung der skizzierten fünf Evaluationsprojekte behandelt jeweils die folgenden Punkte:

- die Fragestellung der Evaluation;
- die Art, wie Variation in den Maßnahmen erzeugt wird;
- die verwendeten Indikatoren;
- die Erfassung von Maßnahmen und Indikatoren in den Daten;
- Kommunikationsstrategie, Implementationsanalyse und Kosten-Nutzen-Analyse;
- besondere Probleme, die bei der Umsetzung auftauchen können;
- eine Einschätzung des Aufwands, den die Untersuchung hinsichtlich des Maßeinsatzes und hinsichtlich der Datenerhebung verursachen würde;
- mögliche Varianten des Evaluationsdesigns;
- die zu erwartenden Ergebnisse einer Evaluation.

Die Darstellung gliedert sich in drei Projekte mit experimentell erzeugter (Abschnitte 4.1 bis 4.3) und zwei Projekte mit natürlicher Variation (Abschnitte 4.4 und 4.5). Eine weitere Unterscheidung besteht hinsichtlich der Ebene, auf der Variation besteht. In zwei Projekten (4.1 und 4.4) variieren die Maßnahmen lediglich zwischen FKS-Standorten, in den übrigen auch zwischen Arbeitsmarkteteiligen, die in die Zuständigkeit des gleichen Standortes fallen. Schließlich wurde bei der Auswahl der Projekte darauf geachtet, dass diese ein relativ weites Spektrum an Maßnahmen von explizit präventiven Maßnahmen wie der Öffentlichkeitsarbeit (Abschnitt 4.1) bis hin zur Ahndung (Abschnitt 4.5) umfassen.

Die Evaluationsprojekte sind so angelegt, dass sie jeweils unabhängig voneinander durchgeführt werden können und nicht aufeinander aufbauen. Wenn mehr als eine Evaluation im gleichen Zeitraum durchgeführt wird, ist im Gegenteil darauf zu achten, dass sich diese Evaluationen nicht gegenseitig beeinflussen. Wenn beispielsweise eine Person oder eine Grup-

pe von Individuen bezüglich zweier Maßnahmen in der Teilnehmergruppe enthalten ist, kann der Effekt nicht mehr kausal auf eine der beiden Maßnahmen zurückgeführt werden. Ähnliche Probleme entstehen, wenn die Nichtteilnehmer eines Programms zum Teil an einem anderen Programm teilnehmen. Tendenziell ist das Problem umso gravierender, je weiter die Wirkung über die Arbeitsmarktbeschäftigten streut und je größer die Zahl derer, die von den Maßnahmen erfasst werden.

Auch wenn Evaluationsprojekte zu unterschiedlichen Zeitpunkten durchgeführt werden, können sich Probleme daraus ergeben, dass bestimmte Individuen oder ganze Regionen bereits von Maßnahmen beeinflusst worden sind, die in anderen Projekten durchgeführt wurden. Auch in diesen Fällen wären die Effekte den Maßnahmen nicht mehr ohne weiteres zuzuordnen. Inwieweit sich diese Interdependenzen ergeben, kann jedoch erst dann geprüft werden, wenn Durchführungszeiträume und betroffene Branchen und Regionen bekannt sind.

4.1 Regionales Prüfklima

4.1.1 Fragestellung

Durch den gezielten Einsatz von Instrumenten der Prüfung (insbesondere verdachtlose bzw. Präventionsprüfungen) wird die Abschreckungswirkung der FKS-Tätigkeit in einer Branche für eine gewisse Zeit um ein bestimmtes, zu definierendes Maß erhöht. Die Maßnahme bzw. das Maßnahmenbündel wird sich auf bestimmte Branchen oder sogar auf bestimmte Tätigkeiten innerhalb von Branchen konzentrieren, damit der Aufwand beim Einsatz der Instrumente begrenzt wird. Die zur Herstellung eines verstärkten Prüfklimas erforderlichen Prüfungen müssen zentral durch die Abteilung FKS angeordnet werden. Der erhöhte Prüfungseinsatz kann durch die Ausdehnung der Präsenz in der Fläche sowie durch Ankündigungen in der Presse oder gezielte Öffentlichkeitsarbeit verstärkt werden. Je nach Zuschnitt des Projektes haben Prüfungen und die anderen präventiven Maßnahmen ein unterschiedliches Gewicht.

Adressaten des Instrumenteneinsatzes können sowohl Unternehmen (also Nutzer von Schwarzarbeit) oder Einzelpersonen (Anbieter von Schwarzarbeit) sein. Bei der Umsetzung ist darauf zu achten, dass die Art der Maßnahmen möglichst ähnlich zu den sonst durchgeführten Maßnahmen in diesem Bereich ist, um die Übertragbarkeit der Ergebnisse der Studie auf die Arbeit der FKS zu gewährleisten. Dies erscheint bei Prüfungen weniger problematisch als bei Presse- und Öffentlichkeitsarbeit, bei der bislang anders vorgegangen wird, als es in diesem Projekt vorgesehen ist.

Der Ablauf des Projekts ist aus der folgenden Übersicht zu ersehen, die in den folgenden Abschnitten im Detail erläutert wird.

Tabelle 4.1: Zeitlicher Ablauf des Evaluationsprojekts „Prüfklima“

Schritt 1 (ca. Monat 1-3)	Erstmalige Datenerhebung durch Schwerpunktprüfung oder Befragung an allen Standorten der FKS. Alternativ: Daten werden aus ProFiS erhoben, was auch gleichzeitig mit Schritt 4 erfolgen kann.
Schritt 2 (ca. Monat 3)	Auswahl der Teilnehmerstandorte, an denen zusätzliche Prüfungen durchgeführt werden, durch einen Zufallsprozess (siehe Abschnitt 4.1.2). Vermittlung des zusätzlichen Aufwands im Rahmen der Kommunikationsstrategie.
Schritt 3 (ca. Monat 6-9)	Durchführung der Prüfungen und der ergänzenden präventiven Maßnahmen an den Teilnehmerstandorten. Dabei Implementationsanalyse.
Schritt 4 (ca. Monat 10-12)	Zweite Datenerhebung nach demselben Verfahren wie in Schritt 1
Schritt 5 (ca. Monat 13-15)	Datenauswertung, falls die Zahl der Verdachtsfälle als Indikator verwendet wird. Im Anschluss daran Kosten-Nutzen-Analyse. Alternative: Verwendung der Zahl der Ahndungen als Indikator. In diesem Fall müsste der zeitliche Rahmen so weit gesteckt werden, dass die weit überwiegende Mehrheit der Verfahren zum Zeitpunkt der Datenauswertung abgeschlossen ist. Entsprechend würde sich die Auswertung um ca. 12-24 Monate nach hinten verschieben.

4.1.2 Art der Variation

Variation entsteht auf der Ebene der Standorte durch Teilnahme oder Nichtteilnahme an einem Bündel an Maßnahmen. Der Teilnahmezustand des einzelnen Standortes wird durch einen Zufallsprozess festgelegt (Schritt 2 in Tabelle 4.1). Eventuell kann der Auswahl der Teilnehmerstandorte eine Einteilung nach Typen vorausgehen, z.B. nach dem Kriterium Großstadt/Ballungsraum vs. Kleinstadt/ländlicher Raum oder nach bestimmten Regionen. Damit würde sichergestellt werden, dass eine ausreichende Anzahl von Standorten bestimmter Typen teilnehmen bzw. nicht teilnehmen.

Der verstärkte Prüfungseinsatz an den Teilnehmerstandorten müsste durch Anweisungen durch die Abteilung FKS zentral gesteuert werden. Zur Erzeugung eines regionalen Prüfklimas können beispielsweise Schwerpunktprüfungen verwendet werden, wobei genaue Vorgaben über die Intensität und den Umfang der durchzuführenden Prüfungen gemacht werden müssten. Diese Vorgaben sind unerlässlich, um eine möglichst exakte „Dosierung“ einer zusätzlichen Aktivität der FKS bei der Bekämpfung von Schwarzarbeit zu erreichen.

4.1.3 Verwendbare Indikatoren

Vermutlich lässt sich der Effekt der untersuchten Maßnahme auf die Entwicklung der legalen Beschäftigung in der Teilnehmer- und Nichtteilnehmergruppe angesichts der relativ geringen Zahl von maximal 113 Beobachtungen (entsprechend der Zahl der Standorte) wegen des indirekten Wirkungszusammenhangs und infolge der Vielfalt der Determinanten der regionalen Beschäftigungssituation kaum hinreichend exakt bestimmen. Daher kommt der Indikator „legale Beschäftigung“ in diesem Projekt nicht in Betracht.

Weitaus besser geeignet ist der Umfang der Schwarzarbeit als Indikator. Hier ist die Wirkung der Maßnahme im Vergleich zu den anderen beiden Indikatoren direkter und damit statistisch leichter zu identifizieren. Dem steht der Nachteil gegenüber, dass dieser Indikator weniger leicht zu erfassen ist als die legale Beschäftigung, so dass in erheblichem Ausmaß zusätzlicher Aufwand entsteht. Zudem ergibt sich bei Schwarzarbeit als Indikator das Problem, dass sich die präventiven Effekte zusätzlicher Prüfungen und die durch zusätzliche Prüfungen erzielten zusätzlichen Ermittlungsergebnisse überlagern (siehe Abschnitt 4.1.4). Die zusätzlichen Ermittlungsergebnisse in Form von Verdachts- oder Ahndungsfällen sind jedoch für sich genommen ebenfalls ein interessanter Indikator, weil sie darüber Aufschluss geben, wie produktiv die Maßnahmen hinsichtlich des Ziels der Aufdeckung von Schwarzarbeit sind und ob zu erwarten ist, dass von den Maßnahmen präventive Effekte ausgehen.

Das Rechtsbewusstsein als Indikatorvariable wirft Schwierigkeiten auf, wenn sich die Maßnahmen in erster Linie an Unternehmen richten, so dass eine Unternehmensbefragung durchgeführt werden muss. Bei Unternehmen als Adressaten ist die Bereitschaft zu wahrheitsgetreuer Beantwortung von Fragen über Schwarzarbeit allgemein als gering einzuschätzen. Allerdings könnte man dieser Tatsache durch eine gezielte Befragungsstrategie entgegenwirken. So könnte man von der Belastung des Faktors Arbeit mit Abgaben ausgehen und nach Gegenstrategien der Unternehmen wie z.B. Einsparung von Personalkosten durch Rationalisierungen, Produktionsverlagerungen, Änderungen im Produktionsvolumen und der Nutzung von Schwarzarbeit als letzter Möglichkeit fragen. Erfahrungen des ZEW bei Unternehmensbefragungen deuten darauf hin, dass das „Framing“ der Fragen entscheidend für die Antwortbereitschaft auch bei Unternehmen ist (Ammermüller et al., 2003). Bei der Befragung sollte nicht darauf hingewiesen werden, dass diese im Auftrag des Bundesministeriums der Finanzen durchgeführt wird, weil Zweifel entstehen könnten, ob die Daten nicht doch zu ermittlungstechnischen Zwecken gebraucht werden. Auf den Datenschutz wäre in jedem Fall hinzuweisen.

4.1.4 Erfassung der Indikatoren in den Daten

Um die Varianz in der Verbreitung der Schwarzarbeit zu mindern, ist es erforderlich, den verwendeten Indikator zweimal zu erheben: einmal vor Durchführung der Maßnahmen und einmal im unmittelbaren Anschluss daran. Verglichen wird dann die Veränderung im Indikator an den Teilnehmerstandorten im Vergleich zu den Nichtteilnehmerstandorten. Bei einer Untersuchung, die nicht auf der Veränderung, sondern auf dem Niveau des Indikators beruht, wäre zusätzlich das unterschiedliche Niveau des Indikators zu erklären, was angesichts

der vielfältigen Ursachen von Schwarzarbeit nur unvollkommen gelingen dürfte. Es verbliebe erhebliche unerklärte Varianz, die die Trennschärfe der geschätzten Effekte erheblich beeinträchtigen würde.

Entscheidend ist ferner, dass die aus den zusätzlichen Prüfungen resultierenden Verdachtsfälle nicht in die gemessenen Indikatoren der Schwarzarbeit eingehen. In diesem Fall könnte man die verhaltenslenkenden Wirkungen der Prüfungen nicht von den Wirkungen auf die Zahl der aktenkundig gewordenen Fälle von Schwarzarbeit trennen. Das bedeutet, dass die im Zuge der zusätzlichen Prüfungen an den Teilnehmerstandorten anfallenden Prüfungsergebnisse im Zuge der Evaluation nicht benutzt werden dürfen, weil diese nicht nur die verhaltenslenkenden Effekte, sondern auch den zusätzlichen Instrumenteneinsatz wiedergeben. Es ist deshalb entscheidend, dass die Prüfungen, die der Datenaufnahme dienen, unabhängig von den Prüfaktivitäten stattfinden, deren Einfluss auf das Verhalten ermittelt werden soll.

Damit ist für die Zwecke dieser Untersuchung die Nutzung vorhandener Daten aus ProFiS problematisch. Dies gilt auch dann, wenn die Daten anderen Zeiträumen entstammen als denen, in denen die zusätzlichen Prüfungen und ergänzenden Maßnahmen durchgeführt werden. Denn es ist nicht auszuschließen, dass die Maßnahme auch über den unmittelbaren Zeitraum ihrer Durchführung hinaus Rückwirkungen auf die Prüfungsaktivität hat. Beispielsweise ist denkbar, dass durch die Durchführung der Maßnahme in den Teilnehmerstandorten zusätzliche Erkenntnisse gewonnen werden, die im Anschluss in der üblichen Prüfungs- und Ermittlungstätigkeit der am Ort tätigen Beamten genutzt werden. Dies würde die Daten, die im Zuge der Auswertung der Maßnahme gewonnen werden (beispielsweise durch eine weitere Schwerpunktprüfung) kontaminieren. Es könnte andererseits auch sein, dass die Zahl der Prüfungen in der betroffenen Branche nach dem Abschluss der Maßnahme „Prüfklima“ zurückgeht, weil nach Einschätzung des Standortes nun andere Branchen und Bereiche „an der Reihe“ sind.

Daher ist es zweckmäßiger, den Indikator Schwarzarbeit im Zuge von Schwerpunktprüfungen zu erfassen. Die Schwerpunktprüfung müsste wiederholt im gleichen Format und mit den gleichen Anweisungen jeweils vor und nach der Maßnahme durchgeführt werden. Was den zeitlichen Abstand zwischen der erstmaligen Datenerhebung und dem Beginn der evaluierten Maßnahmen betrifft, sollte ein Abstand von einigen, mindestens von drei Monaten zur erstmaligen Datenerhebung eingehalten werden, damit die präventiven Effekte der erstmaligen Datenerhebung die präventiven Effekte der zu prüfenden Maßnahmen nicht dominieren. Bei der Schwerpunktprüfung müsste nach dem in Abschnitt 3.2.1 skizzierten Verfahren vorgegangen werden. Statt der in der Schwerpunktprüfung erhobenen Verdachtsfälle von Schwarzarbeit könnten auch die sich aus der Prüfung ergebenden Ahndungen durch Strafen oder Bußgelder erfasst werden, da diese in der Datenbank ProFiS enthalten sind. Dabei könnte nicht allein die Zahl der Vergehen, sondern auch die Schwere der Vergehen in den Indikator eingehen.

Der Indikator „Rechtsbewusstsein“ kann durch eine Befragung in der jeweiligen Branche erfasst werden. Dabei würde man erheben, inwieweit diejenigen Formen der Schwarzarbeit, die von der Maßnahme berührt werden, als ethisch akzeptabel oder inakzeptabel angesehen

werden. Hierbei kommen einerseits Unternehmen in Betracht, zum anderen Arbeitnehmer. Unternehmen sind insofern die bessere Zielgruppe, als sie unmittelbar von der Maßnahme betroffen sind. Wenn in einem Maßnahmenbündel auch Instrumente der Presse- und Öffentlichkeitsarbeit zum Einsatz gekommen sind, könnten eventuell auch Arbeitnehmer befragt werden. Bei einer branchenspezifischen Vorgehensweise könnte es sich allerdings praktisch schwierig gestalten, Arbeitnehmer der jeweiligen Branche gezielt anzusprechen. Die Befragung sollte ebenfalls wiederholt vor und nach der Durchführung der Maßnahme stattfinden. Die Interviews sollten telefonisch durchgeführt werden, um eine möglichst hohe Ausschöpfung des Befragungsfeldes zu gewährleisten und den Aufwand der Befragung so gering wie möglich zu halten.

4.1.5 Kommunikationsstrategie

Eine Kommunikationsstrategie ist zum einen erforderlich, weil die Erzeugung eines regionalen Prüfklimas in einer bestimmten Branche durch zentrale Vorgaben für die FKS eine bisher noch nicht erprobte Vorgehensweise ist. Dabei sind einerseits die Vorgaben bei der Prüfung zu kommunizieren, und andererseits ist zu vermitteln, warum die Maßnahmen nur an bestimmten Standorten durchgeführt werden.

Bei der Ausarbeitung einer Kommunikationsstrategie ist zu beachten, dass im Rahmen des Evaluationsprojektes an manchen Standorten bis zu drei Schwerpunktprüfungen in relativ kurzer zeitlicher Folge durchgeführt werden müssen. Dies betrifft den Fall, in dem die Daten durch wiederholte Schwerpunktprüfungen erhoben werden und die Variation in den Indikatoren durch eine weitere regional zugeschnittene Schwerpunktprüfung erzeugt werden soll. Ein solches Vorgehen weicht deutlich von den üblichen Verfahrensweisen innerhalb der FKS ab und könnte daher bei den betroffenen Mitarbeitern auf erhebliche Vorbehalte stoßen.

