

BIBB-FDZ
Daten- und Methodenberichte

Nr. 7 / 2010

**Kosten und Nutzen der betrieblichen
Berufsausbildung in Deutschland**
- Beschreibung der Datensätze für die
Jahre 2000 und 2007 -

Version 1.1

Holger Alda

Bundesinstitut für Berufsbildung
- Forschungsdatenzentrum -
Robert-Schuman-Platz 3
53175 Bonn

Telefon: 0228 / 107 - 20 41
Fax: 0228 / 107 - 20 20
E-Mail: fdz@bibb.de

www.bibb-fdz.de



Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland - Beschreibung der Datensätze für die Jahre 2000 und 2007

Autor: Holger Alda

Inhaltsverzeichnis

1. Einleitung	3
2. Die Datensätze zu Kosten und Nutzen der betrieblichen Berufsausbildung für die Jahre 2000 und 2007 im Überblick	4
3. Datengenese	7
3.1 Grundgesamtheit/ Untersuchungseinheiten (Stichprobengrundlage)	7
3.2 Erhebungsdesign	8
3.3 Ausschöpfung und Einhaltung des Stichprobenplans	9
4. Variablenspektrum und Datenaufbereitungen	12
4.1 Variablenspektrum	12
4.2 Korrekturen und Plausibilitätsprüfungen	14
4.3 Datenaufbereitungen im BIBB-FDZ	15
4.4 Imputation fehlender Werte	16
4.5 Gewichtung und Hochrechnung	21
5. Zusammenführung der jeweiligen Jahresdatensätze	23
6. Hinweise zur Datennutzung	25
7. FDZ-Angebot für die beiden Datensätze Kosten und Nutzen der Berufsausbildung für die Jahre 2000 und 2007	26
7.1 Arbeitshilfen	26
7.2 Datenzugangswege	27
7.3 Bestimmungen des Datenschutzes	27
7.4. Forschungspotential: Veröffentlichungen (Auswahl)	28
Anhang	30

1. Einleitung

Mit dem vorliegenden BIBB-FDZ Daten- und Methodenbericht werden die beiden Datensätze „Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland“ für die Jahre 2000 und 2007 dokumentiert, um eine eigenständige Auswertung der Daten durch die externe Wissenschaft zu ermöglichen. Beide Datensätze wurden von infas, Institut für angewandte Sozialforschung (Bonn), im Auftrag des Bundesinstituts für Berufsbildung (BIBB) erstellt. Die externe und nicht-kommerzielle Forschung kann über die Datenfernverarbeitung und/ oder Gastaufenthalte auf die Daten zugreifen¹. Da der Geneseprozess für die Daten der Jahre 2000 und 2007 ähnlich ist, entspricht dieser BIBB-FDZ Daten- und Methodenreport in Teilen der Nr. 4/2009.

Grundlage für alle Kosten-Nutzen-Erhebungen der betrieblichen Berufsausbildung ist ein Modell zur Erfassung betrieblicher Ausbildungskosten, wie es von der Sachverständigenkommission „Kosten und Finanzierung der beruflichen Bildung“ erstmals im Jahr 1974 vorgeschlagen wurde (zur Sachverständigenkommission etwa Beicht et al., 2004: 9 und 11). Eine ausführliche Darstellung der methodischen Vorgehensweisen zur Ermittlung und Berechnung von Kosten und Nutzen der betrieblichen Berufsausbildung befinden sich in den Abschlusspublikationen der entsprechenden Forschungsprojekte (Beicht et al., 2004 und Schönfeld et al., 2010). Die Lektüre dieser Publikationen ist daher insbesondere bei Analyseinteressen zu empfehlen, die sich besonders intensiv mit den Kosten- und Nutzenaspekten der betrieblichen Berufsausbildung beschäftigen².

In beiden Forschungsprojekten wurde eine schriftliche Zusatzerhebung bei nicht ausbildenden Betrieben zu einem potentiellen Nutzen der Berufsausbildung und zu den Gründen für einen Ausbildungsverzicht durchgeführt. Auf diese Daten wird im vorliegenden Handbuch *nicht* eingegangen³.

Das Daten- und Methodenhandbuch ist wie folgt gegliedert. Kapitel 2 stellt die zentralen Gemeinsamkeiten und Unterschiede sowie wesentliche Eigenschaften der beiden Untersuchungen zusammen. Zudem werden Unterschiede zu vorherigen Erhebungen (etwa aus dem Jahr 1991) deutlich. Den Prozess der Datengenese beschreibt das dritte Kapitel. In Kapitel 4 wird auf Datenaufbereitungen wie Plausibilitätsprüfungen, ein *Angebot* zur Imputation fehlender Werte sowie auf die Bildung der Gewichtung- bzw. Hochrechnungsfaktoren eingegangen.

¹ Nähere Informationen zu den Datenzugangswegen des BIBB-FDZ befinden sich auf den Internetseiten des BIBB-FDZ (<http://www.bibb-fdz.de>) in der Rubrik Datenzugang.

² Die beiden Datensätze erlauben in einem gewissen Umfang auch die empirische Behandlung anderer Auswertungsinteressen, da sie – neben der detaillierten Erfassung von Kosten- und Nutzenaspekten der betrieblichen Berufsausbildung – eine Vielzahl weiterer interessanter Merkmale für die betriebliche Ebene beinhaltet (vgl. den zusammenfassenden Variablenüberblick in Abschnitt 4.1)

³ Bei einem nachweislichen Interesse der externen Wissenschaft an diesen Datenbeständen kann das FDZ prüfen, ob eine Aufbereitung der Datensätze möglich ist.

Das fünfte Kapitel beschreibt, wie die über das BIBB-FDZ bereitgestellten Datensätze aufgebaut sind. Kapitel 6 geht auf die Handhabung der Datensätze und die Arbeit mit den Variablen der Kosten- und Nutzen-Erhebungen für die Jahre 2000 und 2007 ein. Das FDZ-Angebot der Arbeitshilfen zu den Kosten/Nutzen-Datensätzen wird im siebten Kapitel beschrieben.

Um die Arbeit Externer mit den Daten zu erleichtern, werden mit einer formlosen Begründung zusätzlich einige zusammenfassende Variablen der Modellkostenrechnungen des BIBB für eigene Auswertungen zur Verfügung gestellt. Den Beteiligten im BIBB, insbesondere G. Walden, U. Beicht, H. Pfeifer, F. Wenzelmann, G. Schönfeld und weiteren Auskunft gebenden bzw. Daten zur Verfügung stellenden Personen der Forschungsabteilung 2 des BIBB wird für ihre Unterstützung (und Geduld bei den notwendigen Erklärungen) gedankt.

2. Die Datensätze zu Kosten und Nutzen der betrieblichen Berufsausbildung für die Jahre 2000 und 2007 im Überblick

Die Befragungen zu den Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland haben das Ziel, für ausgewählte Ausbildungsberufe ein möglichst detailliertes Bild zu den Kosten betrieblicher Berufsausbildung zu gewinnen und zentrale Dimensionen des Ausbildungsnutzens zu berechnen.

Seit der grundlegenden Untersuchung der Sachverständigenkommission Kosten und Finanzierung der beruflichen Bildung aus dem Jahr 1974 (vgl. Sachverständigenkommission Kosten und Finanzierung der beruflichen Bildung 1974) erfolgte eine Erhebung der Kosten (und Nutzen) der betrieblichen Berufsausbildung in mehr oder minder regelmäßigen Abständen, teilweise als Aktualisierungen. Wenngleich es sich bei der Kosten-Nutzen-Erhebung für das Jahr 2000 im Prinzip um eine Wiederholungsbefragung des Jahres 1991 handelt, unterscheidet sie sich von dieser in zentralen Dimensionen:

- Für die Jahre 2000 und 2007 wurde – auf unterschiedliche Weise - die gesamte deutsche Wirtschaft sowie der öffentliche Dienst einbezogen, im Jahr 1991 nur die Ausbildungsbereiche Industrie und Handel sowie das Handwerk.
- Die Daten der Erhebungen seit dem Jahr 2000 beziehen sich auf Gesamtdeutschland, diejenigen von 1991 ausschließlich auf die alten Bundesländer.
- Seit dem Jahr 2000 werden pro Betrieb die Ausbildungskosten in einem ausgewählten Ausbildungsberuf erfasst, im Jahr 1991 dagegen die Gesamtkosten für alle im Betrieb angebotenen Ausbildungsberufe.
- Ab dem Jahr 2000 wird der betriebliche Nutzen der Berufsausbildung verstärkt berücksichtigt. So werden etwa repräsentative Daten zu durch Ausbildung eingesparten Rekrutierungskosten externer Fachkräfte erfasst.

Aufgrund dieser Unterschiede sind Vergleiche zwischen den Daten ab dem Jahr 2000 und den Daten aus dem Jahr 1991 (bzw. den auf dieser Basis erstellten Fortschreibungen) nur bedingt möglich. Auch zwischen den Erhebungen aus den Jahren 2000 und 2007 gibt es Unterschiede, die aber weniger gravierend sind als die bisher genannten (vgl. dazu Kapitel 5 in Schönfeld et al., 2010).

Erhebung für das Jahr 2000

Von Mitte Februar bis Mitte August 2001 befragte infas für das BIBB in 2.518 ausbildenden Betrieben die dort für die Berufsausbildung auskunftsfähigen Personen. Die Befragung setzte sich aus zwei Teilen zusammen. Der erste Teil zu den Kosten der betrieblichen Berufsausbildung wurde von geschulten Interviewern face-to-face durchgeführt. Der zweite Teil des Fragebogens mit Fragen zum betrieblichen Ausbildungsnutzen war als im Betrieb hinterlegter schriftlicher Selbstausfüller (Paper-and-Pencil-Methode, PAPI) angelegt, konnte jedoch auch direkt in Anschluss an Teil eins zusammen mit dem Interviewer ausgefüllt werden. Die überwiegende Mehrheit der Betriebe (88 Prozent) machte hiervon Gebrauch.

Die befragten Betriebe wurden über eine geschichtete Stichprobenziehung ausgewählt. Zunächst wurden 52 Ausbildungsberufe bestimmt, die das Gesamtspektrum der beruflichen Ausbildung in Deutschland abdeckten und typisch für die jeweiligen Ausbildungsbereiche sind⁴. Anschließend wurden durch die Kammern bzw. zuständigen Stellen in einer bundesweit gestreuten, aber aus Kostengründen regional konzentrierten Stichprobe etwa 13.000 Betriebe gezogen, die im Jahr 2000 in Deutschland mindestens eine Person in einem der einbezogenen Ausbildungsberufe ausgebildet haben. In einem vom BIBB entwickelten Stichprobenplan wurde die Zahl der pro Ausbildungsberuf und Kammerbezirk zu realisierenden Interviews festgelegt. Durch ein zweistufiges Gewichtungsverfahren einerseits auf Basis der Auszubildendenzahlen sowie andererseits auf Basis aller deutschen Ausbildungsbetriebe sind Schlüsse auf die jeweilige Grundgesamtheit möglich (alle deutschen Ausbildungsbetriebe bzw. alle Auszubildenden in den 52 ausgewählten Ausbildungsberufen).

