

**BIBB-FDZ**  
**Daten- und Methodenberichte**

**Nr. 4 / 2009**

**Kosten und Nutzen der betrieblichen  
Berufsausbildung in Deutschland 2000**

Version 1.1

Holger Alda  
Stefan Koscheck

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](mailto:fdz@bibb.de)

[www.bibb-fdz.de](http://www.bibb-fdz.de)



Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland 2000,

Autoren: Holger Alda, Stefan Koscheck

<b>1.</b>	<b>Einleitung</b> .....	3
<b>2.</b>	<b>Kosten und Nutzen der betrieblichen Berufsausbildung im Überblick</b> ....	4
<b>3.</b>	<b>Datengenese</b> .....	7
3.1.	Grundgesamtheit / Untersuchungseinheiten .....	7
3.2.	Stichprobenziehung (regionale Konzentrierung) .....	7
3.3.	Erhebungsdesign .....	8
3.4.	Ausschöpfung und Einhaltung des Stichprobenplans .....	10
<b>4.</b>	<b>Variablenspektrum und Datenaufbereitungen</b> .....	11
4.1	Variablenspektrum .....	11
4.2	Korrekturen und Plausibilitätsprüfungen .....	13
4.3	Datenaufbereitungen im BIBB-FDZ .....	15
4.4	Imputation fehlender Werte .....	18
4.5	Gewichtung und Hochrechnung .....	22
<b>5.</b>	<b>Zusammenführung der Datensätze</b> .....	23
<b>6.</b>	<b>Hinweise zur Datennutzung</b> .....	25
<b>7.</b>	<b>FDZ-Angebot zum Datensatz Nutzen und Nettokosten der Berufsausbildung aus dem Jahr 2000</b> .....	26
7.1	Arbeitshilfen .....	27
7.2	Datenzugangswege .....	27
7.3	Bestimmungen des Datenschutzes .....	28
7.4	Forschungspotential: Veröffentlichungen (Auswahl) .....	28
<b>8</b>	<b>Ausblick</b> .....	29

## 1. Einleitung

Das vorliegende Daten- und Methodenbericht beschreibt den Datensatz „Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland 2000“ und dient externen WissenschaftlerInnen als Arbeitshilfe für die Auswertung der Daten. Die Daten wurden von infas, Institut für angewandte Sozialforschung (Bonn), im Auftrag des Bundesinstituts für Berufsbildung (BIBB) erhoben. Das Forschungsdatenzentrum des BIBB (BIBB-FDZ) hat weitere Aufbereitungsschritte an den Daten vorgenommen, die im vorliegenden Handbuch dokumentiert sind. Für die externe nicht-kommerzielle Forschung bestehen Datenzugangsmöglichkeiten über das BIBB-FDZ mittels der Datenfernverarbeitung und/ oder Gastaufenthalten<sup>1</sup>.

Anlass und Grundlage für alle Kosten-Nutzen-Erhebungen der betrieblichen Berufsausbildung ist ein Modell zur Erfassung betrieblicher Ausbildungskosten, wie sie 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 des methodischen Vorgehens zur Ermittlung und Berechnung von Kosten und Nutzen der betrieblichen Berufsausbildung – etwa für das Jahr 2001 und für den hier beschriebenen Datensatz - befindet sich in der Abschlusspublikation des entsprechenden Forschungsprojektes (Beicht et al., 2004). Dort wird auch die zugrunde gelegte methodische Vorgehensweise bei Voll- und Teilkostenrechnung der betrieblichen Berufsausbildung detailliert erläutert. Die Lektüre der Abschlusspublikation ist daher insbesondere bei Analyseinteressen zu empfehlen, die sich besonders intensiv mit den Kosten- und Nutzenaspekten der betrieblichen Berufsausbildung beschäftigen<sup>2</sup>.

Ein weiterer Teil des Forschungsprojekts aus dem Jahr 2001 war eine schriftliche Zusatzerhebung bei nicht ausbildenden Betrieben zu einem potentiellen Nutzen der Berufsausbildung. Auf diese Daten wird im vorliegenden Handbuch *nicht* eingegangen<sup>3</sup>.

Das Daten- und Methodenhandbuch ist wie folgt gegliedert. Kapitel 2 fasst die zentralen Ziele und wesentlichen Eigenschaften der Untersuchung zusammen. Zudem werden Unterschiede zu vorherigen Erhebungen (etwa aus dem Jahr 1991) herausgestellt. 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.