Eine Schwierigkeit bei der Kommunikationsstrategie besteht darin, dass ihre Adressaten Beamte an allen Standorten der FKS sind. Das bedeutet, dass die Kommunikation im persönlichen Kontakt einen erheblichen Aufwand verursacht. Andererseits sind die Kommunikationserfordernisse nur bei den Teilnehmerstandorten besonders hoch. Daher bietet sich die Durchführung einer Informationsveranstaltung an, bei der möglichst die Leiter des Arbeitsbereichs Prävention der Teilnehmerstandorte, also 50 bis 60 Personen, vertreten sein sollten. Die weitere Kommunikation könnte dann auch schriftlichem Wege erfolgen.

4.1.6 Implementationsanalyse

Begleitet werden sollte das Projekt durch Befragungen von beteiligten Mitarbeitern der FKS an den einzelnen Standorten. Dabei sollte überprüft werden, ob die Vorgaben der Evaluation auf lokaler Ebene auch umgesetzt werden. Eine weitere Frage ist, ob die verstärkten Prüfungen von der Zielgruppe auch wahrgenommen werden. Hierbei sollten sowohl mit einer kleinen Zahl potentieller Anbieter oder Nachfrager von Schwarzarbeit in der betroffenen Branche Interviews geführt werden (jeweils ca. 10 bis 20 Interviews).

4.1.7 Kosten-Nutzen-Analyse

Da legale Beschäftigung als Indikator zur Erfassung der verhaltenslenkenden Effekte ungeeignet erscheint, wird sich die Kosten-Nutzen-Analyse auf die fiskalischen Effekte einer Verminderung des Volumens der geleisteten Schwarzarbeit beschränken.

Die Kosten der Schaffung eines regionalen Prüfklimas sind aus der Kosten- und Leistungsrechnung zu bestimmen. Hier geht es im Wesentlichen darum, den zusätzlichen Ressourcenaufwand in Form eines erhöhten Personaleinsatzes möglichst exakt zu bestimmen. Da die Variation auf der Ebene der Standorte erzeugt wird und zur Abschätzung der zusätzlichen Kosten keine Informationen über den Einsatz einzelner Mitarbeiter erforderlich sind, sind hierbei keine grundsätzlichen Probleme zu erwarten. Um die Kosten des Maßeinsatzes zu bestimmen, müssten lediglich die tatsächlichen Kosten für jeden Teilnehmerstandort mit den Kosten verglichen werden, die entstanden wären, wenn der betreffende Standort nicht Teilnehmerstandort gewesen wäre. Diese zusätzlichen Kosten lassen sich am einfachsten bestimmen, indem man von den tatsächlichen Kosten diejenigen Kosten abzieht, die in einem geeigneten Vergleichszeitraum tatsächlich entstanden sind. Im einfachsten Fall kann die Zeit unmittelbar vor Durchführung der Maßnahme für diesen Vergleich herangezogen werden.

In einigen Fällen wird sich dieses Vorgehen allerdings verbieten, insbesondere, wenn die Kosten bereits durch die Vorbereitung der Maßnahme beeinflusst werden oder falls die Kosten saisonalen Schwankungen unterliegen. In diesen Fällen könnte man die Kosten als Vergleichspunkt heranziehen, die im Vorjahr in vergleichbarer Zeit entstanden sind. Selbstverständlich sollte in diesen Fällen für allgemeine Kostensteigerungen (z.B. Lohn- und Gehaltssteigerungen) sowie für die Veränderung der allgemeinen Preisniveaus kontrolliert werden. Ebenso ist zu berücksichtigen, dass sich eventuell die Anzahl der Mitarbeiter über den betrachteten Zeitraum verändert hat.

Eventuell sind neben den Personalkosten auch weitere Kosten zu berücksichtigen, die sich beispielsweise aus der verstärkten Nutzung von Fahrzeugen und anderen Ressourcen ergeben. Selbstverständlich sind auch Kosten zu berücksichtigen, die entstehen, wenn das regional erzeugte Prüfklima durch Maßnahmen der Presse- und Öffentlichkeitsarbeit unterstützt werden. Hier ist beispielsweise an Plakatierungsaktionen oder auch an die Produktion eines Handzettels zu denken, der an die regional ansässigen Unternehmen verteilt bzw. verschickt wird.

Falls in dem Bündel der Instrumente, deren Einsatz evaluiert werden soll, auch Maßnahmen der Presse- und Öffentlichkeitsarbeit enthalten sind, muss bei der Durchführung einer Kosten-Nutzen-Analyse ein weiteres Problem gelöst werden. Dieses besteht darin, dass Kostenbestandteile die Effektivität des Maßnahmebündels beeinflussen, die nicht vollständig durch die Kosten der Maßnahmen selbst erfasst werden. Es müssen daher geeignete Verfahren für die Zurechnung solcher Kosten ermittelt und angewandt werden. Unter den zu berücksichtigenden Kosten sind alle Kosten zu verstehen, die bei der Prüfung und Ermittlung sowie der Ahndung entstehen, auf die in der Presse- und Öffentlichkeitsarbeit hingewiesen wird.

Da die Kosten-Nutzen-Analyse nicht den Nutzen des Evaluationsprojektes, sondern den Nutzen der Tätigkeit der FKS bzw. der Ausdehnung der Tätigkeit der FKS um ein bestimmtes Maß bestimmen soll, sind Kosten, die durch die Steuerung und Administration des Evaluationsprojektes entstehen, nicht zu berücksichtigen. Darunter fallen insbesondere Kosten, die durch die Planung des Projektes und die zentrale Steuerung durch die Abteilung FKS bei der OFD Köln entstehen.

Der Nutzen einer Veränderung der Rechtsbewusstseins ist generell schwierig zu bewerten, da sich die fiskalischen und sonstigen Auswirkungen vor allem langfristig bemerkbar machen dürften.

Im Folgenden wird davon ausgegangen, dass der Umfang der Schwarzarbeit als Indikator verwendet wird. Das Projekt gibt dann Aufschluss über den Effekt, den die Erzeugung eines regionalen Prüfklimas auf den Umfang der Schwarzarbeit hat. Es ist zu erwarten, dass damit präzise Aussagen über die Verminderung von Schwarzarbeit durch die Tätigkeit der FKS getroffen werden können. Entsprechend sind auch verlässliche Aussagen über den Nutzen der Maßnahme im Hinblick auf eine Reduktion des Umfangs von zu unrecht bezogenen Sozialleistungen zu erwarten. Um die Abschätzung des Nutzens zu verbessern, sollte bei der Erhebung der Daten erfasst werden, wie viele der illegal beschäftigten Individuen welche Sozialleistungen bezogen haben. Falls es nicht möglich sein sollte, diese genauer zu erfassen, ist auch die Erfassung eines Indikators für den Leistungsbezug (Ja/Nein) zur Präzisierung der Schätzungen sinnvoll.

Da das Projekt ein repräsentatives Bild der Schwarzarbeit in der betreffenden Branche zeichnet, kann der Nutzen der Maßnahme, der durch die Verminderung von zu unrecht bezogenen Sozialleistungen entsteht, auf Basis der erhobenen Daten quantifiziert werden. Dazu wird die Verminderung des Umfangs der Schwarzarbeit, gemessen als Anzahl von in Schwarzarbeit tätigen Personen, ermittelt. Idealerweise werden dabei Vollzeitäquivalente bestimmt, d.h. es wird die Anzahl von Personen bestimmt, die bei Vollzeitbeschäftigung in Schwarzarbeit nötig gewesen wäre, die geleistete Schwarzarbeit zu verrichten. Anschließend kann der Umfang der zu unrecht bezogenen Sozialleistungen abgeschätzt werden. Dies geschieht entweder auf Basis der im Zuge der Maßnahme erhobenen Daten über den Umfang der bezogenen Sozialleistungen, oder es werden die Vollzeitäquivalente benutzt, um mit Hilfe der Indikatorvariablen für den Leistungsbezug und unter Ansetzung von Durchschnittsätzen den Umfang des unrechtmäßigen Leistungsbezugs abzuschätzen. Dabei muss angesichts der beschränkt vorhandenen Information in jedem Fall mit der Annahme gearbeitet werden, dass die Personen, die wegen Schwarzarbeit geahndet werden, keinerlei Anspruch auf Sozialleistungen haben. Dies bedeutet, dass der gesamte beobachtete bzw. geschätzte Sozialleistungsbezug unrechtmäßig ist.

Bei der Bestimmung des Nutzens der Maßnahme sollten neben den zu unrecht bezogenen Sozialleistungen das zusätzlich generierte Aufkommen von Beiträgen zu den Sozialversicherungssystemen und das zusätzlich generierte Steueraufkommen berücksichtigt werden. Hierbei ergibt sich das Problem, dass die geschätzten kausalen Effekte die Schwarzarbeit betreffen und keine präzisen Aussagen darüber getroffen werden können, welchen Effekt die

Maßnahme auf die legale Beschäftigung hat. Da ein Großteil des Nutzens der Maßnahme aus zusätzlicher legaler Beschäftigung entstehen dürfte, ist die Kosten-Nutzen-Analyse mit Unsicherheiten belastet, die durch das Design des Evaluationsprojektes nicht vollkommen zu beseitigen sein dürften.

Da verlässliche Indikatoren für die legale Beschäftigung in diesem Projekt nicht zu ermitteln sind, müssen hier notgedrungen Schätzwerte angesetzt werden. Das zusätzlich generierte Aufkommen von Beiträgen zu den Sozialversicherungssystemen lässt sich abschätzen, indem man die Verminderung im Umfang der Schwarzarbeit in Vollzeitäquivalenten mit dem branchenüblichen regionalen Durchschnitt der Sozialversicherungsbeiträge multipliziert. Die durchschnittlichen Sozialversicherungsbeiträge können mit Hilfe des branchenüblichen regionalen sozialversicherungspflichtigen Einkommens bestimmt werden. Eine Schätzung des zusätzlichen Steueraufkommens kann in ähnlicher Weise erfolgen. Dazu kann das Aufkommen für die Einkommensteuer (einschließlich Solidaritätszuschlag) für den durchschnittlichen Beschäftigten in der betroffenen Branche durch Multiplikation des durchschnittlichen steuerpflichtigen Einkommens mit einem Durchschnittssteuersatz abgeschätzt werden. Das gesamte durch die Maßnahme generierte zusätzliche Aufkommen direkter Steuern wird zwar auch Anteile von Gewerbe- und Körperschaftssteuer enthalten, aber das Gros der Steueraufkommenseffekte sollte sich in der Einkommensteuer niederschlagen, so dass aufwändigere Schätzverfahren kaum einen Mehrwert haben dürften.

Die Bestimmung des zusätzlich generierten Aufkommens zu den Sozialversicherungssystemen und des zusätzlich generierten Steueraufkommens beruht auf der Annahme, dass die Verminderung des Umfangs von Schwarzarbeit in gleichem Umfang sozialversicherungspflichtige Beschäftigung schafft. Darüber hinaus wird unterstellt, dass die Charakteristika, insbesondere die Entlohnung, der zusätzlichen legalen Beschäftigung in etwa denjenigen des durchschnittlichen Beschäftigten in der betreffenden Branche entspricht. Dies sind Annahmen, die die Aussagekraft der Kosten-Nutzen-Analyse erheblich einschränken. Insbesondere ist zu beachten, dass die Annahmen tendenziell zu einer Überschätzung des Nutzens der Maßnahme führen. Daher sollte hier mit alternativen Szenarien gearbeitet werden, die die Sensitivität der Ergebnisse der Kosten-Nutzen-Analyse hinsichtlich alternativer Annahme über die Substitution von regulärer Arbeit und Schwarzarbeit erfassen.

4.1.8 Aufwand

Es entsteht erheblicher Aufwand bei dem zu evaluierenden Instrumenteneinsatz. Um den Aufwand begrenzt zu halten, sollten die zusätzlichen Prüfungen innerhalb eines eng umgrenzten Bereichs innerhalb einer Branche durchgeführt werden. Das bedeutet zum Beispiel, dass man die Maßnahmen beispielsweise nicht im gesamten Transportgewerbe durchführt, sondern lediglich bei den Taxi- oder bei den Kurierfahrern, oder nicht im gesamten Gastgewerbe, sondern nur im Bereich der Imbissgastronomie. Falls begleitend zu den zusätzlichen Prüfungen Maßnahmen der Öffentlichkeitsarbeit oder Pressearbeit durchgeführt werden, fällt hierfür weiterer Aufwand an.

Es entsteht ferner erheblicher Aufwand bei der Erhebung der Daten durch Schwerpunktprüfungen oder durch Befragungen zum Rechtsbewusstsein. Wenn in letzterem Fall in zwei Wellen aus jedem Standort jeweils 10 Unternehmen oder Personen befragt werden, wären zweimal 1130 Interviews notwendig.

4.1.9 Besondere Probleme bei der Durchführung des Projekts

Der Ausstrahlungsbereich der Maßnahme dürfte in vielen Fällen nicht randscharf sein, z.B. können auch Anbieter oder Nachfrager aus benachbarten Nichtteilnehmer-Standorten von der Veränderung des regionalen Prüfklimas beeinflusst werden. Gerade beim Zusammenwirken von mehreren Maßnahmen ergibt sich dieses Problem, da die Wahrscheinlichkeit, dass mindestens eine Maßnahme über die Teilnehmerstandorte hinaus ausstrahlt, größer wird. Insbesondere können Maßnahmen der Presse- und teilweise auch der Öffentlichkeitsarbeit nicht zielgenau auf einen bestimmten Kreis von Adressaten zugeschnitten werden.

Das Problem der Beeinflussung von Arbeitsmarktbeteiligten außerhalb des jeweiligen Standortes stellt sich insbesondere bei überregional tätigen Firmen oder in nichtstationären Gewerben wie dem Transportgewerbe. Daher ist die Evaluation in stationären Branchen mit kleinen, lokal arbeitenden Unternehmen besonders sinnvoll, wie z.B. im Gaststättengewerbe oder im Handwerksbereich.

Weiterhin stellt sich das Problem, dass in den ausgewählten Regionen drei branchenspezifische Schwerpunktprüfungen innerhalb kurzer Zeit durchgeführt werden müssten. Hierbei sind schwerwiegende Akzeptanzprobleme bei den Mitarbeitern zu erwarten. Eventuell ist eine Kompensation der betroffenen Standorte durch verringerte Anforderungen bei normalen Schwerpunktprüfungen möglich (siehe Abschnitt 2.3.1). Falls auch Maßnahmen der Öffentlichkeitsarbeit (z.B. Plakatierungen) berücksichtigt werden sollen, stellt sich zudem die Frage, ob und inwieweit die regionale Variation in den Maßnahmen innerhalb der FKS und bei den Arbeitsmarktbeteiligten bemerkt wird. Gegebenfalls müssen die Unterschiede zwischen Standorten durch eine Kommunikationsstrategie begründet werden.

Bei Einsatz eines Maßnahmebündels erfasst man den gemeinsamen Effekt der Maßnahmen. Dabei würde die Trennschärfe im Hinblick auf die Evaluation eines bestimmten Maßnahme verloren gehen, d.h. es wäre nicht möglich, die Effekte der einzelnen Maßnahmen gesondert zu evaluieren. Der Vorteil der Durchführung von Maßnahmenbündeln ist dagegen die vermutlich höhere Wirksamkeit, die es erleichtert, Effekte der Tätigkeit der FKS auf das Verhalten der Arbeitsmarktbeteiligten zu identifizieren. Falls eine solche Evaluation eines Maßnahmebündels wegen der fehlenden Trennschärfe im Hinblick auf einzelne Instrumente nicht gewünscht wird, müsste man z.B. Prüfungen und Maßnahmen der Presse- und Öffentlichkeitsarbeit getrennt evaluieren.

4.1.10 Mögliche Varianten

Als Alternative zur Erzeugung eines regionalen Prüfklimas könnten auch die Wirkungen einer regional differenzierten Pressearbeit geprüft werden. Dabei bliebe das Niveau der tatsächlich

durchgeführten Prüfungen unverändert. Diese Möglichkeit ist jedoch aufgrund des indirekten Wirkungszusammenhangs zwischen der Pressearbeit der FKS und den in Frage kommenden Indikatoren problematisch. Hierbei stellt sich speziell das Problem, dass Presseberichte in Lokalzeitungen unter Umständen auch eine weitere Verbreitung in der Zielgruppe haben, als es dem unmittelbaren Einzugsbereich des Mediums entspricht. Insbesondere wird es zu dem Problem kommen, dass Berichte in bestimmten Medien von anderen Medien aufgegriffen werden und damit der Kreis der „Empfänger“ in unkontrollierbarer Weise ausgedehnt wird. Zwar ließe sich durch eine Medienbeobachtung im Nachhinein zumindest für die Printmedien überprüfen, welche Medien in welcher Intensität über die FKS berichtet haben, aber das Evaluationsprojekt wäre von einer erheblichen Unsicherheit belastet, ob die erzeugte Variation bei der Presseaktivität für eine Überprüfung der Effektivität der FKS hinreichend ist. Dessen ungeachtet entspricht der Verbreitungskreis der lokalen Medien nicht den Zuständigkeitsbereichen der Standorte der FKS. Es kann also nicht unmittelbar kontrolliert werden, wie stark in der Presse tatsächlich berichtet wird.