Erhebung für das Jahr 2007

Die Feldphase der Erhebung lief vom 22. April bis zum 15. August 2008. In diesem Zeitraum befragte infas computerunterstützt (CAPI) in 2.986 ausbildenden Betrieben auskunftsfähige Personen zu Kosten- und Nutzenaspekten der Berufsausbildung. Daten zu beiden Aspekten wurden in face-to-face-Interviews (F2F) ermittelt. Im Unterschied zur Vorgängeruntersuchung erfolgte die Stichprobenziehung nicht über die Kammern, sondern über die Betriebsdatei der Bundesagentur für Arbeit (Stichtag 30. Juni 2006). Die daraus resultierenden Unterschiede auf der Datenebene zwischen der 2000er- und 2007er-Untersuchung werden im dritten Kapitel beschrieben.

⁴ Eine vollständige Auflistung dieser Ausbildungsberufe befindet sich im Anhang A1.

Andere zentrale Charakteristika der Kosten- und Nutzenerhebungen für die Jahre 2000 und 2007 sind ähnlich. Im Anhang A1 sind etwa die ausgewählten 52 bzw. 51 Ausbildungsberufe der 2000er- und 2007er-Erhebung aufgelistet. Für beide Auswertungsjahre gibt es außerdem je einen Design- und einen Hochrechnungsfaktor, wobei sich die Berechnungsweisen in beiden Erhebungsjahren aufgrund veränderter Stichprobengrundlagen geringfügig unterscheiden (vgl. dazu die Abschnitte 3.2 und 4.5). Tabelle 1 stellt die zentralen Eigenschaften der Kosten-Nutzen-Erhebungen der Jahre 2000 und 2007 gegenüber.

Tabelle 1: Zentrale Eigenschaften der Kosten-Nutzen-Erhebungen aus den Jahren 2001 und 2007

	Erhebung für das Jahr 2000	Erhebung für das Jahr 2007
Befragungs-ID	716	853
Befragungstitel	Kosten-Nutzen der betrieblichen Berufsausbildung im Jahr 2000	BIBB-Kosten-Nutzen-Erhebung 2007
Welle	Kosten-Nutzen betrieblicher Berufsausbildung	
Erhebungsjahr	2001	2008
Grundgesamtheit	alle Ausbildungsbetriebe in Deutschland (Kammerauswahl)	alle Betriebe in Deutschland (Betriebsdatei der BA)
Befragungseinheit	Ausbildungsbetriebe	
Rücklauf	2.518 (Rücklaufquote: 40,2 Prozent)	2.986 (Rücklaufquote 35,2 Prozent)
Nettozahl	6.257	8.907
Verfahren	persönliche Befragung	
Befragungsstil	Primärerhebung	
Erhebungsdesign	Längsschnittdesign (unverbundene Querschnitte als zeitliche Trendstudie)	
Kurzbeschreibung	Detaillierte Erfassung von Kosten und (monetärem und nicht monetär bewertetem) Nutzen der betrieblichen Berufsausbildung in Deutschland pro Auszubildendem und Ausbildungsjahr für die 52 (im Jahr 2000) bzw. 51 (im Jahr 2007) am stärksten besetzten Ausbildungsberufe.	
Befragungsart	Stichprobe(quantitativ)	
Anzahl Variablen	570	545
Gewichtung, Hochrechnung	Anpassungsgewichtung (Iterative Proportional Fitting) und Hochrechnungsvariable	
Forschungsfragen	Kosten der betrieblichen Berufsausbildung; Ausbildungsnutzen (monetär und nicht-monetär) für die Betriebe, jeweils pro Auszubildendem und Ausbildungsjahr für die am stärksten besetzten Ausbildungsberufe	
repräsentative Region	Ost- und Westdeutschland	
primäre Lebenszykluszuordnung	3: Ausbildung	
Feldphase	Februar 2001 bis August 2001	April 2008 bis August 2008
Interviewdauer in min	etwa 40	etwa 85
Kurztitel*	BIBB-CBS 2000	BIBB-CBS 2007

* CBS = Cost-Benefit-Survey

Auswahlverfahren	Zufallslistenauswahl der Kammern	Zufallsauswahl aus der Betriebsdatei der Bundesagentur für Arbeit mit nachgelagertem Screening
Datenaufbereitung	Datenbereinigungen; Gewichtung und Imputation; Anonymisierungen; Klassifizierungen	
Stichworte	Ausbildungskosten; Nutzen der Ausbildung für Betriebe; betriebliche Berufsausbildung; Ausbildungsbetrieb; Fachkräftebedarf; Nachwuchsbedarf und -sicherung; Personalentwicklung; Ausbildungsbereitschaft	
Datenzugangswege	Gastwissenschaftleraufenthalt; Datenfernverarbeitung	

3. Datengenese

Im Folgenden wird das in Kapitel 2 Gesagte bezüglich der Datenerhebungen der Forschungsprojekte zu Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland für die Jahre 2000 und 2007 konkretisiert. Zuerst wird auf die Grundgesamtheit bzw. die Untersuchungseinheiten der Studie eingegangen (3.1) und anschließend das Verfahren der Stichprobenziehung erläutert (3.2). Nach der Darstellung von Erhebungsdesign und Erhebungsmethode (3.3) gibt es noch Informationen zum Rücklauf der Befragung und der Erfüllung der Vorgaben des Stichprobenplans.

3.1. Grundgesamtheit / Untersuchungseinheiten (Stichprobengrundlage)

Im Gegensatz zur vorangegangenen Kosten/Nutzen-Befragung des Jahres 1991 wurden ab der Erhebung für das Jahr 2000 pro Betrieb nicht die Gesamtkosten für *alle* Ausbildungsberufe ermittelt. Vielmehr wurden Kosten und Erträge für jeweils einen ausgewählten Ausbildungsberuf differenziert erhoben (Beicht et al., 2004: 12 und Schönfeld et al., 2010: 21ff.).

Für das Jahr 2000 wurden 52 Ausbildungsberufe aus die gesamte Wirtschaft⁵ abdeckenden Ausbildungsbereichen (Industrie und Handel, Handwerk, Landwirtschaft, Freie Berufe sowie Öffentlicher Dienst) in eine Schichtungsmatrix überführt. Die ausgewählten Ausbildungsberufe sollten für den jeweiligen Ausbildungsbereich typisch sein⁶. Daneben fanden neue und (zur damaligen Zeit) zukunftssträchtige Berufe Berücksichtigung. Mit der geschilderten Auswahl werden für einzelne Ausbildungsbereiche Repräsentationsgrade – das ist der Anteil der Auszubildenden in den ausgewählten Berufen an allen Auszubildenden des Ausbildungsbereichs - zwischen 65 und 85 Prozent (insgesamt: 71 Prozent) erreicht (ebd.: 15). Im Anhang A1 befindet sich eine Auflistung der 52 Ausbildungsberufe. Die Grundgesamtheit der Erhebung ergibt sich dann aus allen deutschen Betrieben, die im Jahr 2000 im Rahmen der dualen betrieblichen Berufsausbildung mindestens eine Person in einem der einbezogenen 52 Ausbildungsberufe ausbilden⁷.

⁵ Die (sehr kleinen) Ausbildungsbereiche Seeschifffahrt und Hauswirtschaft werden nicht erfasst.

⁶ Zentrale Kriterien für die Auswahl sind die Besetzungstärke sowie die Abdeckung eines breiten inhaltlichen Ausbildungsspektrums.

⁷ Mit den in Abschnitt 4.5 vorgestellten Hochrechnungsfaktoren werden Analyseergebnisse auf die Gesamtanzahl aller Ausbildungsbetriebe in Deutschland gewichtet.

Aufgrund der veränderten Stichprobenziehung über die Betriebsdatei der Bundesagentur für Arbeit zum Stichtag 30. Juni 2006 (nachfolgend: Betriebsdatei) wird für die Erhebung für das Jahr 2007 etwas anders vorgegangen. Aus der Betriebsdatei⁸ wurde zunächst ermittelt, ob die in die Auswahl gelangten Betriebe überhaupt zum Kontaktzeitpunkt⁹ in einem der ausgewählten Ausbildungsberufe ausbilden. Da Betriebe in mehreren Berufen ausbilden können und die in BA-Daten verfügbaren Dreisteller für die Auswahl der Berufe nicht immer ausreichend tief genug gegliedert waren, schlossen sich mehr oder minder anspruchsvolle Screeningphasen (telefonisch und schriftlich) an die Lieferung der von der BA gelieferten etwa 32 Tsd. Betriebsadressen an. Es verblieben etwa dreizehn Tsd. Betriebsadressen für die persönlich geführten Interviews.

3.2. Erhebungsdesign

Kosten und Nutzen der betrieblichen Berufsausbildung für das Jahr 2000

Die Befragung zu den Kosten und Nutzen der betrieblichen Berufsausbildung für das Jahr 2000 setzt sich aus zwei Teilen zusammen. Die Kosten wurden mittels Face-to-face-Interviews von geschulten Interviewern ermittelt. Der Fragebogen zum Nutzen der Ausbildung war als schriftlicher Selbstausfüller angelegt, der im Betrieb hinterlassen wird. Es war den Interviewern jedoch freigestellt, ihn in Anschluss an die Kostenbefragung ebenfalls als Face-to-Face-Interview durchzuführen. In den meisten Fällen (88 Prozent) wurde der Nutzenfragebogen im Beisein des Interviewers ausgefüllt (infas 2001: 16).

Da Kosten und Erträge der Berufsausbildung von Unternehmen in der Regel nicht gesondert erfasst werden, besteht der Hauptteil der Befragung in der detaillierten Erfassung der einzelnen Kosten- und Ertragskomponenten (Beicht et al., 2004: 21). Dies erfolgte in den Fragebögen jeweils für alle Auszubildenden eines befragten Betriebs, getrennt nach einzelnen Ausbildungsjahren. Zum Interviewzeitpunkt beziehen sich die Angaben jeweils auf das Vorjahr (das Jahr 2000).

Nach einem Pretest mit 22 Fällen im Herbst 2000 wurde die Befragung von Mitte Februar bis Mitte August 2001 durchgeführt. Die Datenerhebung erfolgte durch infas, Institut für angewandte Sozialforschung Bonn (infas 2001: 16). Der Feldeinsatz fand insgesamt in drei Wellen statt (vgl. 3.4).

Als Interviewer wurden Personen mit besonderen Erfahrungen in Betriebsbefragungen eingesetzt. Die Schulung der Interviewer erfolgte anhand eines Handbuchs, das in Zweifelsfällen zum Nachschlagen im Feld verwendet wurde, sowie durch die Feldeinsatzleitung von infas. Für die Durchführung eines Interviews waren durchschnittlich 3,1 Kontakte mit den Betrieben erforderlich. In Kleinbetrieben wurden in der Regel Leiter und Inhaber befragt, in mittleren und großen Betrieben die für die Ausbildung hauptverantwortlichen Personen (Beicht et al., 2004: 18).

⁸ Den Aufbau der BA-Betriebsdatei(en) beschreibt ein IAB-FDZ-Datenreport: Spengler, A. (2007): The establishment history panel 1975 – 2007 * handbook version 2.0.0. FDZ-Datenreport Nr. 4/2007. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

⁹ In der Betriebsdatei ist der Stand zum 30. Juni 2006 dokumentiert, weshalb sich gegenüber dem Befragungsjahr 2008 Abweichungen ergeben können.