---

<sup>1</sup> 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.

<sup>2</sup> Der Datensatz erlaubt in einem gewissen Umfang auch die empirische Behandlung anderer Auswertungsinteressen, da er – 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)

<sup>3</sup> Zudem liegen die Daten der schriftlichen Zusatzerhebung bisher nicht dem BIBB-FDZ vor.

Da bis zur Aufbereitung der Daten im BIBB-FDZ kein Datensatz existierte, der alle Variablen der Kosten-Nutzen-Erhebung 2001 enthielt, beschreibt das fünfte Kapitel, wie aus den einzelnen Teildatensätzen der über das BIBB-FDZ für die externe Wissenschaft aufbereitete Gesamtdatensatz aufgebaut ist. Kapitel 6 geht auf die Handhabung des Datensatzes und die Arbeit mit den Variablen der Kosten- und Nutzen-Erhebung 2001 ein. Das FDZ-Angebot der Arbeitshilfen zum Kosten/Nutzen-Datensatz wird im siebten Kapitel beschrieben. Abschließend wird im Ausblick (Kapitel 8) ein kurzer Ausblick auf die Daten der Kosten/Nutzen-Untersuchung des Jahres 2007 gegeben. Diese werden im Jahr 2010 ebenfalls der externen Wissenschaft für Auswertungen zur Verfügung gestellt.

Für die vorliegende Datendokumentation und den damit korrespondierenden Datensatz sei angemerkt, dass es sich dabei um die Erschließung eines Datenbestandes handelt, der relativ weit in der Vergangenheit erhoben und ausgewertet wurde. Die Rekonstruktion der Originaldaten erfolgte in Kooperation mit den noch im Institut beschäftigten und damaligen ProjektmitarbeiterInnen. Aus verschiedenen Teildatensätzen ist ein Gesamtdatensatz entstanden (vgl. dazu auch Kapitel 5). Um die Arbeit externer WissenschaftlerInnen mit den Daten zu erleichtern, werden auf Antrag zusätzlich einige zusammenfassende Variablen der Modellkostenrechnungen des BIBB für eigene Auswertungen zur Verfügung gestellt.

In diesem Daten- und Methodenbericht sind alle uns bekannten Informationen zusammengestellt, die wir während dieses Rekonstruktionsprozesses erhalten haben. Den Beteiligten im BIBB, insbesondere Herrn Dr. Walden, Frau U. Beicht, Herrn H. Pfeifer, F. Wenzelmann 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. An einigen wenigen Stellen konnten Informationsdefizite nicht vollständig befriedigend aufgelöst werden. Bei entsprechenden Nachfragen bitten wir Interessierte, sich zunächst an das FDZ zu wenden.

## **2. Kosten und Nutzen der betrieblichen Berufsausbildung im Überblick**

Die Befragung „Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland“ aus dem Jahr 2001 hatte das Ziel, für ausgewählte Ausbildungsberufe ein möglichst detailliertes Bild über die Kosten der betrieblichen Ausbildung 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 Nutzen der beruflichen Bildung 1974) erfolgt eine Erhebung der Kosten (und Nutzen) der betrieblichen Berufsausbildung durch das BIBB etwa alle zehn Jahre. Die letzte Untersuchung von Ausbildungskosten für Betriebe stammte aus dem Jahr 1991 (Bardeleben et al., 1995). Auf Basis dieser Untersuchung wurden Fortschreibungen für die Jahre 1995 und 1997 vorgenommen. Aufgrund der Veränderungen in der Praxis der betrieblichen Berufsausbildung, die auch Auswirkungen auf die Höhe und Struktur der Ausbildungskosten hatten, waren anschließende Fortschreibungen nicht mehr tragfähig (Beicht et al., 2004: 9).