Ein weiteres Problem besteht darin, dass die bisherige Pressearbeit der FKS eher damit beschäftigt war, im Nachgang von Ermittlungen öffentliche Aufmerksamkeit herzustellen und durch die Darstellung von Ermittlungserfolgen präventiv zu wirken. Würde man zu Evaluationszwecken dagegen eine regional differenzierte Pressekampagne durchführen, die unabhängig von Ermittlungserfolgen an den Maßnahmestandorten wäre, käme es hier zum Einsatz eines in dieser Weise noch nicht erprobten Instrumentes. Daher würde man über die Wirksamkeit der gegenwärtigen Instrumente der FKS kaum etwas lernen können.

4.1.11 Erwartete Ergebnisse des Evaluationsprojekts

- Bei Verwendung des Indikators Schwarzarbeit kann ein repräsentatives Bild des Umfangs der Schwarzarbeit im geprüften Bereich gezeichnet werden. Der Umfang der Schwarzarbeit wird dabei nicht allein an den Verdachtsfällen, sondern auch an den Fällen der Ahndung sichtbar. Hierbei werden auch regionale, betriebsgrößenspezifische und andere Unterschiede in der Intensität der Schwarzarbeit sichtbar. Diese deskriptiven Auswertungen wären ein wichtiges zu dokumentierendes Zwischenergebnis, das die deskriptive Auswertung von ProFiS (siehe Abschnitt 3.1) ergänzen könnte.
- Bei Verwendung des Indikators Rechtsbewusstsein kann ein repräsentatives Bild der Einstellungen der Arbeitsmarktbeteiligten im geprüften Bereich gewonnen werden. Auch hier lässt sich der deskriptive Befund nach Eigenschaften des oder der Befragten aufgliedern.
- Das Projekt gibt Aufschluss über den kausalen Effekt der Gesamtheit der durchgeführten Maßnahmen auf den Umfang der Schwarzarbeit bzw. das Rechtsbewusstsein der Arbeitsmarktbeteiligten. Hieraus wird erkennbar, in welchem Umfang die in die Evaluation einbezogenen Maßnahmen der FKS in ihrer Gesamtheit den gewählten Indikator beeinflussen.
- Eine Gegenüberstellung der Kosten der Durchführung der Maßnahmen und des vermiedenen Umfangs der Schwarzarbeit (falls der gewählte Indikator der Umfang der Schwarzarbeit ist) im Sinne einer Kosten-Nutzen-Analyse gibt Aufschluss darüber, ob der Einsatz der Instrumente der FKS einen positiven fiskalischen Nettoeffekt hat.

4.2 Ankündigung von Prüfungen

4.2.1 Fragestellung

Nachfrager von Schwarzarbeit werden angeschrieben, wobei eine verstärkte Aktivität von Prüfungen (z.B. von Buchprüfungen oder Vor-Ort-Prüfungen) in einem bestimmten Zeitraum angekündigt wird. Von Interesse ist der präventive Effekt dieser Ankündigung. Gegen Ende dieses Zeitraumes werden die Prüfungen tatsächlich in angemessenem Umfang vorgenommen.

Beispiel für ein Anschreiben

Sehr geehrter [Firmeninhaber],

die Finanzkontrolle Schwarzarbeit bei der Zollverwaltung hat den gesetzlichen Auftrag, die Schwarzarbeit zu bekämpfen. Wir möchten Ihnen mitteilen, dass im Zeitraum vom TT.TT.TTTT bis zum TT.TT.TTTT in Ihrer Branche und Region verstärkt Prüfungen durchgeführt werden. Dazu haben wir unsere sonst in diesem Bereich vorgehaltenen Kapazitäten verdoppelt. [Die Art der Prüfungen, z.B. Buchprüfungen, könnte an dieser Stelle näher beschrieben werden.]

Diese Prüfungen sind im Interesse aller ehrlichen Arbeitgeber und Arbeitnehmer, da Schwarzarbeit und illegale Beschäftigung auf Kosten legaler Beschäftigung gehen. Wir bitten Sie daher um Verständnis, wenn Ihnen durch eine Prüfung Aufwand entsteht.

Im Falle einer Prüfung in Ihrem Unternehmen bitten wir Sie um Ihre Mithilfe. [Es könnte hier eine Auflistung von Verhaltensmaßregeln folgen sowie eine Belehrung über die Rechte und Pflichten des Arbeitgebers, die im Umgang mit den Angehörigen der FKS gelten.]

Falls Sie Rückfragen zu diesen Informationen haben, können Sie uns unter obiger Adresse erreichen.

Mit freundlichen Grüßen

Das Anschreiben soll die Wahrscheinlichkeit einer Prüfung in der Einschätzung der Unternehmen mittelfristig um ein bestimmtes Maß erhöhen, es soll aber nicht Prüfungen mit Sicherheit ankündigen. Falls Prüfungen (etwa Buchprüfungen) kurzfristig angekündigt werden, ist nicht dieselbe Verhaltensänderung zu erwarten wie im Fall einer mittelfristigen allgemeinen Ankündigung einer intensivierten Prüfungstätigkeit. Letztere kann Verhaltensänderungen auslösen, während die Ankündigung unmittelbar im Vorfeld einer Buchprüfung allenfalls „Verschleierungsaktivitäten“ bewirkt. Beides muss voneinander getrennt werden, um die prä-

ventive Wirkung von Bekämpfungsmaßnahmen identifizieren zu können. Die Fristigkeit der Betrachtung ist ein grundsätzliches Problem, das in diesem Projekt besonders wichtig ist. Lösen Bekämpfungsmaßnahmen z.B. lediglich eine zeitliche (oder räumliche) Verlagerung von Aktivitäten aus, oder führen sie zu längerfristigen Änderungen im Verhalten?

Je präziser das Maß, in dem die subjektive Aufdeckungswahrscheinlichkeit erhöht wird, desto besser kann die Verhaltensreaktion quantifiziert werden. Daher ist zu prüfen, ob im Anschreiben die Erhöhung der Aufdeckungswahrscheinlichkeit um ein bestimmtes Maß angekündigt werden kann oder der zusätzlich Input seitens der FKS präzisiert werden kann (beispielsweise durch die Ankündigung einer Verdoppelung des Einsatzes der Maßnahmen oder ähnliches).

Die Evaluation könnte auf bestimmte Branchen oder auf bestimmte Arten von Schwarzarbeit oder illegaler Beschäftigung, wie z.B. illegale Ausländerbeschäftigung oder illegale Arbeitnehmerüberlassung, beschränkt werden.

Tabelle 4.2: Zeitlicher Ablauf des Evaluationsprojekts „Ankündigungen“

Schritt 1 (ca. Monat 1-4)	Ziehung der Stichprobe gemäß einem Branchenregister oder ähnlichen Informationen. Auswahl der Betriebe, an die Ankündigungen versendet werden, durch einen Zufallsprozess oder ein anderes geeignetes Verfahren (siehe Abschnitt 4.2.2). Versand der Anschreiben. Durchführung der Kommunikationsstrategie an den beteiligten Standorten.
Schritt 2 (ca. Monat 5-7)	Durchführung der Prüfungen bei den Teilnehmer- und den Kontrollbetrieben an ca. 10 bis 20 Standorten. Durchführung der Implementationsanalyse.
Schritt 3 (ca. Monat 8-24 oder später)	ggf. Nachverfolgung der Verdachtsfälle in der Ahndung
Schritt 4 (3 Monate im Anschluss an Schritt 2 bzw. Schritt 3)	Datenauswertung Kosten-Nutzen-Analyse

4.2.2 Art der Variation

Die Variation entsteht auf der Ebene der Unternehmen. Die Stichprobe der Teilnehmer und Nichtteilnehmer wird durch ein Branchenregister oder ähnliche Unterlagen aus der Grundgesamtheit aller Betriebe gezogen. Die angeschriebenen Unternehmen bilden die Teilnehmergruppe, während die nicht angeschriebenen Unternehmen als Kontrollgruppe fungieren. Der Teilnehmerstatus wird im Idealfall durch einen Zufallsprozess festgelegt.

Ein Anschreiben, das verstärkte Prüfungen nur gegenüber einigen Unternehmen ankündigt, könnte allerdings für die Akzeptanz der FKS und die Durchführung der Evaluation problema-

tisch sein (siehe Abschnitt 4.2.9). Dies wäre insbesondere dann der Fall, wenn die Akteure in den betroffenen Branchen häufig miteinander interagieren, etwa auf der Ebene von Verbänden, und die angekündigten Maßnahmen sich auch in der Kontrollgruppe herumsprechen würden. In diesem Fall müsste man eine Variable finden, die die Zuordnung zur Teilnehmer- und Kontrollgruppe glaubwürdig macht und rechtfertigt, ohne für das Ausmaß der Schwarzarbeit bestimmend zu sein. Auf diese Weise würde eine Instrumentvariable (siehe den Anhang über die Evaluationsmethoden) erzeugt, die mit dem Teilnehmerstatus, nicht aber mit dem Niveau der Schwarzarbeit korreliert ist. Allerdings wird so zusätzlich Variation erzeugt, die die Ergebnisse weniger trennscharf werden lässt als bei einer reinen Zufallsauswahl.

Beispiele für solche Instrumentvariablen:

- In der Stichprobe befinden sich nur Firmen von 10 bis 20 Beschäftigten, es werden nur Firmen ab 15 Beschäftigten angeschrieben. Hierbei ist darauf zu achten, ob Schwellenwertregelungen in anderen Bereichen (Kündigungsschutz, Schwerbehindertengesetz) nicht zu unterschiedlichem Verhalten in den beiden Gruppen führen.
- Es werden nur Firmen mit Namen geprüft, die mit A-K beginnen. Auch dies ist problematisch, wenn die Namen der Firmen in systematischer Weise verteilt sind (Beispiel: alle Pizzerien beginnen mit „P“).
- Es werden nur Betriebe geprüft, die sich in einer bestimmten geographischen Situation befinden (z.B. NO vom Standort, nicht SW).

Eine weitere Möglichkeit der Begründung gegenüber dem Adressatenkreis wäre es, die partielle Abdeckung der betroffenen Branchen oder des betroffenen Bereiches mit Kapazitätsgründen zu rechtfertigen.

In dem Maße, in dem sich diese Möglichkeiten als nicht durchführbar erweisen, ist es notwendig, stärker auf die Variation zwischen den Standorten zu setzen. Unabhängig davon sollten in der Implementationsanalyse auch Nichtteilnehmerbetriebe befragt werden, ob sie von den Ankündigungen von Prüfungen gehört haben.

4.2.3 Verwendbare Indikatoren

Indikator wäre hier die tatsächliche Nutzung von Schwarzarbeit gemäß der Ergebnisse der Prüfung bzw. der Weiterverfolgung durch die Ahndung. Wie im Projekt 4.1 könnte dabei nicht allein die Zahl, sondern auch die Schwere der Vergehen durch die Nachverfolgung in der Ahndung erfasst werden.

4.2.4 Erfassung der Indikatoren in den Daten

Sinnvollstes Vorgehen bei der Erhebung der Indikatoren ist eine Schwerpunktprüfung, wobei sichergestellt werden muss, dass die vorher festgelegte Teilnehmer- und Kontrollgruppe möglichst vollständig geprüft wird. Die Auswahl der Prüfobjekte darf insofern nicht dem Standort überlassen werden, was andererseits Akzeptanzprobleme hervorrufen kann (siehe Abschnitt 4.2.5). Bei den vorhandenen bzw. durch die Tätigkeit der FKS angelegten und in ProFiS enthaltenen Datenbeständen ist dagegen nicht davon auszugehen, dass ein großer Teil der Teilnehmer- und Kontrollgruppe erfasst wird.

Teilnehmer- und Nichtteilnehmerstatus sowie Prüfungsergebnis sollten mit den Informationen in ProFiS verknüpft werden, um mit der Ahndung kombiniert werden zu können. Wegen der viel größeren Zahl an Beobachtungen durch die Variation auf der Unternehmensebene reicht hier im Gegensatz zum Projekt 4.1 eine einmalige Prüfung. Die Evaluation braucht nicht an allen Standorten der FKS durchgeführt zu werden, sondern es würde ausreichen, wenn ca. 10 bis 20 Standorte beteiligt wären. Diese Zahl ergibt sich daraus, dass Zielsetzung und Durchführung des Evaluationsvorhabens im persönlichen Kontakt zwischen der Abteilung FKS und dem Evaluator auf der einen und den beteiligten Standorten auf der anderen Seite kommuniziert werden müssen (siehe Abschnitt 4.2.5).

4.2.5 Kommunikationsstrategie

Die Tatsache, dass Teilnehmer- und Kontrollgruppe möglichst vollständig durch die Schwerpunktprüfung erfasst werden sollen, muss begründet werden, weil die Auswahl der Prüfobjekte üblicherweise den Mitarbeitern an den Standorten obliegt.

Auch die Versendung der Anschreiben lediglich an bestimmte Betriebe muss intern kommuniziert werden, damit nicht Anschreiben an die Kontrollgruppe gelangen bzw. Betriebe der Teilnehmergruppe nicht angeschrieben werden. Der Standort hat sonst den Anreiz, die Anschreiben gerade an die Betriebe zu senden, bei denen die Wahrscheinlichkeit der Nutzung von Schwarzarbeit als hoch eingeschätzt wird. Eventuell könnte die Versendung der Anschreiben auch zentral durchgeführt werden, so dass sich hier eine Kommunikationsstrategie erübrigt.

Günstig ist die Tatsache, dass die Evaluation nicht an allen Standorten durchgeführt werden muss. Man könnte daher Standorte ausklammern, an denen eine besondere Belastung durch andere Aufgaben vorliegt. Hinzu kommen mögliche Teilnahmeanreize (siehe Abschnitt 3.2.1).

Zur Durchführung erscheint insbesondere die Organisation von ein oder zwei internen Workshops sinnvoll zu sein. Auf dem ersten Workshop würde die Zielsetzung der Evaluation dargestellt. Zugleich besteht hier die Möglichkeit, dass die Teilnehmer Verbesserungsvorschläge für die Durchführung der Evaluation machen. Im Zuge der Durchführung der Maßnahme könnte ein zweiter Workshop durchgeführt werden, der Probleme bei der Umsetzung aufdecken könnte. Dieser Workshop dient insofern nicht allein der Kommunikation an die örtlichen Dienststellen, sondern ist zugleich Teil der Implementationsanalyse.

4.2.6 Implementationsanalyse

Die Implementationsanalyse hätte zu prüfen, ob die Vorgaben bei der Versendung der Anschreiben und bei der anschließenden Prüfung erfüllt werden. Dazu würden die Mitarbeiter an einigen Standorten durch den Evaluator befragt werden. Ein Workshop der FKS im Zeitraum der Durchführung würde ebenfalls Informationen liefern, wobei darauf zu achten ist, dass Informationen auch vor Ort bei den direkt im Einsatz stehenden Beamten eingeholt werden. Eine Reihe von angeschriebenen Unternehmen würde zudem nach ihrer Einschät-

zung des Anschreibens befragt werden: Nehmen sie dieses ernst und erhöht dieses das subjektiv wahrgenommene Risiko, von einer Prüfung erfasst zu werden?

4.2.7 Kosten-Nutzen-Analyse

Die Ankündigung von Prüfungen in einer bestimmten Intensität ist langfristig nur dann glaubhaft, wenn die Arbeitsmarktbeteiligten anhand objektiver Informationen eine Übereinstimmung von angekündigter und tatsächlicher Prüfungsintensität feststellen. Daher muss eine aussagekräftige Kosten-Nutzen-Analyse die Kosten so ermitteln, dass sie den Aufwand abbilden, der notwendig ist, um das von den Arbeitsmarktbeteiligten empfundene Risiko einer Prüfung langfristig zu beeinflussen. Ausgangspunkt der Kosten-Nutzen-Analyse ist also die Annahme, dass die Arbeitsmarktbeteiligten bei unveränderten Rahmenbedingungen das Risiko, Gegenstand einer Prüfung zu werden, realistisch einschätzen. Das Evaluationsprojekt simuliert gewissermaßen eine tatsächliche Erhöhung der Prüfungsintensität, indem zusätzliche Prüfungen angekündigt werden. Die Kosten-Nutzen-Analyse versucht also, die Kosten einer Erhöhung der Prüfungsintensität zu erfassen, die in ihrer Wirkung auf die Arbeitsmarktbeteiligten der durch angekündigte Prüfungen verursachten Verhaltensänderung gerade entspricht. Hier liegen naturgemäß auch die Hauptschwierigkeiten einer aussagekräftigen Kosten-Nutzen-Analyse.

Zunächst muss eine möglichst präzise Abschätzung für die von den Arbeitsmarktbeteiligten wahrgenommene Veränderung im Prüfungsrisiko vorgenommen werden. Dies kann im Wesentlichen auf zwei Wegen geschehen. Erstens kann das Anschreiben, mit dem die zusätzlichen Prüfungen angekündigt werden, eine exakte Angabe über die Wahrscheinlichkeit machen, mit der ein gegebenes Unternehmen geprüft wird. Dies könnte z.B. so geschehen, dass die Häufigkeit von Prüfungen in der betreffenden Branche zum Status quo mitgeteilt wird. Dann könnte ein Hinweis erfolgen, in welchem Ausmaß die Prüfungen intensiviert werden. Unter der Annahme, dass die Unternehmen die Information über die Prüfungsintensität aufnehmen und für glaubhaft halten, ist die kommunizierte Veränderung in der Prüfungsintensität damit bestimmt. Der zweite Weg führt über Befragungen der Unternehmen. Solche Befragungen könnten in Rahmen der Implementationsanalyse vorgenommen werden, oder sie sind zusätzlich durchzuführen. Eine Kombination beider Methoden führt vermutlich zu den besten Ergebnissen in dem Sinne, dass die wahrgenommene Veränderung in der Prüfungsintensität am besten abgeschätzt werden kann. Dabei würde das Anschreiben in der beschriebenen Weise konzipiert werden, und eine nachfolgende Befragung von Unternehmen aus der Teilnehmergruppe würde Aufschluss darüber geben, ob die im Anschreiben enthaltene Information über die Intensivierung der Tätigkeit der FKS von den Unternehmen wie geplant aufgenommen wurde.