Rund zwölf Prozent der Befragten konnten nicht alle Fragen an Ort und Stelle beantworten (infas 2001: 15). In diesen Fällen wurden die Informationen entweder von den Befragten während des Interviews von anderen Stellen eingeholt oder nach Beendigung des Interviews nachgereicht. Insgesamt hatten etwa 16 Prozent der Befragungseinheiten Schwierigkeiten mit der unmittelbaren Beantwortung einzelner Fragen. Hierbei handelte es sich in erster Linie um detaillierte Kostenangaben für einzelne Ausbildungsberufe und Ausbildungsjahrgänge. Sofern Befragte nicht in der Lage waren, exakte Wertangaben zu machen, wurden auch Schätzungen akzeptiert.

Kosten und Nutzen der betrieblichen Berufsausbildung für das Jahr 2007

Im Unterschied zur Erhebung für das Jahr 2000 gibt es nur einen Fragebogen für die Kosten- und Nutzenaspekte der betrieblichen Berufsausbildung. Außerdem wurden diesmal alle Fragen persönlich-mündlich gestellt. Die durchschnittliche Interviewdauer betrug etwa 72 Minuten. Wie bereits bei der 2000er-Erhebung wurden Kosten und Nutzen der Berufsausbildung detailliert für alle Ausbildungen im jeweiligen Zielberuf pro Ausbildungsjahr erhoben.

Der Pretest umfasste 31 Interviews. Ansonsten entspricht die Vorgehensweise der für die Vorerhebung, d.h. die Befragung wurde nach einer Schulung von hauptberuflichen Interviewern durchgeführt, die in kleinen und mittleren Firmen die Geschäftsleitung und in Großbetrieben die für die Ausbildung hauptverantwortlichen Personen der ausgewählten Betriebe befragten. Wenn auch die überwiegende Mehrheit der Interviews mit einer Person pro Betrieb durchgeführt wurde, so wurden in größeren Betrieben gegebenenfalls mehrere Personen (bis zu fünf) befragt.

3.3. Ausschöpfung und Einhaltung des Stichprobenplans

Kosten und Nutzen der betrieblichen Berufsausbildung für das Jahr 2000

Von den rund 13.000 durch die Kammern und zuständigen Stellen gezogenen Betriebsadressen wurden von infas für die Durchführung der Erhebung insgesamt 7.125 Adressen eingesetzt (Bruttostichprobe). Nach Abzug der stichprobenneutralen Ausfälle (etwa aufgrund von Betriebsauflösungen oder dem (mehr oder minder spontanem) Einstellen aller Ausbildungsaktivitäten) enthält die bereinigte Bruttostichprobe 6.257 Ausbildungsbetriebe. Das führte zu 2.518 auswertbaren Interviews und entspricht einer Ausschöpfungsquote von 40,2 Prozent. In 22 Fällen wurde lediglich der Fragebogenteil zu den Ausbildungskosten beantwortet, sodass für die Nutzendimension 2.496 Betriebe analysiert werden können. Details zur Ausschöpfung der Stichprobe sowie Ausfallgründen befinden sich in Tabelle 2a.

Tabelle 2a: Ausschöpfung und Ausfallgründe bei der Betriebsbefragung „Nutzen und Nettokosten der Berufsausbildung“ für das Jahr 2000

	absolut	Prozent
Bruttostichprobe	7.125	100
davon stichprobenneutrale Ausfälle:		
Adressprobleme (Betrieb existiert nicht unter angegebener Adresse, Privathaushalt unter angegebener Adresse, Betrieb wird aufgelöst)	391	5,5
kein Ausbildungsbetrieb	477	6,7
bereinigte Bruttostichprobe	6.257	100
davon Ausfälle:		
nicht erreicht (Ansprechpartner nicht erreicht, Termin nicht realisiert)	415	6,6
Verweigerung durch Betrieb (z.B. Betrieb nimmt grundsätzlich nicht teil, zu viele Umfragen, wegen Datenschutz)	658	10,5
Verweigerung durch Ansprechpartner (z.B. keine Zeit, Länge des Interviews, kein Interesse, grundsätzliche Ablehnung)	2.406	38,5
Betriebsstätte nicht auskunftsfähig	144	2,3
Interviews realisiert insgesamt	2.634	42,1
davon nicht auswertbar:		
abgebrochene Interviews	50	0,8
keinen Auszubildenden zum Stichtag 30.9.2000	12	0,2
Bildungsträger	12	0,2
andere Gründe	42	0,7
auswertbare Interviews	2.518	40,2

Quelle: Beicht et al., 2004: 17

Die Einhaltung der Soll-Vorgaben für die zu realisierenden Interviews konnten mittels der dreistufigen Stichprobenziehung weitgehend erreicht werden (infas 2001: 21 ff.). In knapp 50 Prozent der Zellen wurden die Vorgaben der zu realisierenden Interviews exakt erfüllt. In 29 Prozent der Zellen wurden die Vorgaben übererfüllt, in 19 Prozent der Zellen lieferten die Ist-Werte im Verhältnis zum Soll Werte zwischen 25 bis 75 Prozent. In fünf Prozent der Zellen betrug der Grad der Erreichung der Soll-Vorgaben weniger als 25 Prozent. Wesentliche Über- und Untererfüllungen traten insbesondere in Zellen auf, in welchen lediglich ein bis drei Interviews zu realisieren waren.

Bezüglich der Ausbildungsberufe wurde die Mehrzahl der Sollvorgaben zu 100 Prozent (plus/minus zehn Prozentpunkte) eingehalten. Für sechs Ausbildungsberufe beträgt die Realisierungsquote weniger als 90 Prozent. Hierbei handelt es sich insbesondere um die vorwiegend in Großbetrieben ausgebildeten Berufe, für welche der Stichprobenplan überproportional hohe Sollvorgaben festlegte. Das Mindestsoll von 40 Interviews pro Ausbildungsberuf wurde in fünf Fällen um maximal drei Interviews unterschritten.

Kosten und Nutzen der betrieblichen Berufsausbildung aus dem Jahr 2007

Nach den Screeningprozessen verblieben in der Ausgangsstichprobe etwa 8.900 Betriebsadressen. Davon waren etwa fünf Prozent (426 Betriebe) stichprobenneutrale Ausfälle (Adressausfälle bzw. keiner oder ausschließlicher Ausbildungsbetrieb). Mit den verbleibenden etwa 8.500 Adressen wurden insgesamt 3.211 Interviews begonnen. Im Einzelnen ergibt sich für die Ausschöpfung und Einhaltung des Stichprobenplans das folgende Bild.

Tabelle 2b: Ausschöpfung und Ausfallgründe bei der Betriebsbefragung „Kosten und Nutzen der Berufsausbildung“ für das Jahr 2007

	absolut	Prozent
Bruttostichprobe	8.907	100
• Betrieb existiert nicht unter angegebener Adresse, Betrieb wird aufgelöst	155	1,7
• kein oder reiner Ausbildungsbetrieb, Betrieb bildet nicht im ausgewählten Beruf aus	271	3,0
Bereinigte Bruttostichprobe	8.481	100
• Betrieb/ Ansprechpartner nicht erreicht	502	5,9
• Betriebsstätte nicht auskunftsfähig	141	1,7
• Verweigerung durch Betrieb	755	8,9
• Verweigerung durch Ansprechpartner	3.809	44,9
• Sonstige Gründe	63	0,7
Realisierte Interviews	3.211	37,9
• Abgebrochen	117	1,4
• Nicht auswertbar	108	1,3
auswertbare Interviews	2.986	35,2

Quelle: Schönfeld et al. (2010): 30

Nicht zuletzt bedingt durch die unsicheren Adressen, die den Interviewern zur Abklärung vor Ort mitgegeben wurden, gibt es 426 Adressen von Betrieben, die entweder nicht in einem der ausgewählten Berufe ausbilden oder unter der Adresse nicht (mehr) existent waren. Das führt zu einer bereinigten Bruttostichprobe von 8.481 potenziellen Befragungseinheiten. Vom bereinigten Stichprobenansatz konnten 502 Einheiten wegen Abwesenheit des Ansprechpartners und 141 wegen unzureichender Auskunftsfähigkeit des Betriebs nicht befragt werden. Hinter der letztgenannten Gruppierung stehen insbesondere Betriebsstätten größerer Unternehmen, insbesondere von Filialen, die für eine einzelne Betriebsstätte nicht auskunftsfähig, teilweise auch nicht auskunftsberechtigt sind.

In 755 Fällen verweigerte der Betrieb ein Interview, 3.809 Mal der für das Thema kompetente Ansprechpartner. In 63 Fällen fiel das Interview aus anderen Gründen aus (z.B. Termine nicht eingehalten oder kurzfristig abgesagt). In 3.211 Betrieben wurde ein Interview geführt,

welches bei 117 Betrieben nicht bis zu Ende geführt wurde. Sofern in einem solchen Fall die Daten größere Lücken aufweisen, wurden die entsprechenden Interviews nicht in den Datensatz aufgenommen. 108 Interviews mussten als nicht auswertbar aus dem Datensatz entfernt werden. Der endgültige und im BIBB-FDZ vorliegende Analysedatensatz enthält 2.986 Datenzeilen.

4. Variablenspektrum und Datenaufbereitungen

Zuerst wird ein Überblick über das Variablenspektrum der Kosten-Nutzen-Erhebung für das Jahr 2000 gegeben. Abschnitt 4.2 beschreibt die Korrekturen und Plausibilitätsprüfungen, die im Anschluss an die Feldphase von infas und im BIBB vorgenommen wurden. Die (zusätzlichen) Aufbereitungsschritte der Daten im BIBB-FDZ werden in Abschnitt 4.3 zusammengestellt. Abschnitt 4.4 enthält Informationen zu einer möglichen Imputation fehlender Werte. Auf die Gewichtung und Hochrechnung der Daten geht Abschnitt 4.5 ein. Schließlich werden in 4.6 zusammenfassende Kosten- und Nutzenvariablen erläutert.

4.1 Variablenspektrum

Für einen kompletten und detaillierten Überblick über die einzelnen Variablen wird auf die FDZ-Arbeitshilfen verwiesen (Fragebögen, Spieldaten). Die Kosten-Nutzen-Erhebung für das Jahr 2000 enthält für die folgenden Themenblöcke Variablen.

Tabelle 3a: Themen- und Variablenübersicht der Kosten-Nutzen-Erhebung für das Jahr 2000

Themenblock	Anzahl Variablen	Variablenbereich	Fragebogen-seiten (von – bis)
I. Kosten der Berufsausbildung			
Allgemeine Angaben zum Betrieb	18	f1k – f7_4k	2 – 3
Allgemeine Angaben zur Ausbildung im Betrieb insgesamt sowie im ausgewählten Beruf	72	f8_1k – f20_4k	4 – 8
Angaben zu den betrieblichen Ausbildungskosten im ausgewählten Beruf	207	f21_1k – f51k	8 – 19
Angaben zur betrieblichen Lehrwerkstatt bzw. sonstigen betrieblichen Lehreinrichtungen	35	f52k – f63k	19 – 21
Angaben zum innerbetrieblichen Unterricht	19	f64k – f71k	21 – 22
Angaben zur Kostenerstattung bzw. finanziellen Förderung	9	f72k – f77_2k	23

(Fortsetzung nächste Seite)

Themenblock	Anzahl Variablen	Variablenbereich	Fragebogen-seiten (von – bis)
II. Nutzen der Berufsausbildung			
Allgemeine Angaben zur Situation des Betriebs und zur Ausbildung insgesamt	125	f1 – f15_6	2 - 6
Angaben zum Nutzen der Ausbildung	61	f15a – f35_15	7 - 10
III. Zusatzvariablen	24	lfd – ausbstd_azubi	trifft nicht zu; eine Übersicht befindet sich in Tabelle 5
IV. synthetische Variablen	6	x1 – x6	trifft nicht zu; eine Beschreibung befindet sich in Tabelle 5

Anmerkung: Kosten und Nutzen werden je in einem separaten Fragebogen erhoben. Deswegen wiederholen sich die Seitenangaben.