Wenngleich es sich bei der Kosten-Nutzen-Erhebung aus dem Jahr 2001 im Prinzip um eine Wiederholungsbefragung des Jahres 1991 handelt, unterscheidet sie sich von dieser in zentralen Dimensionen:

- Im Jahr 2001 wurde die gesamte deutsche Wirtschaft sowie der öffentliche Dienst einbezogen, im Jahr 1991 nur die Ausbildungsbereiche Industrie und Handel sowie das Handwerk.
- Die Daten der Erhebung 2001 beziehen sich auf Gesamtdeutschland, diejenigen von 1991 ausschließlich auf die alten Bundesländer.
- Im Jahr 2001 wurden pro Betrieb die Ausbildungskosten in einem ausgewählten Ausbildungsberuf erfasst, im Jahr 1991 dagegen die Gesamtkosten für alle im Betrieb angebotenen Ausbildungsberufe.
- Im Jahr 2001 erfolgte eine wesentlich umfassendere Untersuchung des betrieblichen Nutzens der Berufsausbildung als im Jahr 1991. Erstmals wurden repräsentative Daten zu durch Ausbildung eingesparten Rekrutierungskosten externer Fachkräfte erfasst.

Aufgrund dieser Unterschiede sind Vergleiche zwischen den Daten aus dem Jahr 2001 und den Daten aus dem Jahr 1991 (bzw. den auf dieser Basis erstellten Fortschreibungen) nur bedingt möglich.

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 setzt 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 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 als typisch für die jeweiligen Ausbildungsbereiche angesehen wurden<sup>4</sup>. 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 Gewichtungungsverfahren 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).

In Tabelle 1 sind die zentralen Eigenschaften der Kosten-Nutzen-Erhebung des Jahres 2001 zusammengefasst.

<sup>4</sup> Eine vollständige Auflistung dieser Ausbildungsberufe befindet sich im Anhang A1.

**Tabelle 1: Zentrale Eigenschaften der Kosten-Nutzen-Erhebung aus dem Jahr 2001**

<b>Befragungstitel</b>	<b>Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland 2000</b>			
DOI	<a href="https://doi.org/10.7803/370.07.1.2.10">doi:10.7803/370.07.1.2.10</a>			
Kurzbeschreibung	Detaillierte Erfassung von Kosten und (monetärem und nicht monetär bewertetem) Nutzen der betrieblichen Berufsausbildung in Deutschland pro Auszubildendem und Ausbildungsjahr für die 52 am stärksten besetzten Ausbildungsberufe.			
Erhebungsjahr	2001			
Erhebungseinheit	Ausbildungsbetriebe			
Themenschwerpunkt	Ausbildung			
Datenzugangsmöglichkeiten	Gastwissenschaftleraufenthalt/Datenfernverarbeitung, Zusatzvariablen (für Volltext-, Sonder- od. regionale Variablen)			
Variablenanzahl <sup>5</sup>	GWA/DFV: 487	VT: 3	SV: 41	REG: 7
Grundgesamtheit	alle Ausbildungsbetriebe in Deutschland, Kammerauswahl (2000)			
Gewichtung/Hochrechnung	Anpassungsgewichtung (Iterative Proportional Fitting) und Hochrechnungsvariable			
repräsentative Region	Ost/West			
Fallzahl	2.518			
Erhebungsverfahren	persönliche Befragung			
Auswahlverfahren	Zufallslistenauswahl der Kammern (2000)			
Erhebungsdesign	Längsschnittdesign (unverbundene Querschnitte als zeitliche Trendstudie)			
Bemerkung	International vergleichende Daten zu den betrieblichen Ausbildungskosten liefert die Betriebsbefragung EU-Labour-Cost-Survey; Schweiz: Kosten und Nutzen der Lehrlingsausbildung aus der Sicht Schweizer Betriebe; die Nettoszahl basiert auf einem anspruchsvollen Screening und wurde nicht voll ausgeschöpft			
Links	BIBB-FDZ Metadatenportal: <a href="http://metadaten.bibb.de/metadaten/16">http://metadaten.bibb.de/metadaten/16</a> Projektseite: <a href="https://www.bibb.de/de/11060.php">https://www.bibb.de/de/11060.php</a>			
Stichworte	Ausbildungskosten, Nutzen der Ausbildung für Betriebe, betriebliche Berufsausbildung, Ausbildungsbetrieb, Fachkräftebedarf, Nachwuchsbedarf, Nachwuchssicherung, Personalentwicklung, Ausbildungsbereitschaft			

<sup>5</sup> Jeweils ohne Personen-/Spellkennzeichen.