Wenn die Information über die Veränderung des von den Arbeitsmarktbeteiligten wahrgenommenen Prüfungsrisikos vorliegt, können die Kosten bestimmt werden, die mit der entsprechenden Veränderung der tatsächlichen Prüfungsintensität einhergehen würden. Dies sind, wie oben ausgeführt, die für die Kosten-Nutzen-Analyse relevanten Kosten. Zur Bestimmung dieser Kosten ist wieder Rückgriff auf die Kosten- und Leistungsrechnung der FKS zu nehmen. Zunächst muss eine Abschätzung vorgenommen werden, wie viel zusätzliches

Personal eingesetzt werden müsste, um die Prüfungsintensität in der beabsichtigten Weise zu erhöhen. Dabei kann vereinfachend davon ausgegangen werden, dass die Produktivität des einzelnen Mitarbeiters bei einer Vergrößerung der Standorte und einer Intensivierung der Tätigkeit der FKS konstant bleibt. Bei moderaten Veränderungen in der Größe der Standorte sollte diese Annahme insgesamt nicht unrealistisch sein. Auf dieser Basis können die zusätzlichen Kosten (reine Personalkosten, Kosten für zusätzliche Bürofläche, Fahrzeuge usw.) relativ präzise bestimmt werden.

Bei der Bestimmung des Nutzens der Intensivierung des Maßnahmeinsatzes werden die Ergebnisse der Evaluation verwendet. Ähnlich wie im Abschnitt zur Kosten-Nutzen-Analyse im Projekt 4.1 beschrieben geht es darum, den fiskalischen Nutzen der Verminderung der Schwarzarbeit möglichst präzise abzuschätzen. Dabei sollten der verminderte Leistungsmissbrauch und zusätzlich zu erwartende Einnahmen bei der Sozialversicherung und bei der Einkommensteuer berücksichtigt werden. Das Vorgehen ist im Wesentlichen wie in Projekt 4.1 und wird daher hier nicht noch einmal detailliert erläutert.

4.2.8 Aufwand

In Teilnehmer- und Kontrollgruppe sollten so viele Betriebe enthalten sein, dass in der Kontrollgruppe eine Anzahl von Beanstandungen bzw. Ahndungen zu erwarten ist, die bei Vorliegen eines Effekts der Maßnahme auch ein signifikantes Schätzergebnis erwarten lässt. Im Einschub auf der folgenden Seite wird hierfür eine Modellrechnung durchgeführt.

Da das Verhältnis von Ahndungen zu Prüfungen von der jeweiligen Branche abhängt, dürfte der Aufwand bei der Erhebung stark zwischen Branchen variieren. So zeigt die Dokumentation der Ergebnisse 2004 der FKS vom Februar 2005 (FKS, 2005), dass das Verhältnis von Befragungen zu Verdachtsfällen in den sechs Branchen, in denen Schwerpunktprüfungen durchgeführt wurden (Spielhallen, Großbaustellen, Güterkraftverkehr, Kurier- Express- und Paketdienste, Garten- und Landschaftsbau sowie Weihnachtsmärkte) zwischen 9% und 20% lag. Diese Zahlen können jedoch nicht direkt auf die hier diskutierten Prüfungen übertragen werden, da diese – anders als im normalen Arbeitsalltag der FKS -- nicht nach einer Risikoeinschätzung der Beamten vorgenommen werden.

Bei der Auswahl von Teilnehmer- und Kontrollgruppe und der Versendung der Anschreiben fällt Aufwand an, der jedoch verhältnismäßig gering sein dürfte. Bei der Prüfung fällt an den Standorten, in denen die Evaluation durchgeführt wird, zusätzlicher Aufwand in erheblichem Ausmaß an, vor allem bei der möglichst vollständigen Erfassung von Teilnehmer- und Kontrollgruppe. Dies setzt voraus, dass die Ziele der Evaluation an die beteiligten Standorte kommuniziert werden.

Verglichen mit dem Projekt „Prüfklima“ handelt es sich hierbei dennoch um ein relativ kostengünstiges Projekt, da die Evaluation nur an einigen Standorten stattfindet, die Prüfungen nur einmal und nicht zweimal erfolgen müssen und im Bereich der Maßnahmendurchführung nur verhältnismäßig geringer Aufwand entsteht.

Abschätzung der erforderlichen Feldgröße

Angenommen wird, dass in einer Branche 10% der Betriebe nachweisbar Schwarzarbeiter beschäftigen, wenn die Maßnahme nicht durchgeführt wird. Angenommen wird ferner, dass die Maßnahme die Zahl der Betriebe mit Schwarzarbeit um 25% senkt. Wie viele Betriebe braucht man in der Stichprobe, damit dieser Effekt noch statistisch signifikant nachgewiesen werden kann?

Dies lässt sich näherungsweise durch einen Zwei-Stichproben-Test abschätzen, dem als Approximation für große Stichproben die Normalverteilung zugrunde liegt. Dieser Test testet hier die Hypothese, dass die Wahrscheinlichkeit von Verdachtsfällen in der Maßnahme- und der Kontrollgruppe gleich hoch ist.

Beträgt die Feldgröße z.B. jeweils 1000 Unternehmen in Maßnahme- und Kontrollgruppe, findet man in der Teilnehmergruppe unter den oben genannten Annahmen 75 und in der Nichtteilnehmergruppe 100 Unternehmen mit Schwarzarbeit. Dieser Unterschied ist zu einem Signifikanzniveau von 5% statistisch signifikant, da die Wahrscheinlichkeit, dass beide Stichproben der gleichen Grundgesamtheit entstammen, mit 0.048 kleiner ist als 5%.

Die folgende Tabelle enthält weitere Beispiele für Annahmen über die Wirkung der Maßnahme und das Ausmaß der Schwarzarbeit. Die Tabelle zeigt jeweils, welche Feldzahlen benötigt werden, um signifikante Ergebnisse zu erzeugen. Augenfällig ist: je wirksamer die Maßnahme und je höher die Verbreitung von Schwarzarbeit, desto geringer die Fallzahlen, die benötigt werden, um signifikante Ergebnisse zu erhalten.

Feldgröße	Prozentsatz der Unternehmen mit Schwarzarbeit	Verminderung der Schwarzarbeit durch die Maßnahme in Prozent	Signifikanzniveau
1000	10	25	0,048
500	20	25	0,037
1500	20	15	0,034
5000	10	10	0,025

4.2.9 Besondere Probleme bei der Durchführung des Projekts

Die Trennschärfe beim Maßnahmeeinsatz steht in Frage, weil die Unternehmen aus Treatment- und Kontrollgruppe wahrscheinlich miteinander kommunizieren. Dies ist vor allem dann der Fall, wenn diese in Verbänden organisiert sind. Diese Art der Kommunikation gefährdet die Zielsetzung der Evaluation, weil Unternehmen aus der Kontrollgruppe auf diese Weise von der Maßnahme erfasst werden und sich dadurch die Unterschiede in den Indikatoren zwischen beiden Gruppen vermindern. Die Evaluation ist also in Branchen sinnvoll, wo diese Kommunikation vermutlich weniger stattfindet. Ähnlich wie im Projekt 4.1 ist dies insbesondere in stationären Gewerben mit vielen einzelnen Unternehmen zu erwarten (Gaststätten, kleine Handwerksbetriebe). Zwischen privaten Haushalten wird vermutlich am wenigsten kommuniziert, allerdings können hier keine Prüfungen angekündigt werden und das Ausmaß der Schwarzarbeit nicht mit einer Prüfung erhoben werden. Daher ist das hier skizzierte Projekt für den Haushaltsbereich nicht geeignet.

Ein weiteres Problem der Kommunikation zwischen Teilnehmer- und Kontrollgruppe besteht darin, dass die Glaubwürdigkeit der Maßnahmen der FKS in Gefahr gerät, wenn die Zielgruppe dieser Maßnahmen den Eindruck erhält, dass diese Maßnahmen willkürlich durchgeführt werden. Mögliche Lösungen für dieses Problem sind in Abschnitt 4.2.2 skizziert. Ferner ist darauf zu achten, dass die Anschreiben nicht in dem Sinne als „Warnung“ verstanden werden, dass nicht angeschriebene Firmen aufgrund des Ausbleibens einer vorherigen Warnung rechtlich gegen die Prüfungen bzw. ihre Folgen vorgehen.

Bei der Durchführung muss darauf geachtet werden, dass auf keinen Fall „durchsickern“ darf, dass die Prüfungen tatsächlich sowohl in der Maßnahme- als auch in der Kontrollgruppe stattfinden. Diese Tatsache wiederum ist notwendig, damit der Umfang der Schwarzarbeit in beiden Gruppen gleich gemessen wird. Dieses Vorgehen beschränkt zugleich die Möglichkeiten einer internen Kommunikationsstrategie.

4.2.10 Mögliche Varianten

Statt eines einheitlichen Anschreibens, das an die Teilnehmergruppe versandt wird, könnten auch unterschiedliche Anschreiben konzipiert werden. Diese könnten sich z.B. darin unterscheiden, ob auf Strafen hingewiesen wird oder ob vorwiegend an die Ehrlichkeit der Zielgruppe appelliert wird. Dies würde allerdings voraussetzen, dass die Größe der Stichprobe erhöht wird.

Dieses Projekt könnte sich prinzipiell auch für Werkverträge eignen. Angeschrieben würden in diesem Fall eine Zufallsauswahl von Unternehmen, die Werkverträge nutzen. Hierbei ist allerdings darauf zu achten, dass infolge der bestehenden Kontingente sowie aufgrund der Sperrung bestimmter Arbeitsagenturbezirke die Größe der Grundgesamtheit aller Unternehmen mit Werkverträgen begrenzt ist.

4.2.11 Erwartete Ergebnisse des Evaluationsprojekts

- Durch die Prüfungen in einer repräsentativen Stichprobe ergibt sich ein Bild des Umfangs der Schwarzarbeit im geprüften Bereich. Der Umfang der Schwarzarbeit wird dabei nicht allein an den Verdachtsfällen, sondern auch an den Fällen der Ahndung sichtbar. Hierbei werden auch regionale, betriebsgrößenspezifische und andere Unterschiede in der Intensität der Schwarzarbeit sichtbar.
- Im Zentrum steht der kausale Effekt der Ankündigung verstärkter Prüfungen in einem bestimmten Maß auf den Umfang der Schwarzarbeit. Hieraus wird erkennbar, in welchem Umfang die angekündigten Maßnahmen der FKS mittels ihrer präventiven Wirkungen den gewählten Indikator wie die Zahl der Verdachtsfälle, die Zahl der Ahndungen oder das Ausmaß der Ahndungen beeinflussen.
- Das Projekt stellt die Kosten, die aus der Erhöhung der Prüfintensität um ein bestimmtes Maß resultieren, den fiskalischen Erträgen in Form nicht zu zahlender Lohnersatzleistungen oder in Form zusätzlichen Aufkommens an Steuern und Sozialversicherungsbeiträgen gegenüber.

4.3 Leistungsempfänger

4.3.1 Fragestellung

Arbeitslose werden gezielt über Maßnahmen zur Aufdeckung von Schwarzarbeit und Leistungsmissbrauch informiert. Dies soll bei der einzelnen Person das subjektiv wahrgenommene Risiko erhöhen, im Falle eines Leistungsmissbrauchs in Form von Schwarzarbeit entdeckt und entsprechend sanktioniert zu werden. Dabei sollte das seitens des Leistungsempfängers subjektiv wahrgenommene Risiko einer Aufdeckung von Schwarzarbeit im Idealfall um einen quantitativ bestimmten Prozentsatz erhöht werden. Hierbei könnte ähnlich vorgegangen werden, wie im Projekt „Ankündigung von Prüfungen“ bereits skizziert. Es könnte beispielsweise während eines Gesprächs mit einem Berater in der Agentur für Arbeit ein entsprechender Hinweis gemacht werden, oder es könnten besondere Informationsmaterialien über die Bekämpfung der Schwarzarbeit und die rechtlichen Folgen von Leistungsmissbrauch überreicht werden.

Zu beachten ist dabei, dass Leistungsempfänger bereits jetzt von der BA Informationsmaterialien zum Thema Leistungsmissbrauch erhalten. In Broschüren über die Rechten und Pflichten der Arbeitslosen und über die Berücksichtigung von Nebeneinkommen werden unter anderem die Pflicht zur Meldung von Beschäftigungsverhältnissen, die Möglichkeit von Rückforderung von zu Unrecht erhaltenen Leistungen und eine mögliche Verfolgung von Missbrauch als Straftat oder Ordnungswidrigkeit angesprochen. Die Tätigkeit der FKS wird jedoch in diesen Broschüren nicht dargestellt. Bei Planung des Projektes ist darauf zu achten, dass die zusätzlich gegebene Information einen Zusatzeffekt auf die Risikoeinschätzung der Betroffenen hat. Hierbei sind unterschiedliche Vorgehensweisen denkbar, neben einer inhaltlichen Differenzierung gegenüber den BA-Broschüren könnten die im Zuge des Projektes zu verteilenden besonderen Informationsmaterialien z.B. auch zu einem anderen Zeitpunkt überreicht werden als die bereits verwendeten und damit einen gesonderten Effekt haben.

Tabelle 4.3: Zeitlicher Ablauf des Evaluationsprojekts „Leistungsempfänger“

Schritt 1 (ca. Monat 1-2)	Auswahl der Stichprobe anhand von BA-Daten in den einbezogenen Standorten, Bestimmung des Teilnehmer- und Kontrollstatus mittels eines Zufallsverfahrens
Schritt 2 (ca. Monat 3-5)	Überreichung der Informationsmaterialien an Leistungsempfänger der Teilnehmergruppe. Durchführung der Implementationsanalyse und Kommunikationsstrategie zu den BA-Angehörigen
Schritt 3 (ca. Monat 5-15 bzw. bei Nachverfolgung in der Ahndung später)	Einholen der Ergebnisse des DALEB-Verfahrens und ggf. Nachverfolgung der Verdachtsfälle in der Ahndung Alternativ: Erhebung des Indikators legale Beschäftigung
Schritt 4 (ca. Monat 15-18 oder später)	Datenauswertung Kosten-Nutzen-Analyse

Auch hier ist es zur Vermeidung unnötigen Aufwands sinnvoll, nicht alle Leistungsempfänger zu betrachten, sondern solche Personen, die eine Ausbildung in „gefährdeten“ Berufen wie z.B. Bauberufen oder im Handwerksbereich besitzen.

4.3.2 Art der Variation

Die Stichprobenziehung muss zunächst aus den Daten der BA (Leistungsempfängerdatei) geschehen. Variation entsteht auf der Ebene der Person, indem aus einer Stichprobe von Leistungsempfängern durch einen Zufallsprozess eine Teilnehmer- und eine Kontrollgruppe bestimmt wird.

4.3.3 Verwendbare Indikatoren

Im Prinzip kommen alle drei Indikatorengruppen (Schwarzarbeit, legale Beschäftigung, Rechtsbewusstsein) in Frage. Die Wirkung auf den Indikator Schwarzarbeit dürfte erheblich direkter und quantitativ umfangreicher sein als die Wirkung auf den Indikator legale Beschäftigung. Wegen der Probleme beim DALEB-Verfahren (siehe Abschnitt 2.2.7) erscheint es allerdings sinnvoll, den Indikator legale Beschäftigung in jedem Fall im Blick zu behalten.

4.3.4 Erfassung der Indikatoren in den Daten

Für den Indikator Schwarzarbeit ist eine einmalige Erhebung ausreichend, da Variation auf der Ebene der Person vorliegt. Die Evaluation muss nicht an allen Standorten durchgeführt werden, es wären ca. 10 bis 20 Standorte ausreichend, so dass Probleme bei der Umsetzung im direkten Kontakt gelöst werden können. Diese Zahl würde sich erhöhen, wenn man die Ergebnisse weiter aufgliedern wünscht (z.B. nach Regionstyp, Bildungsniveau der Betroffenen usw.). Je nachdem, wie hoch die Wahrscheinlichkeit dafür ist, dass ein Leistungsempfänger aus der Kontrollgruppe nach einem Leistungsmissbrauch sanktioniert wird, sollte die Zahl der Personen in der Teilnehmer- und Kontrollgruppe mehr oder weniger hoch sein.

Auf der Basis der vorliegenden Daten konnte bislang nicht exakt festgestellt werden, wie groß, gegeben dass für eine Person Leistungsempfang verzeichnet wird, die Wahrscheinlichkeit einer Zuleitung durch die BA im Rahmen des DALEB-Verfahrens ist. Eine Zahl von 40.000 Zuleitungen in einem Jahr, die zumindest in erster Näherung plausibel erscheint, ergibt bei überschlagsmäßig 2.000.000 Empfängern von Arbeitslosen- oder Unterhaltsgeld eine DALEB-Zuleitungsquote von 2%. Wenn die Maßnahme die Wahrscheinlichkeit der Beanstandung um 20% senkt, ist nach dem im letzten Abschnitt beschriebenen Verfahren eine Stichprobe von jeweils knapp 10.000 Personen in der Teilnehmer- und Kontrollgruppe erforderlich, um den Unterschied statistisch signifikant nachweisen zu können.