Die Aufstellung zeigt die Schwerpunkte der Betriebsbefragung auf. Die meisten Variablen (insgesamt 207) beziehen sich auf Kosten der Berufsausbildung, aber auch deren Nutzen wird detailliert erfasst (61 Variablen).

Für die Erhebung aus dem Jahr 2007 wurden Veränderungen des Erhebungsdesigns vorgenommen. Die Beschreibung des Variablenkanons sieht daher etwas anders aus. Da für die Kosten und den Nutzen betrieblicher Berufsausbildung keine separaten Fragebögen und verschiedene Erhebungsmethoden eingesetzt wurden, ist der Datensatz etwas anders sortiert. Bei Variablenbenennungen entfällt die Kennzeichnung der Kostenvariablen durch das Suffix „k“. Die Variablenbenennung orientiert sich an der Kurzversion des CAPI-Fragebogens, der auf den BIBB-FDZ-Internetseiten eingestellt ist.

Tabelle 3b: Themen- und Variablenübersicht der Kosten-Nutzen-Erhebung für das Jahr 2007

Themenblock	Anzahl Variablen	Variablenbereich	CAPI-Kurzversion des Fragebogens (von – bis)
I. original erhobene Variablen			
allgemeine Angaben zum Betrieb und zur Personalstruktur	30	v2 bis v12	1 - 5
allgemeine Angaben zur Ausbildung im Betrieb insgesamt sowie im ausgewählten Beruf	11	v13 bis v19	5 - 7

(Fortsetzung nächste Seite)

Themenblock	Anzahl Variablen	Variablenbereich	CAPI-Kurzversion des Fragebogens (von – bis)
Angaben zu den betrieblichen Ausbildungskosten im ausgewählten Beruf	109	v20 bis v47_4	5 – 16
Angaben zu produktiven Zeiten der Auszubildenden (inklusive Äquivalenzerträge)	98	v48_1 bis v72	17 - 26
Angaben zur Lehrwerkstatt und zum innerbetrieblichen Unterricht	40	v73 bis v92	26 – 28
Angaben zur Kostenerstattung bzw. finanziellen Förderung	3	v93 bis v95	31
Angaben zum Nutzen betrieblicher Berufsausbildung	90	v96 bis v133_12	32 - 40
Rekrutierungskosten (vom externen Arbeitsmarkt)	113	v134 bis v161	41 - 46
Subjektive Einschätzungsfragen/ Maßnahmen des Betriebes bezüglich der Berufsausbildung		v162 – v170	47 - 50
II. Zusatzvariablen	9	beschg bis wzw	trifft nicht zu; eine Aufstellung befindet sich in Tabelle 5
III. synthetische Variablen	33	loh001j1 bis kos076j4_org	trifft nicht zu; eine Aufstellung befindet sich in Übersicht 3

4.2 Korrekturen und Plausibilitätsprüfungen

Die Darstellung der Korrekturen und Plausibilitätsprüfungen befindet sich aus Übersichtsgründen nicht in diesem Daten- und Methodenreport. Sie sind dem Untersuchungsgegenstand und dem Erhebungsdesign angemessen vielschichtig und detailliert. Für die Kosten-Nutzen-Erhebung für das Jahr 2000 können sich Interessierte die Korrekturen und Plausibilitätsprüfungen im BIBB-FDZ Daten- und Methodenreport Nr. 4/ 2009 ansehen. Besondere Auffälligkeiten gibt es dabei nicht. Noch ausführlicher sind die Korrekturen und Plausibilitätsprüfungen in den Feldberichten dargestellt, die auf Anfrage bzw. mit dem Abschluss eines Nutzungsvertrages natürlich auch für die Erhebung aus dem Jahr 2007 eingesehen werden können.

Bei der Kosten-Nutzen-Erhebung für das Jahr 2007 wurden in fünf „Kategorien“ Datenprüfungen vorgenommen:

- Bereinigung von Jahresangaben (etwa das Gründungsjahr) und monetären Größen (etwa die durchschnittlichen Arbeitslöhne von Fachkräften)

- Bereinigung weiterer, nicht standardisierter Angaben (etwa Arbeitszeiten oder aber auch Anschaffungskosten für Lehrmaterial und ähnliches)
- Bereinigung von Ausreißern (etwa Höhe der Sozialleistungen, Essensgeld)
- Über- und Unterschreiten von Prüfwerten (etwa bei der Arbeitszeit, wenn nicht bezahlte Überstunden zur wöchentlichen Arbeitszeit hinzugerechnet wurden und so Wochenarbeitszeiten deutlich über 42 Stunden pro Woche liegen).
- Variablen mit einem hohen Anteil an Fehlwerten (etwa Abschreibungskosten für Maschinen, kalkulatorische Miete der Lehrwerkstatt und ähnliches).

Die Feld- und Methodenberichte beschreiben all dies im Detail inklusive der entsprechenden Editierregeln. Teilweise wurden (für die 2007er-Erhebung) auch die Informationen aus den so genannten Drop-Off-Fragebogen verwendet¹⁰ und zusätzliche Datenbereinigungen vom Forschungsprojekt vorgenommen. Nähere Auskünfte hierzu erteilt auf Anfrage gerne das BIBB-FDZ.

4.3 Datenaufbereitungen im BIBB-FDZ (nur Daten für das Jahr 2000)

Die Variablen der Erhebung „Kosten und Nutzen der betrieblichen Berufsausbildung im Jahr 2000“ waren im BIBB auf einzelne Teildatensätze verteilt. Eine Zusammenführung der einzelnen Teildatensätze erfolgt im Zuge der Datenaufbereitung durch das BIBB-FDZ. In einigen Teildatensätzen waren differenzierte Missing-Value-Deklarationen nicht enthalten. Sie wurden – soweit möglich – im BIBB-FDZ nachträglich rekonstruiert. Im Zuge der Datenzusammenführung und Datenaufbereitung hat das BIBB-FDZ in der Erhebung für das Jahr 2000 die fehlenden Werte neu deklariert. Fehlende und ungültige Werte wurden bei der Erhebung für das Jahr 2000 auf die Standarddeklarierung fehlender Werte des BIBB-FDZ umkodiert.

Übersicht 1 Standarddeklaration des BIBB-FDZ für verschiedene Arten fehlender Antwortwerte (Erhebung für das Jahr 2000)

-9	keine Angabe
-8	weiß nicht
-7	trifft nicht zu
-6	ungültig (z.B. Ausschluss durch Datenbereinigung)
-5	Verweigert
-1	Filter

Anmerkung: datensatzspezifisch sind weitere Werte möglich.

In der Kosten-Nutzen-Erhebung für das Jahr 2000 ist der Wert -2 ein datenspezifischer und wichtiger *zusätzlicher* fehlender Wert, der sich auf einzelne Ausbildungsjahre bezieht. Da die Kosten und Nutzenkomponenten der betrieblichen Berufsausbildung für jedes Ausbildungsjahr getrennt erfasst werden, gibt es in den entsprechenden einzelnen Kosten-

¹⁰ Sofern Betriebe keine Angaben zu für das Kostenmodell wichtigen Größen machten, wurde im Zuge einer (schriftlichen) Nachbefragung (gleich Drop-Off-Fragebogen) versucht, diese Angaben doch noch zu bekommen. Insgesamt gab es 652 Rückläufe für diese Fragebögen (etwa 22 Prozent der Nettofallzahl im 2007er-Datensatz).

und Nutzenvariablen in nicht besetzten Ausbildungsjahren aus logischen Gründen fehlende Werte, die im FDZ-Datensatz mit dem Wert -2 gekennzeichnet sind. Das ist eine nicht unbedeutende Besonderheit des Datensatzes, die sich aus der Erkenntnislogik der Erhebung ergibt. Insbesondere in kleineren Betrieben bzw. grundsätzlich in allen Betrieben, die weniger als drei Auszubildende beschäftigen, tauchen diese fehlenden Werte mindestens auf. Für das erste Lehrjahr beispielsweise können 1.041 Betriebe keine Angaben zu Kosten und Nutzen machen (zweites Lehrjahr: 883; drittes Lehrjahr: 977; viertes Lehrjahr: 2.260).

Um zu einem vollständigen Datensatz mit Besetzungen in allen Lehrjahren für alle Betriebe zu kommen, besteht die Möglichkeit, (i) die logisch fehlenden Werte -2 durch Imputationsroutinen zu ersetzen (vgl. dazu 4.4) oder (ii) synthetische Variablen zu verwenden (Abschnitt 4.5), für die nicht nur bereits Datenimputationen nach (i) durchgeführt wurden, sondern einzelne Kosten (Teil- und Vollkosten) bzw. Nutzen (Teil- und Vollkosten) berechnet wurden¹¹.

Bei Variablen, die keine differenzierte Erfassung der fehlenden Werte mehr enthielten, hat das BIBB-FDZ Systemmissings auf -9 (keine Angabe) umkodiert. Sofern Filter-Missings nicht mehr enthalten waren, wurden diese neu angelegt. Detaillierte Informationen zu weiteren Umkodierungen des Kosten-Fragebogens aus dem Jahr 2000 befinden sich im FDZ-Daten- und Methodenreport Nr. 4/ 2009.

Für die Daten des Jahres 2007 gibt es keine Umkodierung fehlender Werte auf die FDZ-Standards, da eine Zusammenführung von Teildatensätzen nicht notwendig ist. Für das Jahr 2007 gelten demnach die fehlenden Werte wie im Fragebogen bzw. den Wertelabels einzelner Variablen angegeben.

In Datensätzen des BIBB-FDZ werden grundsätzlich keine MISSING VALUES im Vorhinein definiert. Das entscheiden Anwender selbstständig.

4.4 Imputation fehlender Werte

Auf die Imputation fehlender Werte wird aufgrund der Eigenarten des Untersuchungsgegenstands eingegangen, denn eine nach Ausbildungsjahren getrennte Erfassung der Kosten- und Nutzen-Aspekte der Berufsausbildung führt insbesondere in kleineren Betrieben zu einer relativ hohen Anzahl an fehlenden Werten, weil die entsprechenden Betriebe nur Auszubildende in einem oder zwei Lehrjahren beschäftigen und demzufolge keine Angaben für nicht besetzte Lehrjahre machen (können). In der Folge erhält man für alle Variablen „logische“ fehlende Werte, die sich auf Kosten- und Nutzenaspekte in bei der konkreten Befragungseinheit nicht besetzten Lehrjahren beziehen.