### 3. Datengenese

Im Folgenden wird das in Kapitel 2 gesagte bezüglich der Datenerhebung des Forschungsprojekts Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland 2001 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) werden schließlich Informationen zum Rücklauf der Befragung und der Erfüllung der Vorgaben des Stichprobenplans gegeben.

#### 3.1. Grundgesamtheit / Untersuchungseinheiten

Im Gegensatz zur vorangegangenen Kosten/Nutzen-Befragung des Jahres 1991 wurden in der Erhebung 2001 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). Aus diesem Grund wurden zunächst 52 Ausbildungsberufe aus den die gesamte Wirtschaft<sup>6</sup> abdeckenden Ausbildungsbereichen (Industrie und Handel, Handwerk, Landwirtschaft, Freie Berufe sowie Öffentlicher Dienst) in eine Schichtungsmatrix überführt. Dies geschah nach der Maßgabe, dass es sich bei den ausgewählten Ausbildungsberufen um für die jeweiligen Ausbildungsbereiche als „typisch“ angesehene handeln soll<sup>7</sup>. Daneben fanden aber auch 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., 2003: 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 ausbilden<sup>8</sup>.

#### 3.2. Stichprobenziehung (regionale Konzentrierung)

Zwar erfolgte die Stichprobenziehung wie eben beschrieben nach wichtigen Ausbildungsbereichen und –berufen, aber aus Kostengründen erfolgte zusätzlich eine regionale Konzentration der Erhebung. Da die Stichprobenziehung der Betriebsadressen durch die Kammern bzw. die zuständigen Stellen erfolgte, wurde zunächst nach einem Zufallsverfahren eine Auswahl einzelner Kammerbezirke vorgenommen (Bundesinstitut für Berufsbildung 2002: 3; Beicht et al. 2004: 15). Sehr kleine Kammern - gemessen an der Zahl

<sup>6</sup> Die (sehr kleinen) Wirtschaftszweige Seeschifffahrt und Hauswirtschaft werden nicht erfasst.

<sup>7</sup> Zentrale Kriterien für die Auswahl sind die Besetzungstärke sowie die Abdeckung eines breiten inhaltlichen Ausbildungsspektrums.

<sup>8</sup> Mit den in Abschnitt 4.5 vorgestellten Hochrechnungsfaktoren werden Analyseergebnisse auf die Gesamtanzahl aller Ausbildungsbetriebe und aller Auszubildenden in Deutschland gewichtet.

der Auszubildenden - wurden nicht berücksichtigt. Eine Übersicht über die Anzahl der für die jeweiligen Ausbildungsbereiche ausgewählten Kammern bzw. zuständigen Stellen sowie den dabei erzielten Repräsentationsgrade ist Beicht et al. (2004: 15 f.) zu entnehmen.

Von den einbezogenen Kammern und zuständigen Stellen wurden auf Grundlage der Ausbildungsrollen per Zufallsauswahl 13.000 Betriebsadressen gezogen, welche in den ausgewählten Berufen ausbilden. In einem vom BIBB entwickelten Stichprobenplan mit insgesamt 708 Zellen wurde die Zahl der pro Ausbildungsberuf und Kammerbezirk zu realisierenden Interviews festgelegt. Als durchschnittliche Sollzahl pro Beruf wurden etwa 50 Interviews angestrebt, für einige Berufe mit relativ homogenen Ausbildungsstrukturen reichten etwa 40 Interviews (ebd.: 13), was einer Nettostichprobe von etwa 2.500 Betrieben entspricht. Eine Ausnahme der Sollvorgaben einer Besetzungsstärke von 50 Fällen sind vier Berufe, die überwiegend Großbetriebe ausbilden.<sup>9</sup> Für diese wurden überproportionale Sollzahlen von 100 bzw. 110 Interviews festgelegt, um große Mittelständler und Großbetriebe in der Stichprobe mit ausreichenden Fallzahlen zu repräsentieren (infas, 2001: 7).