Der Indikator Schwarzarbeit kann durch das DALEB-Verfahren bzw. durch Verknüpfung mit ProFiS in Form einer anschließenden Verurteilung/Verhängung eines Bußgeldes erfasst werden. Dabei ist zu berücksichtigen, dass die Zuverlässigkeit des DALEB-Verfahrens nicht vollkommen ist. Hier ist es sinnvoll, sich auf solche Standorte zu beschränken, an denen das DALEB-Verfahren in der Vergangenheit ohne schwerwiegende Probleme abgelaufen ist.

Der Indikator legale Beschäftigung kann durch die Beschäftigtenmeldungen aus der Beschäftigtenstatistik der BA auf der Ebene der Person gewonnen werden. Da die BA zugleich die Informationsmaterialien verteilt, können Maßnahme- wie auch Ergebnisindikator durch die BA einander zugespielt werden. Da der Effekt der Maßnahme auf die legale Beschäftigung geringer sein dürfte, müsste der Umfang der Stichprobe entsprechend erhöht werden.

Für den Indikator Rechtsbewusstsein sollte eine Befragung bei jeweils 500 bis 1000 Personen der Teilnehmer- und Kontrollgruppe durchgeführt werden. Dabei ist zu prüfen, ob Datenschutzgesichtspunkte die Weitergabe von Telefonnummern oder anderen Kontaktmöglichkeiten der einbezogenen Personengruppe erlauben. Wenn dies nicht der Fall ist, lässt sich dieser Indikator nicht mit dem Projekt verknüpfen.

4.3.5 Kommunikationsstrategie

Im Vorfeld der Evaluation wäre die Kooperation der BA sicherzustellen (Bereitstellung von Daten und Verteilung der Informationsmaterialien).

Im Verlauf der Evaluation sollte eine spezielle Kommunikationsstrategie für die Mitarbeiter der Agentur für Arbeit zur Anwendung kommen, die die Informationsschrift verteilen. Dabei ist sicherzustellen, dass die Informationsmaterialien nicht nach persönlicher Risikoanalyse des Vermittlers, sondern strikt nach den Vorgaben der Evaluation verteilt werden.

Die Erhebung des Indikators Schwarzarbeit erfolgt aus ProFiS, so dass hier auf Seiten der FKS keine besondere Kommunikationsstrategie benötigt wird.

4.3.6 Implementationsanalyse

Es wird durch Interviews ermittelt, ob die Teilnehmer tatsächlich die Informationsmaterialien erhalten haben. Bei den Agenturen für Arbeit wird erfragt, wie die Verteilung der Informationsmaterialien erfolgt (werden diese überhaupt von der Teilnehmergruppe angenommen?). Auch die Funktionsweise und der Stand der Umsetzung des DALEB-Verfahrens wird in Interviews bei den örtlichen Agenturen für Arbeit angesprochen.

4.3.7 Kosten-Nutzen-Analyse

Eine aussagekräftige Kosten-Nutzen-Analyse ist in diesem Projekt vermutlich nicht mit einem so umfassenden Ansatz möglich wie in den zuvor diskutierten Projekten. Es ist davon auszugehen, dass im Bezug auf die Frage, ob der Instrumenteneinsatz fiskalisch lohnt oder nicht, nur eine unter restriktiven Annahmen gültige Einschätzung abgegeben werden kann. Dies liegt an diversen Sachverhalten und Problemen, die im Folgenden erläutert werden.

Zunächst einmal sollte die Kosten-Nutzen-Analyse nicht wie bei den zuvor behandelten Projekten davon ausgehen, dass das tatsächliche Entdeckungsrisiko bei Leistungsmisbrauch im Wesentlichen richtig eingeschätzt wird. Dies lässt sich bereits aus der Tatsache ersehen, dass das DALEB-Verfahren regelmäßig Leistungsmisbrauch in erheblichem Umfang auf-

deckt. Dies ist als Indiz dafür zu werten, dass viele Betroffenen über den Datenabgleich nicht informiert sind oder die Konsequenzen eines solchen Abgleichs für das Risiko, bei Leistungsmisbrauch entdeckt und sanktioniert zu werden, nicht korrekt bewerten. Daher sollte bei der Kosten-Nutzen-Analyse in diesem Projekt die zu bewertende Tätigkeit der FKS darin gesehen werden, dass das durch Prüfungen und Ermittlungen gegebene Risiko, bei Leistungsmisbrauch entdeckt zu werden, in der Teilnehmergruppe kommuniziert wird.

Das Hauptproblem der Kosten-Nutzen-Analyse in diesem Projekt besteht darin, dass die Kommunikation eines gewissen Entdeckungsrisikos zu sehr niedrigen Kosten möglich ist. So ist etwa die Produktion von Informationsmaterialien und die Überreichung an Leistungsempfänger zu geringen Gesamtkosten möglich. Selbstverständlich ist aber der Hinweis auf Prüfungen und Sanktionen nur dann wirksam, wenn diese tatsächlich durchgeführt werden. Es ergibt sich also das Problem, dass die erheblichen Kosten, die bei der FKS durch Prüfungen, Ermittlungen und Aktivitäten zur Ahndung von Leistungsmisbrauch entstehen, in der Kosten-Nutzen-Analyse anteilig berücksichtigt werden müssten. Es wird in der Praxis problematisch sein, zu einer überzeugenden Zurechnung von Kostenbestandteilen zu gelangen. Das bedeutet, dass eine umfassende Quantifizierung der Kosten insgesamt nicht verlässlich möglich sein dürfte.

Dennoch sind Aussagen über die fiskalischen Nettoeffekte der Maßnahme möglich. Sie sollten sich allerdings darauf beschränken, den unmittelbaren fiskalischen Effekt der Maßnahme zu bestimmen. Die Kosten-Nutzen-Analyse würde dann klären, wie hoch der fiskalische Nutzen der Maßnahme ist, Leistungsempfänger gezielt über die Arbeit der FKS zu informieren. Dabei müsste unterstellt werden, dass die Kapazitäten der FKS zur Prüfung, Ermittlung und Ahndung von Leistungsmisbrauch ohnehin vorgehalten werden. Vermutlich wird man unter diesen Umständen bereits bei sehr geringen Effekten der Maßnahme auf den Umfang des Leistungsmisbrauchs zu einer positiven Bewertung der Maßnahme kommen. Bei der Interpretation dieser zu erwartenden positiven Effekte ist selbstverständlich auf die methodischen Probleme und die getroffenen Annahmen Rücksicht zu nehmen.

Falls das DALEB-Verfahren verwendet wird, um das Ausmaß des Leistungsmisbrauchs zu identifizieren, sollte man sich bei der Ermittlung des fiskalischen Nutzens auf die Betrachtung der Verminderung des Leistungsmisbrauchs beschränken. Das bedeutet, dass der fiskalische Nutzen der Maßnahme mit relativ geringem Aufwand bestimmt werden kann. Dazu wird aus den im DALEB-Verfahren erhobenen Daten der Umfang der laufend missbräuchlich bezogenen Leistungen in der Teilnehmer- und der Kontrollgruppe geschätzt. Aus der Differenz ergibt sich dann eine Abschätzung für den fiskalischen Nutzen im Sinne einer Verminderung im laufenden Leistungsmisbrauch. Falls der Indikator legale Beschäftigung verwendet wird, so kann analog zu den Ausführungen zur Kosten-Nutzen-Analyse für Projekt 4.1 sowohl die Verminderung des missbräuchlichen Leistungsbezugs als auch die Zunahme von Beiträgen zur Sozialversicherung und im Steueraufkommen berücksichtigt werden.

Um den restriktiven Annahmen Rechnung zu tragen, unter denen die Kosten-Nutzen-Analyse in diesem Projekt erstellt wird, sollte abschließend der fiskalische Nutzen des Instrumenteneinsatzes mit den Kosten verglichen werden, die die Bekämpfung des Leis-

tungsmisbrauchs insgesamt bei der FKS verursacht. Damit würde die Frage untersucht werden, ob der Instrumenteneinsatz, d.h. die speziellen Kommunikationsmaßnahmen der FKS gegenüber Leistungsempfängern bei Konstanzhaltung der eigentlichen Prüfungs-, Ermittlungs- und Ahndungsaktivitäten, die gesamten Kosten der Bekämpfung des Leistungsmisbrauchs kompensiert. Falls dies der Fall sein sollte, würde sich eine methodisch anspruchsvollere Kosten-Nutzen-Analyse erübrigen. Auch wenn das Ergebnis dieses Vergleichs zeigen sollte, dass die gesamten Kosten der Bekämpfung des Leistungsmisbrauchs den fiskalischen Nutzen der speziellen Kommunikationsmaßnahmen übersteigen, aber ein substantieller Teil der Gesamtkosten durch die Kommunikationsmaßnahmen kompensiert wird, könnte dies für den fiskalischen Nutzen der untersuchten Maßnahme sprechen.

4.3.8 Aufwand

Der Aufwand auf der Seite der FKS ist insgesamt sehr gering. Es entsteht insofern Aufwand, als Einblick in die Datenbank ProFiS gegeben werden muss. Darüber hinaus muss das Vorgehen bei der Evaluation mit der BA abgestimmt werden.

Die von der BA zu verteilenden Informationsmaterialien müssten in geeigneter Weise konzipiert und produziert werden. Ferner entsteht auf der Seite der Agenturen für Arbeit Aufwand bei der Verteilung der Informationsmaterialien an die Teilnehmergruppe, der Vermerkung dieser Tatsache in den Daten sowie bei der späteren Bereitstellung der Daten an den Evaluator. Da bei einer unterstellten DALEB-Zuleitungsquote von 2% die Teilnehmer- und Nichtteilnehmergruppe jeweils 10.000 Personen umfassen, ist dieser Aufwand nicht zu vernachlässigen.

Da alle Ergebnisindikatoren aus bereits existierenden Datenbeständen entnommen werden können, ist der Aufwand bei diesem Projekt insgesamt relativ gering. Auch eine Ausdehnung der Zahl der Teilnehmer und Nichtteilnehmer wäre auf Seiten der FKS oder des Evaluators mit moderaten Zusatzkosten verbunden, so dass in Betracht zu ziehen wäre, ob die Ergebnisse nach bestimmten Strukturmerkmalen differenziert werden können. Dagegen steigen die Kosten vermutlich relativ stark in der Zahl der einbezogenen Standorte.

4.3.9 Besondere Probleme bei der Durchführung des Projekts

Das Projekt beruht darauf, dass ein bestimmter Kreis von Leistungsempfängern in besonderer Weise über die Tätigkeit der FKS bei der Bekämpfung des Leistungsmisbrauchs aufgeklärt wird. Es kommt entscheidend darauf an, dass diese zusätzliche Information den betreffenden Personen zielgenau übermittelt wird. Dies ist angesichts der Vielzahl von Informationen, die Leistungsempfängern von der BA übermittelt werden, nur durch sorgfältige Planung des Vorgehens zu bewerkstelligen.

Als weitere besondere Herausforderung für das Gelingen des Projektes ist die notwendige enge Kooperation mit der Bundesagentur für Arbeit zu nennen. Eine Kooperation mit der BA müsste durch das BMF auf oberer Ebene sichergestellt werden. Dies erscheint indes grundsätzlich nicht unrealistisch zu sein, da die Zielsetzung des Projekts in der Evaluation von

Maßnahmen besteht, die die legale Beschäftigung erhöhen. Es wäre zu prüfen, ob der Status der BA als Zusammenarbeitsbehörde hierbei den erforderlichen Rahmen bilden könnte.

Die in der jüngeren Vergangenheit festzustellende Erosion des DALEB-Verfahrens stellt ein besonderes Problem dar. Angesichts der hohen Belastung durch die Arbeitsmarktreformen ist die Bereitschaft der BA, (ggf. zusätzliche) Aufgaben im Rahmen eines Evaluationsprojektes zu übernehmen, als kritisch einzuschätzen. Dabei geht es vor allem um die Datenbereitstellung über das DALEB-Verfahren. Wie im Projekt „Ankündigungen“ sollte man sich hier auf solche Standorte beschränken, an denen das DALEB-Verfahren bislang ohne große Reibungsverluste abläuft.

4.3.10 Mögliche Varianten

Statt einheitlicher Informationsmaterialien, die an die Teilnehmergruppe verteilt werden, könnten auch unterschiedliche Informationsmaterialien verteilt werden. Diese könnten sich z.B. darin unterscheiden, ob auf Strafen hingewiesen wird oder ob vorwiegend an die Ehrlichkeit der Zielgruppe appelliert wird. Auch diese würde eine Erhöhung des Stichprobenumfangs und damit der Anzahl der verteilten Informationsmaterialien voraussetzen.

4.3.11 Erwartete Ergebnisse des Evaluationsprojekts

- Bisher wird nicht systematisch erhoben, welches das Verhältnis von Leistungsempfängern einerseits und Zuleitungen im Wege des DALEB-Verfahrens bzw. Ahndungen wegen Leistungsmissbrauchs andererseits ist. Die Erhebung in der Kontrollgruppe würde hier quantitative Evidenz bereitstellen. Diese Evidenz könnte mit Hilfe der über die BA erhobenen Daten auch hinsichtlich bestimmter Charakteristika der Leistungsempfänger (Alter, Familienstand, genaue Berufszugehörigkeit usw.) differenziert werden, um so ein Risikoprofil des Leistungsmissbrauchs zu erstellen.
- Das Evaluationsprojekt würde den kausalen Effekt der Information auf den Umfang des Leistungsmissbrauchs bzw. auf den Umfang der legalen Beschäftigung darstellen. Je nach Formulierung in den Informationsmaterialien und den verwendeten Varianten würde dieser Effekt sich auf eine vermehrte Aufdeckungswahrscheinlichkeit (z.B. durch Information über das DALEB-Verfahren), auf die Information bezüglich des Strafmaßes oder auf den Effekt der Appellation an ethische Normen beziehen. Der Vergleich der beiden Indikatoren Schwarzarbeit und legale Beschäftigung würde zudem eine Einschätzung darüber erlauben, inwieweit die Bekämpfung der Schwarzarbeit mit einer Erhöhung der legalen Beschäftigung einhergeht.
- Eine Gegenüberstellung der Kosten der Durchführung der Maßnahmen und des vermiedenen Umfangs von Leistungsmissbrauch bzw. der zusätzlich geschaffenen legalen Beschäftigung im Sinne einer Kosten-Nutzen-Analyse.

4.4 Personelle Ausstattung

4.4.1 Fragestellung

Dieses Evaluationsprojekt zielt auf die Frage, wie sich die personelle Ausstattung der Standorte auf die Abschreckungswirkung der FKS auswirkt. Ist die Abschreckung an Standorten höher, die personell bereits gut ausgestattet sind? Wie viel bewirkt die Erhöhung des Personalaufwands bei der FKS um durchschnittlich eine Stelle hinsichtlich der betrachteten Indikatoren?

Aufbauend auf einer wiederholten Erhebung könnte die Abschätzung, in welchem Maße der Umfang der Schwarzarbeit von der Personalstärke beeinflusst wird, die Entwicklung des Indikators über die Zeit ausnutzen.

Tabelle 4.4: Zeitlicher Ablauf des Evaluationsprojekts „Personelle Ausstattung“

Schritt 1 (ca. Monat 1-3)	Erstmalig Datenerhebung durch Schwerpunktprüfung oder Befragung an allen Standorten der FKS. Hierbei Kommunikationsstrategie und Implementationsanalyse.
Schritt 2 (ca. Monat 12-15)	Zweite Datenerhebung nach demselben Verfahren wie in Schritt 1.
Schritt 3 (ca. Monat 15-18)	Datenauswertung, ggf. unter Verwendung von Ergebnissen der Ahndung. In diesem Fall müsste der zeitliche Rahmen so weit gesteckt werden, dass die weit überwiegende Mehrheit der Verfahren zum Zeitpunkt der Datenauswertung abgeschlossen ist. Kosten-Nutzen-Analyse.

4.4.2 Art der Variation

Als Variation wird die im Verlauf des Ausbaus der FKS regional unterschiedliche personelle Ausstattung der Standorte im Bereich Prävention und Prüfung und Ermittlung verwendet (Ist-Stärke relativ zu Soll-Stärke). Dabei wird unterstellt, dass diese Variation im Wesentlichen auf der Attraktivität der Standorte für Bewerber beruht und insofern unabhängig vom vermuteten Umfang der Schwarzarbeit ist. Angesichts der Tatsache, dass der Personalbestand an den Standorten weitgehend angebotsseitig gesteuert wird, erscheint diese Annahme als sinnvoll. Alle freien Stellen werden ausgeschrieben, lediglich bei der Terminierung gibt es einen möglichen Einfluss der Dringlichkeit. Ob die Stellen besetzt werden, hängt dann in erster Linie von der Attraktivität des Standortes ab. Darunter fällt beispielsweise das örtliche Mietniveau. Zwangsmaßnahmen wie Versetzungen sind sehr selten und spielen insgesamt keine Rolle bei der Auffüllung zur Sollstärke. Daher dürfte der Grad, in dem die Sollstärke erreicht wird, nicht wesentlich vom Ausmaß der Schwarzarbeit abhängen.

Wenn das Verhältnis Sollstärke zu Iststärke der Indikator für die Maßnahme ist, muss sichergestellt sein, dass die Sollstärke über den Zeitraum der Evaluation nicht verändert wird.

Dies ist beobachtbar und die entsprechenden Standorte können aus der Evaluation herausgenommen werden.