Insbesondere bei der Auswertung von Teilpopulationen kann das zu kleinen Fallzahlen führen bzw. Selektionsverzerrungen verursachen. Es gibt verschiedene Möglichkeiten der

¹¹ Entsprechende Ausführungen befinden sich in Beicht et al. (2004): 21 - 28

Datenimputation, die sich teilweise beträchtlich im methodischen Vorgehen und dem Ergebnis unterscheiden¹².

Imputationen mit Durchschnittswerten

Für die repräsentative Berechnung der Ausbildungskosten und die Hochrechnung auf alle Ausbildungsbetriebe haben Beicht et al. (2004) fehlende Werte der für die Kostenberechnung relevanten Variablen durch Durchschnittswerte der vorhandenen Werte ersetzt (ebd.: 20). Dabei wurde so weit wie möglich nach den relevanten Einflussfaktoren für die jeweilige Variable (Trennung nach Ausbildungsbereich, Betriebsgröße oder Berufsgruppe) vorgegangen. Die Quote fehlender Werte lag in den meisten Fällen unter drei Prozent. Solche Ersetzungen sind im Datensatz des BIBB-FDZ nicht enthalten und müssten von Anwendern gegebenenfalls eigenständig vorgenommen werden.

Imputationen im BIBB-FDZ

Im Zuge der Datenaufbereitungen stellt das BIBB-FDZ für ausgewählte und wenige Variablen Variableninhalte mit imputierten Werten zur Verfügung. Die Auswahl basiert auf bereits durchgeführten Forschungsarbeiten mit den Kosten-Nutzen-Daten.

Übersicht 2: Originalvariablennamen und korrespondierende Variablennamen bei imputierten Werten bzw. synthetischen Variablen im Datensatz für das Jahr 2000

originaler Variablenname	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
f21_1k	Ausbildungsvergütung im ersten Lehrjahr	1469 (941)	f21_1k_i	2496 (910)
f21_2k	Ausbildungsvergütung im zweiten Lehrjahr	1624 (1051)	f21_2k_i	2496 (1031)
f21_3k	Ausbildungsvergütung im dritten Lehrjahr	1530 (1161)	f21_3k_i	2496 (1153)
f21_4k	Ausbildungsvergütung im vierten Lehrjahr	221 (1251)	f21_4k_i	656 (1186)
-	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (erstes Lehrjahr)	-	t67_1i	2496 (15118)
-	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (zweites Lehrjahr)	-	t67_2i	2496 (16963)
-	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (drittes Lehrjahr)	-	t67_3i	2496 (18779)
-	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (viertes Lehrjahr)	-	t67_4i	656 (10149)

(Fortsetzung nächste Seite)

¹² Einen kurzen Überblick über gängige Imputationsverfahren gibt das BIBB-FDZ in der Dokumentation der Imputation der Lohnvariable der BIBB-BAuA-Erhebung des Jahres 2006.

originaler Variablenname	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
-	Durchschnitts-Bruttokosten (Vollkosten) in Euro – Modellrechnung	-	x1	2496 (19178)
-	Durchschnitts-Äquivalenzerträge (Vollkosten) in Euro – Modellrechnung	-	x2	2496 (7891)
-	Durchschnitts-Nettokosten (Vollkosten) in Euro – Modellrechnung (= x1 – x2) ¹³	-	x3	2496 (11287)
-	Durchschnitts-Bruttokosten (Teilkosten) in Euro – Modellrechnung	-	x4	2496 (11202)
-	Durchschnitts-Nettokosten (Teilkosten) in Euro – Modellrechnung (= x4 – x2) ¹¹	-	x5	2496 (3311)
-	Durchschnitts-Substitutionserträge ¹⁴ (Vollkosten) in Euro – Modellrechnung	-	x6	2496 (15790)

Anmerkungen: die ausgewiesenen Mittelwerte basieren auf ungewichteten Angaben, die Ausbildungsvergütungen und Personalkosten sind in DM gemessen; die summarischen Erträge und Kosten des dritten Variablenblocks werden in Euro ausgewiesen; **Quelle:** BIBB-CBS 2000

Die ersten vier Variablen beziehen sich auf die Ausbildungsvergütungen und sind im FDZ-Datensatz sowohl im originalen wie imputierten Wertebereich enthalten. Bei den folgenden vier Variablen im zweiten Block handelt es sich um einen Auszug aus der Vollkostenrechnung des BIBB (Personalkosten pro Auszubildenden). Bei den letzten sechs Variablen des dritten Blocks handelt es sich wie bei den Variablen des zweiten Blocks um vollsynthetische Variablen, die aus bestimmten Werten der Originalvariablen im BIBB generiert wurden. Die jeweiligen Additionen bzw. Rechenoperationen, die zu diesen synthetischen Variablen führen, beschreiben Beicht et al. (2004) in Teil II.

Zu den synthetischen Variablen ist verschiedenes anzumerken:

- 1.) Das FDZ-Angebot zur Verwendung der synthetischen Variablen soll potenziellen Anwendern die (komplexe) Berechnung der zusammenfassenden Kosten- und Nutzenkomponenten der betrieblichen Berufsausbildung erleichtern bzw. diese überflüssig machen.
- 2.) Die synthetischen Variablen sind nach dem von der Sachverständigenkommission Kosten und Finanzierung der beruflichen Bildung vorgeschlagenen Vorgehen gebildet worden. Mit dem FDZ-Auswertungsdatsatz und unter Verwendung der synthetischen Variablen werden nicht exakt die Eckwerte der Publikation von Beicht et al. (2004) erreicht. Die inhaltlichen Gründe hierfür sind vielfältig (etwa: unterschied-

¹³ Die Differenzbildung muss dafür auf der einzelbetrieblichen Ebene erfolgen.

¹⁴ Die Definition von Substitutionserträgen befindet sich in Beicht et al. (2004) auf den Seiten 198 ff.

liche Imputationsverfahren) und in der entsprechenden Publikation nachzulesen (ebd., Teil II). Der entscheidende methodische Grund für die Abweichungen ist die Hochrechnung einmal nach Auszubildendengewichten (Beicht et al.) und Betriebsgewichten (FDZ-Datensatz).

- 3.) Die synthetischen Variablen sind so genannte sensible Zusatzvariablen, d.h. sie werden Anwendern nur auf begründeten Antrag zur Verfügung gestellt und sind demnach ohne explizite Anforderung bzw. Nennung im Nutzungsantrag *nicht* im Auswertungsdatensatz enthalten.
- 4.) Für die Konstruktionsprinzipien der synthetischen Variablen liegt ein FDZ-internes Dokument mit dem Generierungsalgorithmus vor, der die Ausführungen von Beicht et al. (2004) in mathematische Operationen übersetzt. Dieses Dokument ist nicht nur umfangreich und komplex, sondern enthält aufgrund des expliziten Bezugs auf die originalen Variablenwerte datenschutzrechtlich sensible Informationen. Es kann daher leider nicht öffentlich zugänglich gemacht werden. Während eines Gastaufenthalts ist ein Einblick in dieses Dokument auf Anfrage möglich.
- 5.) *Die Verwendung der synthetischen Variablen bzw. der mit imputierten Werten geschieht grundsätzlich in eigener Verantwortung der Anwender! Das FDZ oder das BiBB übernehmen keine Haftung oder Gewähr für die Verwendung dieser Variablen.*

Da es bei der Arbeit mit den Daten auch für andere Variablen sinnvoll sein kann, Werte zu imputieren, wird der Generierungsbefehl in STATA am Beispiel der Imputation der Brutto-Ausbildungsvergütungen im ersten Lehrjahr dargestellt. Die Imputation fehlender Werte bei anderen Variablen kann analog vorgenommen werden.

** 3-jährige Ausbildungsberufe*

```
impute f21_1k bgroesse wzw westost f21_2k f21_3k if lz_3 == 1, ///  
generate(f21_1k_i)
```

**4-jährige Ausbildungsberufe*

```
impute f21_4k bgroesse wzw westost f21_1k f21_2k f21_3k if lz_3 == 0, ///  
generate(f21_1k_i)
```

In den Beispielen wird davon ausgegangen, dass sich die fehlenden Werte über eine Funktion imputieren lassen, die konditionale Effekte von Betriebsgrößenklassen, Wirtschaftszweigen sowie von Ausbildungsvergütungen in anderen Lehrjahren berücksichtigt. Mit ice.ado lassen sich multiple Imputationen realisieren bzw. mehr als 30 Regressoren verwenden. Wegen sehr langer Laufzeiten bei einem im vorliegenden Fall nur mäßig anderem Imputations-Endergebnis wurde den schnelleren und einfacheren impute-Befehlen der Vorzug gegeben. Auch SPSS stellt Imputationsroutinen zur Verfügung, auf die in diesem Handbuch nicht eingegangen wird.

Bei den Daten für das Jahr 2007 wurde der Imputationsalgorithmus ausgebaut und (zusätzlich) auf mögliche konditionale Selektivitäten bei fehlenden Angaben kontrolliert, etwa dass die Betriebsgröße ein relativ guter Prädiktor dafür ist, dass für einzelne Betriebe Angaben zu Kosten und Nutzen der Berufsausbildung nicht für alle möglichen Lehrjahre vorliegen. Dann erhält man – bei metrischen Variablen - einen ähnlichen Fall, wie ihn etwa

Gartner (2005) für die Kappung der Einkommen an der Beitragsbemessungsgrenze in den IAB-Daten feststellt. Die fehlenden Angaben lassen sich dann in einem zweistufigen Modell imputieren, wobei im ersten Schritt eine logistische Regression auf die Ausfallwahrscheinlichkeit und im zweiten eine vom (beobachtbaren) Ausfallmechanismus abhängige Ersetzung der fehlenden Werte bei den jeweiligen Kosten- und Nutzenkomponenten erfolgt.

Die methodische Vorgehensweise entspricht demnach vollständig dem von Gartner (2005) vorgeschlagenen Vorgehen. Auf der Ebene einzelner Variablen wurden teilweise unterschiedliche Regressionsmodelle aufgestellt, da ja für einzelne Sachverhalte verschiedene Ausfallmechanismen zugrunde liegen können. Im Bedarfsfall können die Modelle zur Imputation fehlender Werte auf der Ebene von STATA-Programmcodes zur Verfügung gestellt werden. Weitere Hinweise zur Vorgehensweise und den empirischen Ergebnissen im Kosten-Nutzen-Datensatz 2007 gibt auch jederzeit gerne das BIBB-FDZ.

Analog zu Übersicht 2 gibt es auch für das Jahr 2007 einige synthetische Variablen, die mit einer begründenden Zusatzangabe für Auswertungen zur Verfügung stehen.