Für den Feldeinsatz bzw. die Einhaltung des komplexen Zellenplans wurden von infas in drei Einsatzwellen aus den 13.000 Betriebsadressen insgesamt 7.125 Adressen gezogen (ebd.: 8). Ausgehend von einer gleich hohen Stichprobenausschöpfung in allen Regionen und Berufen wurde anschließend in einem ersten Schritt eine nach Ausbildungsberufen und Betrieben geschichtete Zufallsstichprobe gezogen. Im zweiten Schritt wurden die Adressen von bisher nicht erreichten Betrieben und situativen Verweigerern erneut eingesetzt bzw. im Bedarfsfall eine nach Beruf und Ausbildungsberuf geschichtete und den ersten Schritt ergänzende Zufallsstichprobe gezogen. Im dritten Schritt schließlich wurde nach Vorliegen der meisten Interviews zusätzlich zum erneuten Einsatz der Adressen Nicht-Erreichter und „weicher“ Verweigerer für einige wenige Berufe eine eigenständige Aufstockungstichprobe eingesetzt. Hiermit konnten auch in bis dato nicht ausreichend gefüllten Zellen die angestrebten Mindestzellbesetzungen erreicht werden. Im Anschluss an jede Ziehung erfolgte eine Prüfung, ob Betriebe durch Ausbildung in mehreren der ausgewählten Ausbildungsberufe mehrfach gezogen wurden. Traf dies zu, wurde der Betrieb nur für einen Ausbildungsberuf in der Stichprobe belassen, auch, um den Aufwand für die Betriebe auf ein zumutbares Maß zu begrenzen.

### 3.3. Erhebungsdesign

Die Befragung zu den Kosten und Nutzen der betrieblichen Berufsausbildung aus dem Jahr 2001 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

---

<sup>9</sup> Das sind die Ausbildungsberufe Industriemechaniker/-in Fachrichtung Betriebstechnik, Energieelektroniker/-in Fachrichtung Betriebstechnik, Industriekaufmann/-frau sowie Bürokaufmann/-frau.



als Face-to-Face-Interview durchzuführen. In den meisten Interviews (85,6 Prozent) wurde der Nutzenfragebogen im Beisein des Interviewers ausgefüllt (infas 2001: 16).

Da Kosten und Erträge der Berufsausbildung von Unternehmen i.d.R. 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 für je alle Ausbildungsjahre einzeln. Die Angaben beziehen sich jeweils auf das Vorjahr (das Jahr 2000).

Nach einem Pretest mit 22 Fällen im Herbst 2000<sup>10</sup> 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 insbesondere Personen mit besonderen Erfahrungen in Betriebsbefragungen eingesetzt. Im Durchschnitt wurden pro Interviewer 13 Interviews geführt. Das Durchschnittsalter der Interviewer betrug 45 Jahre. Die Schulung der Interviewer erfolgte anhand eines Handbuchs, das auch 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. Die Auswahl einer geeigneten Person in den Betrieben, die zu der Ausbildung in dem bestimmten Ausbildungsberuf auskunftsfähig war, erfolgte durch die Interviewer. In Kleinbetrieben wurden i.d.R. Leiter und Inhaber befragt, in mittleren und großen Betrieben die für die Ausbildung hauptverantwortlichen Personen (Beicht et al., 2004: 18).

Nach Beendigung des Interviews wurde die Interviewsituation von den Interviewern in einem separaten Fragebogenteil protokolliert. Nach Einschätzung der Interviewer waren zwei Drittel der Befragten sehr stark an der Untersuchung interessiert, ein Drittel der Befragten zeigte etwas Interesse. Die Auskunftsbereitschaft der Befragten wurde von den Interviewern zu etwa 80 Prozent als gut bzw. sehr gut eingestuft.

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.

---

<sup>10</sup> Im Pretest zeichneten sich Zugangsschwierigkeiten bei Vertretern der Freien Berufe (Ärzte und Rechtsanwälte) ab. Um die nach dem Stichprobenplan erforderlichen Interviews zu erzielen, wurde für die Hauptstudie beschlossen, den Bruttostichprobenansatz für die Freien Berufe zu erhöhen. Zudem bereiteten den Befragten genaue Kostenangaben für den ausgewählten Beruf Schwierigkeiten (infas, 2001: 5 f.).

**3.4. Ausschöpfung und Einhaltung des Stichprobenplans**

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 (relativ abrupten) Einstellen aller Ausbildungsaktivitäten) stand eine bereinigte Bruttostichprobe von 6.257 Ausbildungsbetrieben zur Verfügung. Das führte zu 2.518 auswertbaren Interviews, was einer Ausschöpfungsquote von 40,2 Prozent entspricht. In 22 Fällen wurde lediglich der Fragebogenteil zu den Ausbildungskosten beantwortet, sodass für die Nutzendimension 2496 Betriebe analysiert werden können. Details zur Ausschöpfung der Stichprobe sowie Ausfallgründen befinden sich in Tabelle 2.