Voraussetzung des Projekts ist eine hinreichend große Anzahl von Standorten, bei denen sich die Personalstärke im Zeitraum der Evaluation deutlich ändert. Daher ist es erforderlich, dass dieses Projekt relativ bald durchgeführt wird, bevor die meisten Standorte auf Sollstärke angewachsen sind (siehe Abschnitt 2.1.1). Zugleich hat dieses Projekt einen verhältnismäßig weit gespannten zeitlichen Rahmen, weil Änderungen der Ist-Größen innerhalb weniger Monate vermutlich nicht in hinreichendem Umfang zu beobachten sein dürften. Bevor das Projekt durchgeführt wird, sollte genau eingeschätzt werden, welche Änderungen in den Ist-Größen zu erwarten sind. Nur bei verhältnismäßig großen Änderungen sind statistisch signifikante Ergebnisse zu erwarten.

4.4.3 Verwendbare Indikatoren

Als Indikator ist hier vor allem der Umfang der Schwarzarbeit verwendbar. Die Beschäftigungssituation kommt nicht in Frage, da sie auf regionaler Ebene von zu vielen anderen Bestimmungsfaktoren abhängt (vgl. Projekt 4.1).

4.4.4 Erfassung der Indikatoren in den Daten

Bei der Erfassung der Indikatoren ist entscheidend, dass der Umfang und die Intensität der dazu notwendigen Prüfungen gerade nicht von der unterschiedlichen Personalstärke der Standorte abhängen. Daher sind hier separate, am besten im Zuge von Schwerpunktprüfungen an jedem Standort nach demselben Verfahren erhobene Daten erforderlich.

4.4.5 Kommunikationsstrategie

Eine Kommunikationsstrategie ist insofern notwendig, als die Vorgaben für die zur Erhebung der Daten notwendigen Schwerpunktprüfungen an den Standorten möglichst exakt umgesetzt werden sollten.

4.4.6 Implementationsanalyse

Unter Umständen ist auch eine Implementationsanalyse sinnvoll. Diese soll Informationen darüber bringen, wie gut die Umsetzung der Vorgaben für die Schwerpunktprüfungen tatsächlich ist.

Ein weiterer wichtiger Gegenstand der Implementationsanalyse wäre, dass die Erhöhung der Iststärke bei den Arbeitsmarkteteiligen auch wahrgenommen wird. Andernfalls sind keine präventiven Wirkungen zu erwarten.

4.4.7 Kosten-Nutzen-Analyse

Die Bedingungen für eine aussagekräftige Kosten-Nutzen-Analyse sind in diesem Projekt besonders gut. Dies liegt daran, dass die verwendete Variation in der Intensität des Maßnahmeinsatzes der FKS unmittelbar am Personalaufwand anknüpft. Der Personalaufwand ist der wichtigste Kostenfaktor bei der Bekämpfung der Schwarzarbeit, so dass hier eine besonders gute und verlässliche Abschätzung der Kosten des Maßnahmeinsatzes möglich ist.

Die Ermittlung der Kosten setzt also unmittelbar an der beobachteten Veränderung der Personalstärke an den verschiedenen Standorten an. Aus der Kosten- und Leistungsrechnung sind die notwendigen Informationen zu entnehmen, um aus der Veränderung in der Personalstärke die zusätzlich entstandenen Personalkosten zu bestimmen. Darüber hinaus sind Zusatzkosten zu berücksichtigen, die durch zusätzlich benötigte Bürofläche, Fahrzeuge usw. entstanden sind. Bei einer umfassenden Nutzung der Daten der Kosten- und Leistungsrechnung ist eine präzise Quantifizierung der Kosten zu erwarten, die durch die personelle Aufstockung der Standorte entstanden sind.

Die Ermittlung des fiskalischen Nutzens setzt am festgestellten Rückgang des Umfangs der Schwarzarbeit an und entspricht weitgehend dem Vorgehen, das im Abschnitt zur Kosten-Nutzen-Analyse für Projekt 4.1 beschrieben wurde. Es ist aber zu beachten, dass bei diesem Projekt die Veränderung des Umfangs der Schwarzarbeit in allen Branchen berücksichtigt werden sollte. Zur Ermittlung des zusätzlichen Steueraufkommens und der zusätzlichen Beiträge in die Sozialversicherungssysteme sowie der eingesparten Sozialleistungen müssten dann durchschnittliche Löhne und Gehälter für viele Branchen verwendet werden. Die Abschätzung der fiskalischen Effekte wird dadurch verkompliziert und mit zusätzlichen Unsicherheiten belastet.

4.4.8 Aufwand

Aufwand entsteht nur bei der Datenerhebung, dabei gelten die weiter oben beschriebenen Gesichtspunkte. Die Datenerhebung müsste an allen FKS-Standorten geschehen.

4.4.9 Besondere Probleme bei der Durchführung des Projekts

Bei der Erfassung der Indikatoren ist entscheidend, dass die dazu notwendigen Prüfungen nicht von der unterschiedlichen Personalstärke der Standorte abhängen. Dies setzt voraus, dass die schwach ausgestatteten Standorte die Schwerpunktprüfungen mit der gleichen Intensität durchführen wie die besser ausgestatteten. Inwieweit das der Fall ist, müsste in der Implementationsanalyse beobachtet werden.

Das Projekt muss durchgeführt werden, bevor alle FKS-Standorte die Sollstärke erreicht haben. Es kommt also darauf an, den Zeitpunkt für die Evaluation richtig zu wählen. Das Projekt ist grundsätzlich nur zu realisieren, wenn Variation im Verhältnis von Ist- zu Sollstärke besteht. Das Projekt ist entweder so zügig umzusetzen, dass die im Zuge des Aufbaus der FKS-Standorte gegenwärtig vorhandene Variation in der personellen Ausstattung über die

Zeit genutzt werden kann. Andernfalls könnte es nur bei künftigen standortübergreifenden Änderungen in der Sollstärke umgesetzt werden.

4.4.10 Mögliche Varianten

Bei der Erfassung der Personalstärke kann zwischen den unterschiedlichen Arbeitsgebieten differenziert werden. Insbesondere bietet es sich an, den Einsatz von Personal in den Bereichen Prävention und Prüfungen und Ermittlungen separat zu analysieren. Die Fallzahlen setzen allerdings der Differenzierung der Ergebnisse klare Grenzen.

Neben der Personalstärke können gesetzliche Änderungen natürliche Variation hervorrufen. Hier kommt es darauf an, dass unterschiedliche Arten oder Bereiche von Schwarzarbeit unterschiedlich von Gesetzesänderungen betroffen sind. Sieht das Gesetz eine Intensivierung der Bekämpfungsaktivitäten in bestimmten Bereichen vor, in anderen Bereichen jedoch nicht, so kann der Differenzen-von-Differenzen-Schätzer benutzt werden, um den Effekt der Gesetzesänderungen auf den gewählten Indikator zu bestimmen (siehe dazu den Anhang). Eine Schwierigkeit hierbei besteht darin, dass man Daten vor und nach der Änderung benötigt. Eine Evaluationsstudie müsste also die Gesetzesänderung antizipieren. Ein Problem hierbei ist, dass von einer absehbaren Gesetzesänderung auch bereits Verhaltensänderungen ausgehen können. Daher ist der Zustand vor der Änderung möglicherweise nur mit Verzerrungen zu erheben. Günstiger stellt sich die Situation dar, wenn bereits bestehende Daten aus dem Ermittlungsprozess benutzt werden können, etwa die Ergebnisse verdachtsloser Prüfungen. In diesem Fall ist allerdings wieder zu bedenken, ob und inwiefern die Gesetzesänderung auf die Datenerhebung einwirkt. Führt sie zu einer Erhöhung der Zahl der Prüfungen in dem von der Änderung betroffenen Bereich, nicht aber in anderen Bereichen, können der verhaltenslenkende und der Erfassungseffekt nicht voneinander getrennt werden. Da derzeit nicht absehbar ist, welche Gesetzesänderungen in nächster Zeit im Bereich der Schwarzarbeit zu erwarten sind, wird dieser Punkt hier nicht weiter verfolgt. In den meisten Fällen kann jedoch schnell beurteilt werden, ob eine Gesetzesänderung so gestaltet ist, dass sie das hier skizzierte Vorgehen als sinnvoll erscheinen lässt.

4.4.11 Erwartete Ergebnisse des Evaluationsprojekts

- Das Projekt würde zunächst eine Einschätzung über das Ausmaß der Schwarzarbeit im geprüften Bereich ergeben. Entsprechend der verhältnismäßig globalen Definition der Maßnahme könnte auch der Ergebnisindikator relativ breit (also z.B. über mehrere Branchen hinweg) erfasst werden und würde so eher als bei den anderen Projekten eine Aussage über die Verbreitung von Schwarzarbeit in der gesamten Volkswirtschaft gestatten.
- Der kausale Effekt der Personalstärke auf das Ausmaß der Schwarzarbeit erfasst die Wirkungen der FKS bzw. der einzelnen Arbeitsgebiete der Sachgebiete E auf das Ausmaß der Schwarzarbeit. Infolge der begrenzten Zahl an Standorten wird es kaum möglich sein, zwischen Effekten an unterschiedlichen Standorten zu differenzieren.
- Die Kosten-Nutzen-Analyse erlaubt eine Einschätzung des fiskalischen Nutzens zusätzlicher FKS-Mitarbeiter auf das Ausmaß von Leistungsmissbrauch bzw. das zusätzliche

generierte Steueraufkommen. Diese Nutzen sind relativ klar definierten Kosten zusätzlicher FKS-Mitarbeiter gegenüberzustellen.

4.5 Wiederholungstäter

4.5.1 Fragestellung

Dieses Projekt widmet sich der Frage, ob es einen Zusammenhang zwischen der Härte von Sanktionen und der Wahrscheinlichkeit gibt, mit der bereits wegen Schwarzarbeit belangte Personen oder Betriebe wiederum Schwarzarbeit anbieten bzw. nachfragen. Weisen Personen oder Betrieben, die in der Vergangenheit wegen Schwarzarbeit belangt worden sind, eine umso geringere Neigung zur Wiederholungstat auf, je härter die Strafe bzw. das Bußgeld im ersten Fall gewesen ist? Auch die relative Wirksamkeit von Gefängnisstrafen und Geldbußen kann überprüft werden. Die Untersuchung kann sich sowohl mit den Nachfragern als auch den Anbietern von Schwarzarbeit befassen. Um die Variation, die aus der Unterschiedlichkeit von Vergehen resultiert, als Quelle von Unterschieden in der Neigung zu Wiederholungstaten auszuschließen, müsste ein Standardvergehen definiert werden, das in der Datenbasis (ProFiS) häufig vorkommt. Für dieses Standardvergehen würde dann überprüft, ob die Variation in der Härte der Sanktionen einen Effekt auf die Wahrscheinlichkeit von Wiederholungstaten hat.

Tabelle 4.5: Zeitlicher Ablauf des Evaluationsprojekts „Wiederholungstäter“

Schritt 1 (ca. Monat 1-6)	Erhebung von Personen, die während eines zu wählenden Zeitraums in ProFiS wegen Schwarzarbeit erfasst worden sind, sowie der verübten Tatbestände und der verhängten Strafen und Bußgelder
Schritt 2 (ca. Monat 6-12)	Datenauswertung

4.5.2 Art der Variation

Die Variation wird auf der Ebene des Individuums durch unterschiedlich hohe Strafen für ähnliche Vergehen gemessen. Diese Variation kommt ihrerseits dadurch zustande, dass von den Gerichten bzw. den Verwaltungsbehörden fallweise und relativ zueinander unterschiedlich entschieden wird, wobei ein Rahmen für diesen Ermessensspielraum durch den Bußgeldkatalog und andere Regelungen besteht. Die Variation entsteht damit im Wesentlichen auf der Ebene der Standorte oder der HZÄ, obwohl sie auf der Ebene des Individuums festgestellt wird. Die Stichprobe sollte daher Fälle aus allen HZÄ umfassen.

Ein weiterer möglicher Indikator, der die Härte der Strafen auf der Ebene der Standorte messen könnte, ist die Anzahl der Einsprüche. Dieser Indikator könnte auch zur Validierung des Indikators Strafhöhe verwendet werden.

4.5.3 Verwendbare Indikatoren

Die sinnvollen Indikatoren für die Ergebnisvariable beziehen sich auf die Schwarzarbeit. Am einfachsten ist es, eine Null-Eins-Variable für Wiederholungstaten zu verwenden. Unter Umständen könnte auch das Ausmaß der zum zweiten oder wiederholten Mal festgestellten Schwarzarbeit einbezogen werden oder bestimmte Tatbestände separat erfasst werden.

4.5.4 Erfassung der Indikatoren in den Daten

Es wäre zunächst aus ProFiS eine Stichprobe von Personen mit Ahndungen für Schwarzarbeit zu ziehen. Der Indikator für die Maßnahmeintensität (Härte der Strafe) dürfte aus ProFiS hervorgehen. Die Härte der Strafe sollte nicht absolut gemessen werden, sondern im Bezug auf den jeweiligen Tatbestand, der auch aus ProFiS ersichtlich sein dürfte. Da bei der Höhe der Strafen nicht vollständig auf den Tatbestand kontrolliert werden kann, wird der Messfehler in diesem Indikator allerdings relativ hoch sein.

Der Stichprobenumfang müsste bei einer vermutlich nicht allzu hohen Quote von beobachteten Wiederholungstaten verhältnismäßig hoch sein (vgl. Projekt 4.3). Da alle Daten aus vorhandenen Quellen entnommen werden und der Aufwand insofern eher gering ist, wird hier auf eine Berechnung der erforderlichen Stichprobengröße verzichtet.

Der Wirkungsindikator ist eine in ProFiS erfasste Beanstandung oder Ahndung innerhalb eines bestimmten Zeitraumes nach der erstmaligen Ahndung. Praktisch könnte es sich schwierig gestalten, die erforderliche Nachverfolgung einer Stichprobe von Personen mit Ahndungen in der Datenbasis zu realisieren; hier ist auf Seiten des Evaluators entsprechender Personalbedarf anzusetzen.

4.5.5 Kommunikationsstrategie

Nicht erforderlich, da alle Indikatoren aus vorhandenen Datenbeständen entnommen werden.

4.5.6 Implementationsanalyse

Nicht erforderlich.

4.5.7 Kosten-Nutzen-Analyse

Es ist zu erwarten, dass in diesem Projekt keine durchgängig aussagekräftige Kosten-Nutzen-Analyse vorgelegt werden kann. Unter bestimmten einschränkenden Annahmen und Bedingungen lassen sich aber voraussichtlich Abschätzungen für die fiskalischen Nettoeffekte machen.

Das Hauptproblem besteht darin, die Kosten von härteren Sanktionen, falls sie überhaupt existieren, zu ermitteln. Bei Gefängnisstrafen fallen erhebliche Kosten an, die gegebenenfalls auch näherungsweise quantifiziert werden könnten. Offensichtlich würden diese Kosten auch mit der Härte der Strafe variieren. Angesichts der Datenlage ist es aber wahrscheinlich, dass

das zu definierende Standardvergehen eine Ordnungswidrigkeit darstellt, die mit Geldbußen und nicht mit Haftstrafen geahndet wird. Bei Geldbußen haben härtere Sanktionen positive fiskalische Effekte, so dass zunächst einmal keine Kosten definiert werden können. Dies bedeutet, dass eine eingeschränkte Kosten-Nutzen-Analyse für dieses Projekt lediglich die positiven fiskalischen Effekte abschätzen würde. Dies könnte so geschehen, dass die ermittelte Minderung in der Wahrscheinlichkeit für eine Wiederholungstat bei einer bestimmten Variation der Härte der Strafe für ein Erstvergehen multipliziert wird mit den durchschnittlichen fiskalischen Erträgen, die mit einer Verhinderung weiterer Vergehen einhergeht. Allerdings gestaltet sich die Abschätzung der fiskalischen Erträge schwieriger als in den meisten anderen Projekten, solange Wiederholungstäter über alle Branchen bzw. Berufe betrachtet werden. Es müssten dann zur Ermittlung des zusätzlichen Steueraufkommens und der zusätzlichen Beiträge in die Sozialversicherungssysteme durchschnittliche Löhne und Gehälter in vielen Branchen verwendet werden. Eine verlässliche Abschätzung der fiskalischen Effekte wird dadurch erschwert. Als weitere Erträge müssten die zusätzlich generierten Einnahmen aus Bußgeldern berücksichtigt werden.

Eventuell könnten in einer Kosten-Nutzen-Analyse auch Kosten berücksichtigt werden, die bei der Durchsetzung der Sanktionen entstehen. Beispielsweise ist daran zu denken, dass bei härteren Bußgeldern insgesamt mehr Einsprüche erfolgen. Die Bearbeitung dieser Einsprüche verursacht zusätzliche Kosten. Aus ProFiS könnte dazu zusätzlich zur Höhe des Bußgeldes auch die Information entnommen werden, ob ein Einspruch erfolgt ist. Die zusätzlichen Kosten könnten dann abgeschätzt werden, indem die Veränderung in der Wahrscheinlichkeit für einen Einspruch mit den durchschnittlich pro Einspruch entstehenden zusätzlichen Personalkosten multipliziert wird.

Es ist zu erwarten, dass die Kosten-Nutzen-Analyse in diesem Projekt zu dem Ergebnis kommen wird, dass härtere Sanktionen positive fiskalische Effekte nach sich ziehen. Interpretationen dieses Ergebnisses sind mit großer Vorsicht vorzunehmen, denn eine unbeschränkte Verschärfung von Sanktionen ist kein realistisches Szenario. Darüber hinaus würde ein derartiges Vorgehen vermutlich zu negativen Effekten auf die Akzeptanz der FKS und des Rechtssystems insgesamt führen. Derartige Wirkungen von schärferen Sanktionen sind in einer Kosten-Nutzen-Analyse nicht ohne weiteres zu berücksichtigen und würden bei dem beschriebenen Vorgehen in jedem Fall außer Acht gelassen.