Übersicht 3: Originalvariablenamen und korrespondierende Variablenamen bei imputierten Werten bzw. synthetischen Variablen im Datensatz für das Jahr 2007

originaler Variablenname	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
loh001j1	Ausbildungsvergütung im ersten Lehrjahr	1504 (539)	loh001j1_imp	2986 (514)
loh001j2	Ausbildungsvergütung im zweiten Lehrjahr	1729 (594)	loh001j2_imp	2986 (574)
loh001j3	Ausbildungsvergütung im dritten Lehrjahr	1591 (660)	loh001j3_imp	2986 (652)
loh001j4	Ausbildungsvergütung im vierten Lehrjahr	235 (728)	loh001j4_imp	705 (663)
kos001j1	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (erstes Lehrjahr)	1597 (8690)	kos001j1_imp	2986 (8346)
kos001j2	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (zweites Lehrjahr)	1845 (9434)	kos001j2_imp	2986 (9204)
kos001j3	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (drittes Lehrjahr)	1711 (10543)	kos001j3_imp	2986 (10307)
kos001j4	Personalkosten für jeden Auszubildenden pro Ausbildungsjahr (viertes Lehrjahr)	250 (5984)	kos001j4_imp	705 (5759)

(Fortsetzung nächste Seite)

originaler Variablenname	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
kos073j1_org	Bruttokosten der Ausbildung im ersten bis vierten Lehrjahr	1597 (15584)	kos073j1	2986 (15374)
kos073j2_org		1845 (16290)	kos073j2	2986 (16376)
kos073j3_org		1711 (17381)	kos073j3	2986 (17573)
kos073j4_org		250 (5817)	kos073j4	705 (11594)
nuz004j1_org	Äquivalenzerträge der Ausbildung im ersten bis vierten Lehrjahr	1597 (9174)	nuz004j1	2986 (9200)
nuz004j2_org		1845 (11006)	nuz004j2	2986 (10999)
nuz004j3_org		1711 (13566)	nuz004j3	2986 (13879)
nuz004j4_org		250 (5817)	nuz004j4	2986 (4080)
kos076j1_org	Nettokosten der Ausbildung im ersten bis vierten Lehrjahr	1597 (6409)	kos076j1	2986 (6174)
kos076j2_org		1845 (5284)	kos076j2	2986 (5377)
kos076j3_org		3816 (10106)	kos076j3	2986 (3695)
kos076j4_org		250 (5347)	kos076j4	705 (4080)
	Ausbilderstunden (gesamt) <i>pro Auszubildendem</i> in einer Standardwoche	-	ausbstd_azubi	2986 (5,43)

Anmerkung: die ausgewiesenen Mittelwerte sind ungewichtete Angaben; **Quelle:** BIBB-CBS 2007

4.5 Gewichtung und Hochrechnung

Aufgrund der geschichteten Stichprobe nach Ausbildungsbereichen und Berufen (Abschnitt 3.2) erfolgt die Anpassung an die realen Verhältnisse der Grundgesamtheit in den beiden jeweiligen Jahresdatensätzen über eine Gewichtung (Beicht et al., 2004: 19; Schönfeld et al., 2010: 30f.). Die Gewichtung und Hochrechnung wird auf zweierlei Arten durchgeführt. Zum einen eine Gewichtung auf Basis der Ausbildungszahlen und zum anderen eine auf Basis der befragten Betriebe (bezogen auf die Grundgesamtheit aller Ausbildungsbetriebe in Deutschland). Die erste Form der Hochrechnung auf Basis der Auszubildendenzahlen dient in erster Linie einer Abschätzung der Kosten und Nutzen der Berufsausbildung für alle Ausbildungsgänge in Deutschland und bezieht sich primär auf die aus dem Datensatz hervorgehende Modellrechnung. Da es sich hierbei um einen sehr speziellen Anwendungsfall aus dem ursprünglichen Projektzusammenhang handelt, sind die Gewichtungsfaktoren nach Auszubildendenzahlen im FDZ-Datensatz nicht enthalten (vgl. dazu auch Kapitel 5). Aus Gründen der Transparenz wird auf die Konstruktionsprinzipien dieser Gewichtungsfaktoren eingegangen.

In der Gewichtung auf Basis der Ausbildungszahlen wurden Koeffizienten für die repräsentative Anpassung der Stichprobenstruktur an die Verteilung der Grundgesamtheit aller Auszubildenden ermittelt. Es werden jeweils zwei Gewichtungs- und Hochrechnungsfaktoren berechnet: Zum einen wurden die Verteilungen der Auszubildenden in der Stichprobe an die Verteilungen der Auszubildenden in der Grundgesamtheit der 51 bzw. 52 abgedeckten Berufe angepasst, zum anderen wurde eine Anpassung an die Grund-

gesamtheit *aller* (Ausbildungs-)Berufe vorgenommen. Dabei wurden folgende Merkmale einbezogen:

- Ausbildungsberuf differenziert nach Ausbildungsjahren
- Ausbildungsbereich
- neue und alte Bundesländer
- Betriebsgröße

Die Berechnung der Gewichtungskoeffizienten wurde auf Basis der Ausbildungszahlen mittels eines so genannten Iterative Proportional Fitting Algorithmus (IPF) durchgeführt (infas, 2001: 56), um eine gleichzeitigen Anpassung an mehrere, nicht als gemeinsame Verteilungen vorliegender Randverteilungen vorzunehmen.

Für Auswertungen auf Betriebsebene wurden Koeffizienten zur Anpassung der Betriebsstruktur der Stichprobe an die Struktur der Grundgesamtheit der Betriebe aus der Beschäftigtenstatistik gebildet (Beicht et al., 2004: 19; Schönfeld et al., 2010: 31). Einbezogen wurden die Merkmale

- Wirtschaftszweig,
- Betriebsgrößenklasse sowie
- alte und neue Bundesländer.

Auch die Gewichtungskoeffizienten für Betriebe werden mittels eines IPF an die drei aufgelisteten strukturellen Ebenen angepasst. Der FDZ-Datensatz enthält zum einen ein Designgewicht für die Betriebe - hiermit wird der disproportionale bzw. geschichtete Stichprobenansatz ausgeglichen. Die Variable heißt **b_gew** und hat in den alten und neuen Bundesländern je einen Mittelwert von etwa eins. Zur Hochrechnung auf die Grundgesamtheit wird die Variable **b_hoch** verwendet. Der Mittelwert liegt etwas unter 200 und die Gesamtheit der Gewichte ergibt – angewendet auf den FDZ-Datensatz – eine gewichtete Gesamtanzahl von etwa 484 Tsd. (Ausbildungs-)Betrieben. In Tabelle 4 befinden sich Kennwerte der Verteilung der beiden Faktoren.

Tabelle 4: Mittelwert und Verteilungsmaße der Betriebsgewichte in den FDZ-Datensätzen Kosten und Nutzen der Berufsausbildung 2000 und 2007

Kennwert	<i>alte Bundesländer</i>				<i>neue Bundesländer</i>			
	b_gew		b_hoch		b_gew		b_hoch	
	2001	2007	2001	2007	2001	2007	2001	2007
Mittelwert	1,00	1,17	196	189	0,98	0,58	186	94
Standardabweichung	0,95	1,11	185	180	0,81	0,48	153	79
Median	0,59	0,83	116	134	0,67	0,33	126	54
10-Perzentil	0,08	0,17	15	28	0,18	0,15	35	24
25-Perzentil	0,24	0,36	47	58	0,35	0,20	67	33
75-Perzentil	1,47	1,75	288	283	1,65	0,73	313	119
90-Perzentil	2,32	2,11	454	342	1,84	1,22	349	296

Quelle: BIBB-CBS 2000 und 2007

Die Veränderungen der Hochrechnungs-/ Gewichtungsfaktoren zwischen dem Jahr 2000 und dem Jahr 2007 spiegeln die Veränderungen im Erhebungsdesign und der Fallzahlen wieder. Grob gesagt sinken etwa die Hochrechnungsfaktoren in beiden Landesteilen, da im Jahr 2007 für mehr Betriebe Angaben vorliegen als für das Jahr 2000. Die Gewichtungsfaktoren haben in Westdeutschland eher zu- und in Ostdeutschland eher abgenommen. Demzufolge liegt der Informationsgehalt einer einzelnen Datenzeile (gleich ein befragter Betrieb) etwas näher an dem (idealen) Wert eins, d.h. es waren mehr typische Betriebe in der 2007er-Stichprobe, wobei offen bleiben muss, ob das ein Effekt der höheren Fallzahlen (die Sollzahlen können besser auf die einzelnen Gewichtungszellen verteilt werden) oder des veränderten Erhebungsdesigns ist (die Betriebsdatei der Bundesagentur für Arbeit bildet die Grundgesamtheit besser ab als eine (regional geklumpte) Zufallsauswahl aus einzelnen Kammern). Insgesamt sollten die Unterschiede der Gewichtungs- und Hochrechnungsfaktoren zwischen der 2000er- und 2007er-Erhebung nicht überbewertet werden, da vor dem Hintergrund verschiedener Fallzahlen und unterschiedlicher Stichprobendesigns die Verschiebungen in allen Fällen eher gering zu bewerten sind.

5 Zusammenführung der jeweiligen Jahresdatensätze

Erhebung für das Jahr 2000

Das volle Variablenspektrum der Erhebung Nutzen und Nettokosten der Berufsausbildung für das Jahr 2000 entsteht aus der Zusammenführung von Teildatensätzen je für die Kosten und den Nutzen der Berufsausbildung. Die Variablen des Kostenteils sind am Suffix „_k“ im Datensatz identifizierbar. Da eine Korrespondenz der Kosten- und Nutzenteile nur für die Datensätze mit den Betriebsgewichten gegeben ist, wurde der Teildatensatz mit den Kosten auf Basis der Auszubildendengewichte bei der Datenaufbereitung nicht berücksichtigt¹⁵.

Erhebung für das Jahr 2007

In der 2007er-Erhebung gibt es nur einen Gesamtdatensatz, weil die Kosten- und Nutzenaspekte der betrieblichen Berufsausbildung im Rahmen eines einzelnen CAPI-Interviews erhoben werden. Da aber ebenso wie in der 2000er-Erhebung die vom Erhebungsinstitut erhaltenen Rohdaten im BIBB erhebliche Erweiterungen erfahren, um das Kosten-/Nutzenmodell berechnen zu können, war auch hier eine Variablenauswahl jenseits der Originalvariablen für den FDZ-Datensatz erforderlich.

Neben den Variablen der Fragebögen und den bereits angesprochenen zusätzlichen Variablen aus den Abschnitten 4.3 und 4.4 enthalten die Datensätze die folgenden zusätzlichen Merkmale.

¹⁵ Der einzige faktische Unterschied zu dem in den BIBB-FDZ-Datensatz integrierten Kostenteil besteht im Fehlen der Auszubildendengewichte, wobei sich letztere auf alle Auszubildenden in Deutschland beziehen (und nicht nur die in den 52 in der Befragung berücksichtigten Ausbildungsberufen).