4.5.8 Aufwand

Der Aufwand besteht lediglich in der Gewinnung der Daten mit Hilfe der Datenbank ProFiS. Daher ist dieses Projekt als sehr kostengünstig einzustufen.

4.5.9 Besondere Probleme bei der Durchführung des Projekts

Die Evaluation beruht auf einer selektiven Stichprobe, daher sind keine generellen Aussagen über die präventiven Wirkungen von Strafen in der Gesamtpopulation aller potentiellen Anbieter oder Nachfrage von Schwarzarbeit möglich. In ProFiS sind keine Informationen hinterlegt, die eine exakte Identifikation der geahndeten Vergehen ermöglichen. Allerdings können

gewisse Kategorien (z.B. Leistungsmissbrauch) gebildet werden. Bei Fällen von Leistungsmissbrauch könnte auch eine Verknüpfung mit Schadenssummen erreicht werden. Bei Bußgeldern ist die Höhe des Bußgeldes hinterlegt. Bei Straftaten ist das Strafmaß hingegen nicht zuverlässig aufgeführt. Hier wäre eine Aktenauswertung notwendig, was jedoch aus Gründen des zusätzlichen Aufwands vermieden werden sollte. Die Datenbestände reichen bis 1.1.2004 zurück, für den Bereich des Leistungsmissbrauchs könnte auf ältere Daten der BA zurückgegriffen werden.

Das Projekt beruht auf der Annahme, dass die Wahrscheinlichkeit, mit der bereits einmal auffällig gewordene Personen ins Blickfeld der FKS geraten, nicht von der Strenge der verhängten Strafen abhängt. Nach Einschätzung der Abteilung FKS erfolgen jedoch generell keine gezielten Prüfungen von Personen, die bereits wegen Schwarzarbeit oder illegaler Beschäftigung auffällig geworden sind.

4.5.10 Mögliche Varianten

Neben der Analyse der Wirkung von Strafen könnte sich eine Stichprobe von Wiederholungstätern auch dazu eignen, die Wirkung präventiver Maßnahmen zu überprüfen. So könnte – ähnlich wie im Projekt Leistungsempfänger – einem zufällig ausgewählten Teil der Personen Informationen unterschiedlicher Art überreicht werden. Die Gruppe der bereits einmal aktenkundig gewordenen Personen ist insofern interessant, als es hier um einen Ausschnitt der Bevölkerung geht, die vermutlich eine relativ hohe Neigung zur Begehung von Schwarzarbeit hat.

4.5.11 Erwartete Ergebnisse des Evaluationsprojekts

- Bislang ist unbekannt, inwieweit einmal durch die FKS erfasste Personen wiederholt zu Beanstandungen Anlass geben. Dieses Projekt würde hierüber auf repräsentativer Basis Fakten etablieren. Das Ausmaß von Wiederholungstaten könnte dabei hinsichtlich einer Reihe von Charakteristika der betreffenden Personen bzw. Unternehmen aufgegliedert werden. Dies könnte nicht zuletzt dazu helfen, Risikoprofile für die präventive Arbeit der FKS zu entwickeln.
- Als kausale Wirkung wird hier der Effekt von Strafen auf die Neigung zu Wiederholungstaten ermittelt. Die Ergebnisse geben Anhaltspunkte dafür, inwieweit eine intensiverte Ahndung abschreckend auf die Begehung von Schwarzarbeit wirkt. Obwohl dieses Ergebnis nicht mit einer eindeutigen Kosten-Nutzen-Abwägung verbunden werden kann, lassen sich vermutlich Empfehlungen für die Rechtsgestaltung ableiten.

5 Schlussfolgerungen

Bisher liegen auch international nur wenige quantitative Wirkungsanalysen zur Frage der Wirksamkeit von Maßnahmen gegen Schwarzarbeit oder Steuer- und Abgabenhinterziehung vor. Sofern sie durchgeführt worden sind, betreffen sie in der Regel die Vereinigten Staaten, so dass ihre Ergebnisse kaum auf deutsche Verhältnisse übertragen werden können. Die Wirksamkeit von Maßnahmen zur allgemeinen Verbrechensbekämpfung ist häufiger analysiert worden, allerdings beruhen diese Studien kaum jemals auf Daten, die dem Prozess der Aufklärung und Ermittlung entstammen. Mit einer Evaluation der Maßnahmen der FKS auf der Basis von Prüfungsdaten würde also in mehrfacher Hinsicht Neuland betreten.

Bei der Informationserhebung zu diesem Projekt zeigte sich mehrfach, dass schon dadurch zusätzliche Kenntnisse über das Ausmaß von Schwarzarbeit und die Anwendung der Maßnahmen der FKS gewonnen werden könnten, wenn die vorhandenen Datenbestände systematisch ausgewertet würden. Entsprechend sollte ein Evaluationsvorhaben unabhängig von der Fokussierung auf bestimmte Branchen, Begehungsformen oder Zeiträume die Gesamtheit der Daten deskriptiv auswerten.

Da diese Datenbestände jedoch nicht repräsentativ für die Grundgesamtheit der Arbeitsmarkt-beteiligten sind, sollte eine quantitative Evaluation der Tätigkeit der FKS im Regelfall auf Daten beruhen, die im Zuge des Ermittlungsprozesses unter besonders kontrollierten Umständen, in der Regel im Kontext von Schwerpunktprüfungen, erhoben werden. Diese Daten sollten dann mit weiteren Daten der FKS verknüpft werden, insbesondere mit den Ergebnissen der Ahndung. Auf der Basis der in diesem Bericht vorgeschlagenen spezifischen Erhebungen ließen sich nicht nur Evaluationen durchführen, sondern es ergäbe sich zugleich ein repräsentatives Bild der Schwarzarbeit in ausgewählten Bereichen. Damit würden also erstmals für Deutschland verlässliche Informationen über den Umfang der Schwarzarbeit in den von den Evaluationsprojekten erfassten Bereichen gewonnen werden.

Hauptziel der beschriebenen Evaluationsprojekte ist es, verlässliche Evidenz für die Wirksamkeit der Maßnahmen der FKS auf den Umfang der Schwarzarbeit bzw. der legalen Beschäftigung zu schaffen. Für die erfassten Bereiche könnte damit geklärt werden, ob die von der FKS ergriffenen Maßnahmen den Umfang der Schwarzarbeit vermindern und ob dadurch zusätzliche legale Beschäftigung geschaffen wird. Die einzelnen Projekte sind so konzipiert, dass darüber hinaus auch die Größe der Effekte gemessen werden kann. Bei Durchführung der Evaluationsprojekte kann demnach Aufschluss darüber erwartet werden, in welchem Umfang die Schwarzarbeit in Deutschland in den betrachteten Bereichen zurückgedrängt wird, wenn der Umfang der Maßnahmen der FKS um ein bestimmtes Maß variiert wird. Auf der Quantifizierung dieser Effekte könnte eine Kosten-Nutzen-Analyse aufbauen. Diese würde darüber Aufschluss geben, ob sich der Maßnahmeneinsatz der FKS gegen Schwarzarbeit aus fiskalischer Sicht lohnt oder nicht.

Die Probleme bei der Durchführung von Evaluationsprojekten sind vielfältig. Dies hat die Evaluation der Maßnahmen der FKS mit Evaluationsvorhaben in anderen Bereichen gemein. Aufwand entsteht vor allem bei der Erhebung einer Datenbasis, denn die besondere Kontrol-

le der Umstände der Datenerhebung, die zur Sicherstellung der Repräsentativität der Datenbasis notwendig ist, bedeuten einen erheblichen Eingriff in die ansonsten üblichen Arbeitsabläufe innerhalb der FKS. Falls diese Probleme als zu schwerwiegend eingeschätzt werden, sollten andere Indikatoren (legale Beschäftigung oder das Rechtsbewusstsein gemäss repräsentativer Umfragen) verwendet werden.

Des Weiteren verlangen einige der vorgeschlagenen Projekte, dass der Umfang der zu evaluierenden Maßnahmen auf kontrollierte Art und Weise, nämlich durch eine Zufallsauswahl, zwischen Standorten oder Prüfobjekten variiert wird. Auch dies greift sehr weitgehend in die Arbeitsabläufe an den Standorten der FKS ein. Wesentlich ist hier eine sorgfältig konzipierte Kommunikationsstrategie, die Akzeptanzprobleme der Projekte innerhalb der FKS minimieren kann. Zugleich sollte die Umsetzung der zusätzlich durchgeführten Maßnahmen durch eine Implementationsanalyse dokumentiert werden. Diese Projekte sollten am besten nur an einer begrenzten Auswahl von Standorten und nicht flächendeckend durchgeführt werden, da letzteres die Kommunikation zu den beteiligten Standorten erschweren würde. Trotz dieser Probleme sind es gerade die Projekte mit zufällig erzeugter Variation, die die präzisesten und aufschlussreichsten Ergebnisse erwarten lassen.

6 Projektteam

Priv.-Doz. Dr. Bernhard Boockmann (Leitung)

Tel.: 0621/1235-156

Fax: 0621/1235-225

e-mail: boockmann@zew.de

Dipl.-Volksw. Johannes Rincke, ZEW

Tel.: 0621/1235-217

Fax: 0621/1235-223

e-mail: rincke@zew.de

Dr. Friedrich Heinemann, ZEW

beratend tätig:

Prof. Dr. Michael Lechner, Universität St. Gallen, IZA und ZEW

7 Literaturverzeichnis

- Ammermüller, A.; B. Boockmann, A. Garloff, A. Kuckulenz und A. Spermann (2003), Die ZEW-Erhebung bei Zeitarbeitsbetrieben - Dokumentation der Umfrage und Ergebnisse von Analysen, ZEW Dokumentation Nr. 03-07.
- Bundesagentur für Arbeit (2005), Merkblatt Beschäftigung ausländischer Arbeitnehmer aus Staaten außerhalb der Europäischen Union im Rahmen von Werkverträgen in der Bundesrepublik Deutschland, Stand Januar 2005, Bundesagentur für Arbeit, Nürnberg.
- Deutscher Bundestag (Hrsg.) (2005), Zehnter Bericht der Bundesregierung über die Auswirkungen des Gesetzes zur Bekämpfung der illegalen Beschäftigung – BillBG --, Unterrichtung durch die Bundesregierung, Drucksache 15/5934 vom 20.07.2005
- Di Tella, R. und E. Schargrodsky (2004), Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack, *American Economic Review*, 94, 115-33.
- Eide, E. (1999), *Economics of Criminal Behavior*, in: Bouckaert, Boudewijn; Gerrit De Geest (2000), *Encyclopedia of Law and Economics* (Edward Elgar and the University of Ghent)
- Feld, L.P. und C. Larsen (2005), Black Activities in Germany in 2001 and in 2004, A Comparison Based on Survey Data, Rockwool Foundation, Kopenhagen.
- Finanzkontrolle Schwarzarbeit (2005), Ergebnisse 2004, Köln, http://www.zoll.de/e0_downloads/f0_dont_show/fks_ergebnisse.pdf.
- Frey, B.; L.P. Feld (2002), Deterrence and Morale in Taxation: An Empirical Analysis, CESifo Working Paper No. 760.
- Gürtzgen, N. (2004), Makroevaluation mit berechenbaren allgemeinen Gleichgewichts-Modellen, in: T. Hagen, A. Spermann, Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung, ZEW Wirtschaftsanalysen, Bd. 74, Baden-Baden , S. 98-109.
- Heckman, J. J.; Lalonde, R.; J. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, In: *Handbook of Labor Economics*, Volume III, Hrsg. O. Ashenfelter und D. Card, Elsevier, Amsterdam.
- Jacobebbinghaus, P. (2004), Steuer-Transfer-Mikrosimulation, in: T. Hagen, A. Spermann, Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung, ZEW Wirtschaftsanalysen, Bd. 74, Baden-Baden, S. 73-83.
- KfW (2005), KfW-Gründungsmonitor 2005, Jährliche Analyse von Struktur und Dynamik des Gründungsgeschehens in Deutschland, Frankfurt, http://www.kfw.de/DE/Research/KfW-Grndun29/Gruendungsmonitor_2004.pdf.
- Körner, M. und Strotmann, H. (2005), Steuermoral – Das Spannungsfeld von Freiwilligkeit der Steuerzahlung und Steuerhinterziehung, *Monatesbericht des BMF*, März 2005, 49-57.
- Levitt, S.D. (1997), Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime, *American Economic Review*, 87, 270-90.

- Pedersen, S. (2003), *The Shadow Economy in Germany, Great Britain and Scandinavia: A Measurement Based on Questionnaire Surveys*, Rockwool Foundation, Kopenhagen.
- Schneider, F. und D. Enste (2000), *Schattenwirtschaft und Schwarzarbeit*, Oldenbourg.
- Slemrod, J., M. Blumenthal und C. Christian (2000), *Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota*, *Journal of Public Economics* 79, 455–83.
- Slemrod, J. und S. Yitzhaki: *Tax Avoidance, Evasion, and Administration*, in Auerbach, A. and M. Feldstein (eds.) *Handbook of Public Economics*, 3, 2002, Amsterdam: Elsevier Science B.V. Ch. 22, 1423-1470.
- Smith, J.A. (2000), *A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies*, Manuskript, University of Western Ontario.
- Torgler, B. (2004a), *Tax morale in Asian countries*, *Journal of Asian Economics*, 15, 237-266.
- Torgler, B. (2004b), *Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland*, *Economics of Governance*, 5, 235-253.
- Torgler, B. (2005), *Tax Morale in Latin America*, *Public Choice*, 122, 133-157.
- Torgler, B.; F. Schneider (2005), *Attitudes Towards Paying Taxes in Austria: An Empirical Analysis*, *Empirica*, 32, 231-250.
- Trumbull, W.N. (1989), *Estimations of the Economic Model of Crime Using Aggregate and Individual Level Data*, *Southern Economic Journal*, 56, 423-39.
- Wilke, R. und H. Winterhager (2004), *Datenquellen als Voraussetzung für die Evaluationsforschung*, in: T. Hagen, A. Spermann, *Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung*, *ZEW Wirtschaftsanalysen*, Bd. 74, Baden-Baden , S. 98-109.

Anhang: Überblick über die Evaluationsmethoden

Für eine Wirkungsanalyse kommt grundsätzlich eine Vielzahl von unterschiedlichen Verfahren in Frage. Als Evaluationsstudien verstehen wir in diesem Bericht empirische Studien, die auf der Basis von geeignet definierten Indikatoren statistische Aussagen über den kausalen Zusammenhang von Maßnahmen und Ergebnissen treffen. Neben solchen quantitativen Analysen sind auch qualitative Analysen möglich und sinnvoll. Hierbei geht es darum, das Untersuchungsfeld zu beschreiben und die Perzeption der Beteiligten zu erfassen. In vielen derzeit durchgeführten Evaluationsprojekten, wie z.B. der Evaluation der Hartz-Gesetze, gehen qualitative und quantitative Untersuchungsschritte miteinander einher. Dies wäre auch im Fall einer Evaluation der Bekämpfung der Schwarzarbeit eine sinnvolle Vorgehensweise. Dennoch wird im Rahmen dieses Berichts fast ausschließlich auf die Machbarkeit quantitativer Untersuchungen eingegangen, weil die Machbarkeit in diesem Bereich weitaus stärker in Zweifel steht als die Realisierbarkeit von qualitativen Studien.

Unter den quantitativen Studien kann man ex-post-Evaluationen von ex-ante-Evaluationen unterscheiden. Letztere beruhen auf angenommenen Regelmäßigkeiten des Verhaltens, die meistens durch ein theoretisches Verhaltensmodell gestützt werden. Sie werden vor allem dazu genutzt, die wahrscheinlichen Effekte einer Maßnahme vor ihrer Einführung abzuschätzen.¹ Diese Methoden werden üblicherweise vor der Verabschiedung gesetzlicher Maßnahmen eingesetzt, oder sie dienen zu einer ersten Abschätzung ihrer Effekte. Sie sind allerdings im Vergleich zu den auf realisiertem Verhalten beruhenden ex-post-Evaluationen mit größerer Unschärfe behaftet. Ferner eignen sie sich nicht für sehr detaillierte Analysen, bei denen zwischen in ihrer Wirkung ähnlichen Maßnahmen unterschieden werden soll. Sie kommen daher für die in Frage stehende Evaluation von Maßnahmen der FKS kaum in Betracht. Jedoch könnten ex-ante-Evaluationen zur Bewertung künftiger gesetzlicher Maßnahmen im Bereich der Schwarzarbeit verwendet werden.

Das Evaluationsproblem lässt sich durch das folgende Gedankenexperiment beschreiben. Es gelte, die Wirkung einer Maßnahme zu bewerten, deren Durchführung oder Nichtdurchführung einfach durch eine 0-1-Variable beschrieben werden kann. In jedem der beiden Zustände hat jedes Individuum eine Verhaltensweise. Da die Maßnahme nicht gleichzeitig durchgeführt und nicht durchgeführt werden kann, wird sich immer nur eine der beiden Verhaltensweisen realisiert. Potentiell gibt es jedoch beide Verhaltensweisen, und die Wirkung der Maßnahme ist die Differenz zwischen der realisierten und der nicht realisierten der beiden potentiellen Verhaltensweisen.