Tabelle 5: Zusätzliche Variablen in den FDZ-Datensätzen Kosten und Nutzen der Berufsausbildung für das Jahr 2000 bzw. 2007

Variablen-name	Bezeichnung	numerische Vercodung bzw. range	Ausprägungen	Bemerkung
beschg	Beschäftigte inklusive Azubildende	-	kontinuierliches Merkmal	Betriebsgröße
berufart	Berufsart	1	gewerblich-technisch	nur für das Jahr 2000 verfügbar
		2	kaufmännisch-verwaltend	
westost	West-Ost-Kennzeichen für Standort des Betriebs	1	Westdeutschland	
		2	Ostdeutschland	
bgk	Betriebsgrößenklasse	1	unter 50 Beschäftigte	2007 davon abweichend (bis 9, 10 bis 49, 50 bis 499, 500 und mehr)
		2	50 bis unter 250 Beschäftigte	
		3	250 Beschäftigte und mehr	
abereich	Ausbildungsbereich	1	Industrie und Handel	
		2	Handwerk	
		3	Landwirtschaft	
		4	Freie Berufe	
		5	öffentlicher Dienst	
b_art3	Berufsart	1	kaufmännisch	nur für das Jahr 2007 verfügbar
		2	gewerblich	
		3	technisch	
zuststell	zuständige Stelle	1 ; 74	aus Platzgründen nicht abgedruckt	2000: enthält die Kammerbezirke aufgeteilt nach Handwerks-, Industrie- und Handels- sowie weiterer Kammern (Ärzte-, Steuerberaterkammer etc.) 2007: nicht vorhanden, da Auswahl über die Betriebsdatei der BA
bland	Bundesland (Standort des Betriebs)	1 ; 16	aus Platzgründen nicht abgedruckt	2007: aufgrund des Stichprobendesigns enthält die Variable 13 der 16 möglichen Ausprägungen
wzw	(Rekodierung) Wirtschaftszweig	1 ; 17 (2000) bzw. 1 ; 16 (2007)	aus Platzgründen nicht abgedruckt	2001: Rekodierung von f2_1k0 in 17 Wirtschaftszweige 2007: Rekodierung in 16 Wirtschaftszweige
nace_2digit	Wirtschaftszweige auf der Zweisteller-Ebene NACE	1 ; 29	aus Platzgründen nicht abgedruckt	nur für das Jahr 2007 verfügbar

Die aus Platzgründen nicht abgedruckten Ausprägungen werden anhand der frei zugänglichen Spieldaten verfügbar.

6 Hinweise zur Datennutzung

Auf die einzelnen Bereinigungsverfahren bei einzelnen Variablen geht dieser Daten- und Methodenbericht nicht ein. Im sechsten Kapitel der Dokumentation für die Daten der Kosten- und Nutzenerhebung für das Jahr 2000 (BIBB-FDZ-Daten- und Methodenbericht Nr. 4/2009) wird das für die 2000er-Erhebung relativ ausführlich dargestellt. Die Datenbereinigungsverfahren für die 2007er-Daten sind ähnlich. Mit einem abgeschlossenen Nutzungsvertrag können die entsprechenden Datenbereinigungsverfahren im Methodenbericht des Erhebungsinstituts eingesehen werden. Die Änderungen/ Bereinigungsverfahren bewegen sich im üblichen Rahmen und werden daher an dieser Stelle nicht detailliert wiedergegeben. Interessierte an diesen Bereinigungsverfahren können sich auch jederzeit gerne an das BIBB-FDZ wenden.

Bei der Datennutzung gibt es bezogen auf die Handhabung keine Besonderheiten zu beachten. Einzelne Variablen sind nach den Nummern im Fragebogen durchnummeriert. Alle weiteren in den Kapiteln 4 und 5 vorgestellten Variablen sind vollständig belabelt und in diesem Daten- und Methodenreport inhaltlich beschrieben. Das BIBB-FDZ bietet die Daten der Kosten- und Nutzenerhebung für die Jahre 2000 und 2007 jeweils in (jahresweise unterschiedlich strukturierten) Teildatensätzen an, die sich im Grad der Schutzwürdigkeit der Daten unterscheiden. Wie in der Regel für Betriebsdaten üblich, können die Daten der Kosten- und Nutzenerhebungen der betrieblichen Berufsausbildung nicht als Scientific-Use-File angeboten werden. Datenzugangswege sind demnach ausschließlich die Datenfernverarbeitung und Gastwissenschaftleraufenthalte. Angeboten werden mit diesen Datenzugangswegen vom BIBB-FDZ die folgenden Teildatensätze:

- 1.) *Grunddatensatz*: Der jeweilige Grunddatensatz für das Jahr 2000 bzw. 2007 enthält alle originalen Variablen der Kosten- und Nutzenerhebungen, einen Identifikator sowie die Variablen wie in Tabelle 5 beschrieben. Es gibt *keinen* Gesamtdatensatz, der die Jahre 2000 und 2007 umfasst. Aufgrund unterschiedlicher Erhebungsmethoden und gewissen Umstellungen der Berechnungsweisen für einige Konstrukte der Kosten und Nutzen betrieblicher Berufsausbildung sind Vergleiche zwischen den Jahren 2000 und 2007 nur unter bestimmten Voraussetzungen und eingeschränkt möglich (vgl. dazu Schönfeld et al., 2010: Kapitel 5)
- 2.) *Datensatz mit sensiblen Merkmalen*: hierzu gehören alle synthetischen Variablen wie in den Übersichten 2 und 3 beschrieben. Die Abtrennung vom jeweiligen Hauptdatensatz liegt weniger an schutzwürdigen Belangen der Erhebungseinheiten, sondern vielmehr zur Verdeutlichung, dass es sich hierbei um Serviceprodukte der Forschungsprojekte handelt. Sie werden bereitgestellt, um Forschenden die teilweise mühsame und akribische Arbeit bei der Bildung von aggregierten (summarischen) Kosten- und Nutzengrößen zu erleichtern bzw. diese überflüssig zu machen.
- 3.) *Datensatz mit kleinräumigen regionalen Kennziffern*: Für die Kosten-Nutzen-Erhebung aus dem Jahr 2007 gibt es die Kennziffer des Arbeitsamtsbezirks, Kreiskennziffern (auch als Kreiskennung in der NUTS3-Terminologie), Raumordnungsregionen und Regierungsbezirke. Diese Ebenen ermöglichen in der

Regel keine repräsentativen Auswertungen auf der Ebene einzelner konstituierender Einheiten. Denkbare Forschungsinteressen mit einer Verwendung dieser Regionalidentifikatoren können beispielsweise Re-Klassifizierungen sein (etwa Arbeitsagenturbezirke in SGB II-Typen) oder aber auch das Zusammenspielen der Mikrodaten mit Makroindikatoren für deskriptive oder multivariate Analysen.

- 4.) *Datensatz mit Volltexten*: Volltextangaben gibt es in der Kosten-Nutzen-Erhebung für das Jahr 2007 nur sehr wenige: zum einen der Betriebstyp (v91o), wenn bei dieser Frage mit sonstiges, und zwar ... geantwortet wurde (83 Nennungen) und zum anderen die genaue Wirtschaftszweigbezeichnung, wenn bei dieser Frage mit sonstiges, und zwar ... geantwortet wurde (539 Eintragungen). In den Datensatz mit den Volltexten aufgenommen wurde auch die nuts3-Variable, weil sie aus einem Buchstaben- und Zifferncode besteht.

Die Bereitstellung der Teildatensätze 2.) bis 4.) für die jeweiligen Jahre erfolgt formlos. Allerdings ist – im Unterschied zum Grunddatensatz - die Verwendung von Variablen aus den Teildatensätzen bei der Beantragung des Datenzugangs explizit zu nennen und gesondert bezüglich des Verwendungszwecks zu begründen. Alle Teildatensätze verfügen über den gleichen Pseudo-Identifikator für eine Erhebungseinheit (lfd), mit dem sich die Informationen einfach zusammenführen lassen.

Der Vollständigkeit halber sei erwähnt, dass aus Datenschutzgründen einige sehr wenige Variablen (vier) der Kosten- und Nutzenerhebung für das Jahr 2007 der externen Wissenschaft *nicht* für Auswertungen zur Verfügung stehen. Hierzu gehören Fragen nach der Funktion und Betriebszugehörigkeitsdauer der Auskunft gebenden Personen ebenso wie individuelle Betriebsangaben (etwa die E-Mail-Adresse oder die Betriebsnummer).

7 FDZ-Angebot für die beiden Datensätze Kosten und Nutzen der Berufsausbildung für die Jahre 2000 und 2007

Als Arbeitshilfen werden über das BIBB-FDZ die Fragebögen (pdf-Files) sowie Spieldaten zur Verfügung gestellt (Abschnitt 7.1). Als Zugangswege zu den Daten sind die Datenfernverarbeitung sowie Gastaufenthalte möglich (7.2). Für einen Eindruck zum Forschungspotential des Datensatzes werden einige zentrale Publikationen mit Analyseergebnissen der Kosten-Nutzen-Daten aufgelistet (7.3).

7.1 Arbeitshilfen

- **Fragebögen**

Über die Internetseiten des BIBB-FDZ (www.bibb-fdz.de) können die Fragebögen der Kosten-Nutzen-Erhebung für die Jahre 2000 und 2007 abgerufen werden. Beim Kosten-Fragebogen handelt es sich um einen face-to-face eingesetzten Fragebogen, der Nutzen-Fragebogen ist als schriftlicher Selbstausfüller (PAPI) angelegt. Der Fragebogen für das Jahr 2007 ist eine Kurzversion des CAPI-Fragebogens, der um datenschutzrechtlich problematische Fragen (vgl. letzter Absatz aus Kapitel 6) gekürzt ist.

- **Spieldaten**

Auf den Internetseiten des BIBB-FDZ können für Stata und SPSS formal anonymisierte Spieldatensätze der Kosten-Nutzen-Befragungen für die Jahre 2000 und 2007 heruntergeladen werden. Die Spieldaten sind frei zugänglich und gleichen in Aufbau und Merkmalsausprägungen den Originaldaten. Hierdurch besteht hohe Transparenz über das Analysepotential der Daten. Mit den Spieldaten sind jedoch keine inhaltlichen Auswertungen möglich. Vielmehr dienen sie - neben der Gewinnung eines Überblicks über die Daten - im Datenzugangsweg Datenfernverarbeitung Nutzern zur Kontrolle der Lauffähigkeit ihrer Auswertungsprogramme, bevor diese im BIBB-FDZ an den Originaldaten ausgeführt werden.

In den frei zugänglichen Spieldaten auf unseren Web-Seiten wurden 2.518 bzw. 2.986 Datenzeilen generiert, d.h. die Spieldaten haben die gleiche Dateigröße wie die Originaldaten (Stata: etwa 4,5 MB; SPSS: etwa 2,3 MB). Bei Bedarf kann ein von den Fallzahlen her reduzierter Spieldatensatz formlos per E-Mail übermittelt werden.

Die Spieldaten enthalten aus Datenschutzgründen keine Variablen mit Volltexten. Spieldaten mit Volltexten können von Anwendern bei Vorliegen eines Nutzungsvertrages per E-Mail angefordert werden. Die sensiblen Variablen der Kosten-Nutzen-Erhebung für das Jahr 2000 (vgl. Übersicht 2) bzw. 2007 (vgl. Übersicht 3) sind - in inhaltlich verfälschter Form - in den Spieldaten vorhanden.

7.2 Datenzugangswegen

Die Daten der Befragung „Kosten und Nutzen der betrieblichen Berufsausbildung“ für die Jahre 2000 und 2007 sind – wie alle Betriebsdaten – grundsätzlich nur über Datenfernverarbeitung und im Gastwissenschaftlermodell verfügbar. Nähere Informationen zu den Datenzugangswegen befinden sich auf unseren Internetseiten. Sofern Anwender sensible Zusatzmerkmale auswerten möchten, bitten wir dies im Nutzungsantrag an der dafür vorgesehenen Stelle gesondert zu vermerken und zu begründen. Im Anhang A2 sind diese Variablen für das Jahr 2000 an der kursiven Schrift erkennbar bzw. für das Jahr 2007 in Übersicht 3 aufgelistet.

7.3 Bestimmungen des Datenschutzes

Der Zugang zu den Forschungsdaten des BIBB-FDZ erfolgt ausschließlich unter dem Gebot der Datensparsamkeit und der Einhaltung der geltenden Datenschutzregeln nach der Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung, DS-GVO) sowie ergänzender Vorschriften des Bundesdatenschutzgesetzes (BDSG 2018). Demnach dürfen für die Durchführung wissenschaftlicher Vorhaben Daten an die unabhängige wissenschaftliche Forschung weitergegeben werden, wenn die Herstellung eines Bezugs zu einer Erhebungseinheit nicht möglich ist ("Anonymität"). Es wird empfohlen, dass zugangs- und nutzungsberechtigte Forscher/-innen auf die Einhaltung von Datenschutzbestimmungen besonders verpflichtet sind (vgl. Nutzungsrichtlinien des BIBB-FDZ für die Datenfernverarbeitung und Gastaufenthalte).

Die Mitarbeiter/-innen des BIBB-FDZ nehmen nur zum Zwecke der Beratung, der Verbesserung des BIBB-FDZ-Services sowie zur Gewährleistung der Einhaltung des Datenschutzes Einblick in Forschungsfragen, Methoden und Analysen der Wissenschaftlerinnen und Wissenschaftler. Mitarbeiterinnen und Mitarbeiter des BIBB, die nicht dem BIBB-FDZ angehören, erhalten keinen Einblick in die Tätigkeiten der Wissenschaftlerinnen und Wissenschaftler.

7.4 Forschungspotential: Veröffentlichungen (Auswahl)

Die nachfolgende Zusammenstellung ist mehr oder minder willkürlich und gibt nur einen kleinen Ausschnitt aus dem gesamten Analysepotenzial der Kosten-Nutzen-Erhebungen wieder. Für eine vollständige Recherche wird auf die Literaturlistenbank Berufliche Bildung (LdBB) des BIBB verwiesen. Unter Eingabe von „Kosten und Nutzen der Berufsausbildung“ wird eine sehr ausführliche Trefferliste zum Thema angezeigt. Die entsprechend angezeigten Publikationen nutzen in vielen Fällen die Kosten-Nutzen-Erhebung(en) als Datengrundlage, die nachfolgend aufgelisteten allesamt. Zentral sind die beiden Publikationen aus den jeweiligen Forschungsprojekten:

Beicht, U./ Walden, G./ Herget, H. (2003): Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland. Bundesinstitut für Berufsbildung. Der Generalsekretär (Hrsg.): Berichte zur beruflichen Bildung, Heft 264, Bonn

Schönfeld, G./ Wenzelmann, F./ Dionisius, R./ Pfeifer, H./ Walden, G. (2010): Kosten und Nutzen der dualen Ausbildung aus Sicht der Betriebe – Ergebnisse der vierten BIBB-Kosten-Nutzen-Erhebung. Bundesinstitut für Berufsbildung (Hrsg.), Berichte zur beruflichen Bildung, Bertelsmann-Verlag, Bielefeld

Die beiden Publikationen geben einen detaillierten Einblick in die Erhebungsmethodik und die Berechnungsweisen der Kosten- und Nutzenerhebungen für die Jahre 2000 und 2007. Darüber hinaus werden eine Vielzahl an deskriptiven und multivariaten Auswertungen berichtet. In der 2010er-Veröffentlichung befindet sich außerdem ein eigenes Kapitel, das in methodischer und inhaltlicher Perspektive die Untersuchungen für die Jahre 2000 und 2007 gegenüberstellt. Nachfolgend sind ohne Anspruch auf Vollständigkeit einige weitere Publikationen mit den Daten der Kosten-Nutzen-Erhebungen aufgelistet.

Bardeleben, R. von/ Beicht, U./ Fehér, K. (1995): Betriebliche Kosten und Nutzen der Ausbildung: Repräsentative Ergebnisse aus Industrie, Handel und Handwerk. Bundesinstitut für Berufsbildung. Der Generalsekretär (Hrsg.): Berichte zur beruflichen Bildung, Heft 210, Bielefeld

Beicht, U., Herget, H., et al. (2004): Costs and benefits of in-company vocational education and training in Germany. Bielefeld, Bertelsmann

Beicht, U./ Walden, G. (2002): Wirtschaftlichere Durchführung der Berufsausbildung - Untersuchungsergebnisse zu den Ausbildungskosten der Betriebe. In: Berufsbildung in Wissenschaft und Praxis, Heft 6/2002: 38 - 43

- Beicht, U./ Walden, G./ Herget, H. (2004):* Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland.
- Bundesinstitut für Berufsbildung (2002): Abschlussbericht zum Forschungsprojekt 2.9005. Nutzen und Nettokosten der Berufsausbildung für Betriebe, Bonn
- Bundesinstitut für Berufsbildung (2009): Abschlussbericht zum Forschungsprojekt 2.1.203 Kosten und Nutzen der betrieblichen Berufsausbildung. Bonn, URL: http://www2.bibb.de/tools/fodb/pdf/eb_21203.pdf (letzter Zugriff am 29. 07. 2010)
- Dionisius, R., S. Mühlemann, et al. (2009):* Ausbildung aus Produktions- oder Investitionsinteresse? Einschätzungen von Betrieben in Deutschland und der Schweiz. Zeitschrift für Berufs- und Wirtschaftspädagogik **105**(2): 267-284
- Dionisius, R., S. Mühlemann, et al. (2009):* Costs and benefits of apprenticeship training. A comparison of Germany and Switzerland. Applied economics quarterly **55**(1): 7-37.
- infas (2001): Nutzen und Nettokosten der Berufsausbildung. Methodenbericht zu einer repräsentativen Betriebsbefragung. internes Dokument, Bonn
- infas (2008): Betriebsbefragung zu den Kosten und dem Nutzen der betrieblichen Berufsausbildung – Methodenbericht. Internes Dokument, Bonn
- Mohenweiser, J./ Backes-Gellner, U. (2008): Apprenticeship Training - What for? Investment in Human Capital or Substitution of Cheap Labour? Working Paper / Swiss Leading House Economics of Education 17
- Mühlemann, S./ Pfeifer, H./ Walden, G./ Wenzelmann, F./ Wolter, Stefan C. (2010): The financing of apprenticeship training in the light of labor market regulations. Labour economics 17 (5): 799-809
- Sachverständigenkommission „Kosten und Finanzierung der beruflichen Bildung (1974): Kosten und Finanzierung der außerschulischen beruflichen Bildung, Bielefeld
- Pfeifer, H./ Wenzelmann, F./ Schöfeld, G. (2010): Ausbildungskosten und das Übernahmeverhalten von Betrieben - Ein Vergleich der BIBB-Kosten- und Nutzenerhebungen der Jahre 2000 und 2007. Sozialer Fortschritt 59 (6-7): 174-182
- Walden, G. (2007):* Short-term and long-term benefits as determinants of the training behaviour of companies. Zeitschrift für Arbeitsmarktforschung **40**(2/3): 169-191.
- Walden, G./ Herget, H. (2002)* Nutzen der betrieblichen Ausbildung für Betriebe - erste Ergebnisse einer empirischen Erhebung. in: Berufsbildung in Wissenschaft und Praxis, Heft 6/2002: 32 - 37

Anhang

A1: Übersicht der berücksichtigten Ausbildungsberufe in den Kosten-Nutzen-Erhebungen für das Jahr 2000 bzw. 2007

2000	2007
Industrie und Handel	
	Bankkaufmann/frau Bauzeichner/-in Bürokaufmann/frau Chemielaborant/-in Chemikant /-in Drucker/-in
Energieelektroniker/-in - Betriebstechnik Fachinformatiker/-in Fachkraft für Lagerwirtschaft	Energieelektroniker/-in für Betriebstechnik Fachinformatiker/-in - Systemintegration Fachkraft für Lagerlogistik
	Florist /-in Hotelfachmann/frau Industriekaufmann/frau
Industriemechaniker/-in – Betriebstechnik Industriemechaniker/-in – Maschinen- u. Systemtechnik	Industriemechaniker/-in
	Informatik Kaufmann/frau Informations- u. Telekommunikationssystem-Elektroniker/-in Informations- u. Telekommunikationssystem-Kaufmann/frau Kaufmann/frau für Bürokommunikation Kaufmann/frau im Einzelhandel Kaufmann/frau im Groß- u. Außenhandel Koch/Köchin Mechatroniker/-in
Mediengestalter/-in für Digital- u. Printmedien	Mediengestalter/-in für Digital- u. Printmedien - Mediendesign
	Reiseverkehrskaufmann/frau Veranstaltungskaufmann/-frau Kaufmann/-frau für Versicherungen und Finanzen
Versicherungskaufmann/frau	
(Fortsetzung nächste Seite)	
2000	2007
Werkzeugmechaniker/-in – Stanz- und Umformtechnik	Werkzeugmechaniker/-in
Handwerk	
	Augenoptiker Bäcker/-in Bürokaufmann/frau
Elektroinstallateur/-in Fachverkäufer/-in im Nahrungsmittelhandwerk	Elektroniker/-in – Energie- u. Gebäudetechnik Fachverkäufer/-in im Lebensmittelhandwerk
	Fleischer/-in Friseur/-in
Gas- und Wasserinstallateur/-in Zentralheizungs- u. Lüftungsbauer/-in	Anlagenmechaniker/-in für Sanitär-, Heizungs- und Klimatechnik
	Informationselektroniker/-in
Kraftfahrzeugmechaniker/-in	Kraftfahrzeugmechatroniker/-in

Maler/-in und Lackierer/-in

Maurer/-in

Metallbauer/-in

Tischler/-in

Zahntechniker/-in

Landwirtschaft

Gärtner/-in

Landwirt/-in

Freie Berufe

Arzthelfer/-in

Medizinische/-r Fachangestellte/-r

Rechtsanwaltsfachangestellte/-r

Steuerfachangestellte/-r

Zahnarzthelfer/-in

Zahnmedizinische/-r Fachangestellte/-r

Öffentlicher Dienst

Fachangestellte/-r für Bürokommunikation

Sozialversicherungsfachangestellte/-r

Vermessungstechniker/-in

Verwaltungsfachangestellte/-r

Impressum

BIBB-FDZ Daten- und Methodenberichte
Nr. 7/2010
Autor: Holger Alda

Herausgeber:
Bundesinstitut für Berufsbildung
Forschungsdatenzentrum
Robert-Schuman-Platz 3
53175 Bonn

Redaktion: Holger Alda
Redaktionsassistentin: Tanja Stierner

Downloads unter:
www.bibb-fdz.de

Tel.: +49-228-107-2041
Fax: +49-228-107-2020
E-Mail: fdz@bibb.de

ISSN-Nr.: 2190-300X

Der Inhalt dieses Werkes steht unter einer [Creative Commons Lizenz](http://creativecommons.org/licenses/by-nc-nd/4.0/) (Lizenztyp: Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 4.0 Deutschland). Weitere Informationen finden Sie unter www.bibb.de/cc-lizenz.