Die nicht realisierte Verhaltensweise kann nicht beobachtet werden, und der Kern des Evaluationsproblems besteht darin, hierfür glaubwürdige Approximationen zu finden. Wie in der medizinischen Forschung benutzt man dazu oft das Konzept der Kontrollgruppe. Hierbei wird die Maßnahme nur bei einem Teil der Personen angewendet, und man vergleicht das Verhalten der von der Maßnahme Betroffenen und nicht Betroffenen. Üblicherweise geschieht der

¹ Mikrosimulationsmodelle werden in Jacobebbinghaus (2004) und Allgemeine Gleichgewichtsmodelle in Gürtzgen (2004) dargestellt.

Vergleich durch die Bildung eines Mittelwerts in der Maßnahme- und der Kontrollgruppe. Die Auswahl der Personen kann dabei durch die Evaluation selbst gesteuert sein, indem diese durch einen Zufallsprozess festlegt, welche Individuen an der Maßnahme teilnehmen. In diesem Fall spricht man von einer experimentellen Vorgehensweise.

Auf der anderen Seite kann man die Maßnahme- und Kontrollgruppe auch nachträglich den Daten entnehmen. So könnte die Maßnahme darin bestehen, dass die FKS in einer Branche ein bestimmtes Instrument der Prävention angewendet hat. Bei der Evaluation vergleicht man dann Personen aus der Branche, in der das Instrument angewendet wurde, mit Personen, die in anderen Branchen tätig sind. Hierbei ist jedoch zu berücksichtigen, dass die Zuordnung von Personen und Maßnahmen selten zufällig erfolgt. Im Beispiel könnte die FKS die Maßnahme aus Gründen der rationellen Verwendung ihrer Mittel in einer Branche angewendet haben, die ein besonders hohes Niveau der Schwarzarbeit aufweist. Der Vergleich der tatsächlich geleisteten Schwarzarbeit könnte dann ergeben, dass in der Branche, in der die Maßnahme durchgeführt wurde, trotzdem mehr Schwarzarbeit geleistet wird als in der Vergleichsbranche, was dann fälschlich als Misserfolg der Prävention interpretiert wird.

In diesen Fällen kommt es darauf an, aus den vorhandenen Daten eine unverzerrte Kontrollgruppe zu ziehen, die der Gruppe der Teilnehmer möglichst ähnlich ist. Man spricht dann von einem quasi-experimentellen Forschungsdesign. In der folgenden Darstellung werden experimentelle und quasi-experimentelle Forschungsdesigns dargestellt und auf das Problem der Evaluation der Tätigkeit der FKS bezogen.

Kontrollierte Experimente

Ein prinzipiell gut geeignetes Vorgehen bei der Identifikation kausaler Effekte ist das kontrollierte Experiment. Hierbei wird der Teilnahmestatus durch einen Zufallsprozess festgelegt. Hiermit ist sichergestellt, dass die Selektion in eine der beiden Gruppen nicht aufgrund beobachteter oder nicht beobachteter Merkmale der Untersuchungseinheit erfolgt. Der Maßnahmeeffekt wird dann z.B. als Mittelwert der Ergebnisse in der Maßnahme- und Kontrollgruppe gemessen. Damit der Effekt auf die Grundgesamtheit übertragen werden kann, muss die Gesamtheit der Individuen in der Maßnahme- und Kontrollgruppe einen repräsentativen Ausschnitt aus der Grundgesamtheit darstellen.

Ein Beispiel für ein experimentelles Vorgehen wird in Blumenthal et al. (2002) beschrieben. Hierbei ging es um ein Experiment, das 1994 von der Finanzverwaltung im US-Bundesstaat Minnesota durchgeführt wurde. Aus der Grundgesamtheit aller Steuerpflichtigen wurden drei Zufallsstichproben gezogen. Personen in den ersten beiden Gruppen wurden unterschiedlich lautende Briefe zugesandt, die sie an ihrer Pflicht zur korrekten Steuererklärung ermahnten. Die dritte Gruppe erhielt keinen Briefe und diente insofern als Kontrollgruppe. Die Wirkung wurde an den Steuererklärungen der Personen in den drei Gruppen festgestellt. Dazu wurden die Steuererklärungen im Jahr vor dem Experiment und im Jahr nach der Zusendung der Briefe ausgewertet. Die Erwartung war, dass Personen in den ersten beiden Gruppen dazu neigen würden, im Vergleich zum Vorjahr ein höheres Einkommen anzugeben, während sich bei der Kontrollgruppe keine Änderung abzeichnen würde. Diese Erwartung wurde allerdings durch die empirischen Ergebnisse nicht bestätigt. Ein ähnliches Vorgehen wird von Torgler (2004b) gewählt.

Der Vorteil experimenteller Verfahren gegenüber der im folgenden Abschnitt beschriebenen nicht-experimentellen Vorgehensweise besteht darin, dass ersteren ein leicht durchführbares und nachvollziehbares Forschungsdesign zugrunde liegt. Dass Experimente trotz ihrer Vorteile nicht weiter verbreitet sind, hängt nicht allein mit dem erheblichen Aufwand zusammen, den sie innerhalb der Verwaltungen, die die Experimente umsetzen sollen, verursachen. Daneben gibt es oft rechtliche Restriktionen, die ein soziales Experiment nicht erlauben. Beispielsweise ist es nicht durchführbar, beim Niveau von Strafen oder Bußgeldern nach Maßnahme- und Kontrollgruppe zu unterscheiden. Damit scheidet der gesamte Bereich der Ahndung als Anwendungsfeld sozialer Experimente aus.

Auch wenn Experimente bei den Maßnahmen der Prävention grundsätzlich ein gut geeignetes Verfahren darstellen, ist nicht jede experimentelle Studie tatsächlich hilfreich bei der Feststellung kausaler Effekte. Ein üblicher Einwand gegen experimentelle Studien ist, dass meistens Freiwilligkeit bei der Teilnahme am Experiment gegeben sein muss. Freiwillige Teilnehmer sind aber oft keine Zufassauswahl aus der Bevölkerung. In unserem Fall greift dieser Einwand nicht, weil die Maßnahmen nicht mit Zustimmung der Betroffenen durchgeführt werden. Daher sind Experimente gerade in diesem Bereich besonders sinnvoll.

Ein anderer, für die durchzuführende Studie gewichtigerer Einwand ist, dass Experimente in die Abläufe in der Verwaltung eingreifen. So könnten Mitarbeiter der Verwaltung die Evaluation nicht mittragen und die im Rahmen des Experiments vorgesehenen Maßnahmen nicht ordnungsgemäß durchführen. Die Maßnahme, die innerhalb des kontrollierten Experiments durchgeführt wird, ist in diesem Fall nicht mehr die gleiche wie diejenige, die ursprünglich evaluiert werden sollte.: “[Experimental methods] are less effective when evaluating ongoing programs in part because they appear to disrupt established bureaucratic procedures. The threat of disruption leads local bureaucrats to oppose their adoption. To the extent that programs are disrupted, the program evaluated by the method is not the ongoing program that one seeks to evaluate.” (Heckman et al. 1999)

In der Praxis kann es deshalb ratsam sein, von einer zufälligen Zuordnung von Maßnahmen und durchführenden Verwaltungen abzusehen, wenn Evidenz dafür vorliegt, dass bestimmte Verwaltungen weniger bereit sind als andere, die Maßnahme für den Zweck der Evaluation wie geplant durchzuführen. Damit ist andererseits der Zufallscharakter des Experiments und damit die Abwesenheit von Selektionseffekten gefährdet.

Nicht-experimentelle Verfahren

Für bestimmte Bereiche wie die Ahndung scheidet experimentelle Verfahren aus grundsätzlichen Erwägungen aus. In anderen Bereichen können andere Verfahren aus Kostenerwägungen besser als experimentelle Verfahren für eine Evaluation geeignet sein. Im Folgenden werden einige geläufige nicht-experimentelle Verfahren verwendet. Die ersten drei Verfahren beruhen auf der Ausnutzung von Variation in den zu prüfenden Maßnahmen zwischen Gruppen von Individuen und/oder über die Zeit.

- Schon genannt wurde der Mittelwertvergleich zwischen von der Maßnahme Betroffenen und nicht Betroffenen zum gleichen Zeitpunkt. Der Nachteil ist hier, dass sich die verglichenen Personen hinsichtlich einer Vielzahl von Merkmalen unterscheiden, von denen nicht alle erfasst werden können. Das Ergebnis des Vergleichs spiegelt daher sowohl den Maßnahmeneffekt als auch die Unterschiede zwischen beiden Personengruppen wider.
- Eine zweite Möglichkeit ist der Vorher-Nachher-Schätzer, bei dem die Indikatoren für die gleichen Individuen vor und nach der Einführung der Maßnahme verglichen werden. In diesem Fall überlagern zwar die Unterschiede zwischen den Individuen den Vergleich nicht, auf der anderen Seite können die Änderungen zwischen den Zeitpunkten auch von anderen Umständen als der zu evaluierenden Maßnahme verursacht sein; ein Beispiel hierfür sind konjunkturelle Ursachen, ein anderes Änderungen im gesetzlichen Rahmen.
- Der Differenzen-von-Differenzen-Schätzer kombiniert beide Ansätze und verspricht daher, robustere Ergebnisse zu liefern. Eine Gruppe von Personen, die der Maßnahme unterworfen ist, wird hierbei verglichen mit einer anderen Gruppe, für die sich nichts geändert hat. Unter der Annahme, dass sich die eben erwähnten Änderungen über die Zeit im gleichen Verhalten zwischen den beiden Gruppen ausdrücken, kann der Unterschied in den zeitlichen Unterschieden zwischen den beiden Gruppen als der kausale Effekt der Maßnahme interpretiert werden.

Allen drei Schätzern ist die Anforderung gemein, dass die Variation nicht ihrerseits durch die Realisation der Ergebnisvariable hervorgerufen sein darf. Dies bedingt teilweise ein auf den ersten Blick etwas umständliches Vorgehen, wie aus dem folgenden Beispiel deutlich wird.

Als Beispiel sei hier die Studie von Di Tella und Schargrodsky (2004) erwähnt. Hierin ging es um den Effekt von Polizeipräsenz auf die Häufigkeit von Autodiebstählen. Ein Differenzen-in-Differenzen-Vorgehen könnte darauf beruhen, dass die Polizeipräsenz nur in bestimmten Bezirken erhöht wird. Der Effekt ist dann die gemessene Veränderung in der Zahl der Diebstähle in diesen Bezirken abzüglich der Veränderung in den Bezirken mit unveränderter Polizeipräsenz, wobei die Veränderung zwischen zwei Zeitpunkten vor und nach der Verstärkung der Polizeipräsenz gemessen wird. Das Problem ist, dass die Polizeipräsenz normalerweise dort verstärkt werden dürfte, wo das Problem der Autodiebstähle besonders akut ist. Di Tella und Schargrodsky verwenden jedoch Variation in der Polizeipräsenz, die aus anderen Gründen zustande kam. Im Jahr 1994 wurde in Buenos Aires ein Sprengstoffanschlag auf eine jüdische Kultureinrichtung verübt, bei dem 85 Menschen ums Leben kamen. In der Folge wurde die Polizeipräsenz vor allen jüdischen Einrichtungen erheblich verstärkt. Da das Vorkommen von Autodiebstählen nicht mit der Nähe zu jüdischen Einrichtungen zusammenhängt, entsteht so Variation in der Maßnahmeintensität, die nicht mit der Ergebnisvariablen korreliert ist. Die Verfasser weisen einen hohen Effekt der Polizeipräsenz auf die Zahl der Autodiebstähle nach.

Der Differenzen-von-Differenzen-Schätzer setzt wiederholte Beobachtungen der Ergebnisvariable vor und nach Änderungen in der Maßnahme voraus. Wenn Instrumentvariablen verfügbar sind, reicht unter Umständen auch ein einfacher Querschnitt von Beobachtungen aus. Instrumentvariablen für eine Maßnahme sind solche Variablen, die zwar die Maßnahme be-

einflussen, aber nicht von der jeweiligen Ergebnisvariablen beeinflusst sind. Die Instrumentvariable kann z.B. eine Null-Eins-Variable sein, gemäß deren Ausprägung die Individuen von der Maßnahme erfasst oder nicht erfasst werden. Wenn die Ausprägung der Instrumentvariablen nicht von der Ergebnisvariablen beeinflusst wird, kommt die dadurch erzeugte Zuordnung hinsichtlich der Ergebnisvariablen einer zufälligen Zuordnung nahe. Der Kausaleffekt wird dann unter Verwendung der Instrumentvariablen durch ein mehrstufiges Regressionsverfahren errechnet.

Ein Beispiel hierfür ist die Studie von Levitt (1997), in der es um die Wirkung einer Erhöhung des Personalbestands der Polizei auf die Kriminalitätsrate in den US-Bundesstaaten geht. Die verwendete Instrumentvariable ist eine Variable, die in Vormonaten vor Gouverneurswahlen den Wert eins annimmt und null andernfalls. Diese Variable weist einen engen Zusammenhang zur Polizeistärke auf, denn in Wahljahren versuchen die Gouverneure, die innere Sicherheit zu verbessern. Von dieser Variable darf man zweitens aber annehmen, dass sie nicht ihrerseits von der Kriminalitätsrate beeinflusst wird (es sei denn, es würden wegen regelmäßig in Fällen hoher Kriminalität vorzeitige Neuwahlen angesetzt). Damit erfüllt die Variable die Bedingungen für eine Instrumentvariable. Im Endeffekt wird nur diejenige Variation in der Polizeistärke verwendet, die sich auf den Wahlzyklus zurückführen lässt. Dies reicht im Fall von Levitt (1997) aber aus, um einen präventiven Effekt der Polizei nachzuweisen, insbesondere im Bereich der Gewaltkriminalität.

Neben den genannten Verfahren gibt es noch weitere, die in der Evaluation der Arbeitsmarktpolitik benutzt werden. So besteht das Matching-Verfahren als eines der geläufigsten Verfahren darin, für jeden Teilnehmer an einer Maßnahme einen „statistischen Zwilling“ zu finden, der möglichst ähnliche Eigenschaften besitzt, aber nicht an der zu evaluierenden Maßnahme teilgenommen hat. Dieses Verfahren ist nur in Fällen geeignet, in denen die Maßnahme auf der individuellen Ebene variiert werden kann. In diesen Fällen bietet sich aber bei der Evaluation der FKS-Maßnahmen ein experimentelles Design an. Dieses hat den Vorteil, dass die zufällige Zuordnung von Individuen in die Maßnahme- und Kontrollgruppe direkt gesteuert werden kann. Zudem müssen keine sonstigen Eigenschaften der Person zur Verfügung stehen, die im Matching-Verfahren zur Bildung von Vergleichspaaren benötigt werden. Die Präzision der Ergebnisse ist beim experimentellen Verfahren höher als beim Matching-Verfahren.

Multiple Maßnahmenwirkungen und multipler Maßnahmeneinsatz

Die Instrumente der FKS haben aus unterschiedlichen Gründen einen Einfluss auf den gemessenen Umfang der Schwarzarbeit. Falls eine Maßnahme der Prävention verstärkt angewendet wird, hat dies einen Abschreckungseffekt. Maßnahmen der Ermittlung und Prüfung haben vermutlich ebenfalls präventive Effekte, wenn sie angekündigt werden. Durch diese Maßnahmen und die damit verbundene Ahndung werden zugleich potentielle Nutzer oder Anbieter von Schwarzarbeit „aus dem Verkehr gezogen“, was das Niveau der Schwarzarbeit ebenfalls vermindert. Durch die erhöhte Ermittlungsaktivität werden jedoch auch mehr Fälle von Schwarzarbeit aktenkundig, ohne dass sich das Verhalten ändert (reporting bias).

Ein Beispiel: Wie wirkt die verstärkte Prüfung in einer Branche oder einer Region auf die in den Arbeitsstatistiken der FKS festgehaltenen Verstöße? In erster Linie dürften diese die

höhere Prüfungsintensität widerspiegeln und erst in zweiter Linie mögliche verhaltenslenkende Effekte. Dass Fälle aktenkundig werden und geahndet werden können, entspricht dem Auftrag der FKS. Die Zahl der durch zusätzliche Prüfungen ermittelten Verstöße ist daher durchaus ein interessanter Indikator. Im Zentrum der Analyse sollte jedoch stehen, ob die Tätigkeit der FKS verhaltenslenkend in dem Sinne wirkt, dass weniger Schwarzarbeit und mehr legale Arbeit geleistet wird. In jedem Fall sind diese beiden Fragen stets sorgfältig zu trennen.

Ob und wie diese Wirkungen voneinander zu trennen sind, muss jeweils im Einzelfall analysiert werden, wobei es entscheidend darauf ankommt, ob die unterschiedlichen Maßnahmen unabhängig voneinander eingesetzt werden können. Wenn z.B. zu präventiven Zwecken eine verstärkte Prüfung durchgeführt wird, dürften die Ergebnisse dieser Prüfung nicht verwendet werden, wenn ihre präventiven Wirkungen festgestellt werden sollen. Die Datenerhebung sollte vielmehr in einem getrennten Schritt durchgeführt werden, der von der erhöhten Prüfungsaktivität nicht beeinflusst ist.

Ein anderes Problem ist, dass sich der Einsatz anderer Instrumente von Region zu Region systematisch unterscheiden kann. Eine Region, die Präventionsinstrument A nutzt, setzt möglicherweise auch das Instrument B verstärkt ein. Erfasst man den Instrumenteneinsatz B nicht, so wird dessen Effekt fälschlich dem Instrument A zugeschrieben. Auch hier ist das experimentelle Vorgehen dasjenige, das am ehesten geeignet ist, Verzerrungen zu vermeiden.