**Tabelle 2: Ausschöpfung und Ausfallgründe bei der Betriebsbefragung „Nutzen und Nettokosten der Berufsausbildung“ aus dem Jahr 2000**

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

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. Nur 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.

#### **4. Variablenspektrum und Datenaufbereitungen**

Zuerst wird ein Überblick über das Variablenspektrum der Kosten-Nutzen-Erhebung aus dem Jahr 2001 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 aus dem Jahr 2001 enthält für die folgenden Themenblöcke Variablen.

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

Themenblock	Anzahl Variablen	Variablenbereich	Fragebogen-seiten (von – bis)
<b>I. Kosten der Berufsausbildung</b>			
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
<b>II. Nutzen der Berufsausbildung</b>			
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
<b>III. Zusatzvariablen</b>			
	24	lfd – ausbstd_azubi	trifft nicht zu: eine Übersicht befindet sich in Anhang A2
<b>IV. synthetische Variablen</b>			
	6	x1 – x6	trifft nicht zu; zur Beschreibung vgl. 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 (125 Variablen).

#### 4.2 Korrekturen und Plausibilitätsprüfungen

Eine erste Datenprüfung wurde vom Erhebungsinstitut vorgenommen. Vor der Erfassung der Fragebögen wurde zunächst etwa geprüft, ob sich das Interview auch tatsächlich auf den laut Stichprobenplan ausgewählten Ausbildungsberuf bezogen hat, ob der Betrieb zum Stichtag 30.09.2000 tatsächlich Auszubildende beschäftigte, ob es sich bei dem Betrieb evtl. um eine überbetriebliche Ausbildungsstätte handelte sowie, ob der Fragebogen vollständig ausgefüllt wurde (infas 2001: 17). Nach manueller Prüfung der Befragungsdaten auf Vollständigkeit während der Datenerfassung wurden die Daten formal auf Vollständigkeit, Filterfehler und Inkonsistenzen geprüft.

Über die formelle Datenprüfung hinaus erfolgte durch infas anhand einer vom BIBB vorgegebenen Aufstellung zu prüfender Sachverhalte erste Plausibilitätsprüfungen. In erster Linie handelte es sich dabei um die Prüfung der Über- bzw. Unterschreitung als plausibel erachteter Ober- und Untergrenzen. Aufgrund branchentypischer oder regionaler Besonderheiten erwies sich ein Teil der Prüfkriterien als zu eng (infas 2001: 27). Bei Unplausibilitäten, für die weder aus dem Fragebogen noch aus dem Gesamtkontext Erklärungen erschlossen werden konnten, wurde mit den betroffenen Betrieben telefonische Rücksprache gehalten. Die Ergebnisse der Plausibilitätsprüfung wurden in gesonderten Prüfvariablen abgelegt. Für einige Variablen ergeben sich Datenkorrekturen.

- **f32a\_jk sowie f34a\_jk (Anzahl der Teilnehmer an einzelnen betrieblichen Ausbildungsmaßnahmen):** Anhand der Fragebögen und telefonischen Rückfragen wurden auf Missverständnissen beruhende Fehlangaben korrigiert.
- **f39\_1k (Anzahl zusätzlicher Fachkräfte) und f39\_2k (Anzahl zusätzl. An-/Ungelernter):** Prüfkriterium war, dass die Summe beider Angaben nicht größer sein darf als die Anzahl der Auszubildenden im ausgewählten Beruf insgesamt. In Fällen, in denen nur zu einer Mitarbeiterart eine Angabe erfolgte und die Differenz in den Angaben zum Ersatzbedarf und der Gesamtzahl der Auszubildenden auf ein Schätzproblem zurückgeführt werden konnte, wurde die zu hohe Nennung durch die Gesamtzahl der Auszubildenden ersetzt. In Fällen, in denen zu beiden Mitarbeitergruppen Angaben gemacht wurden und bei denen ebenfalls Schätzfehler zu Abweichungen führten, wurde eine anteilige Reduktion der Angaben auf die Gesamtzahl der Auszubildenden im ausgewählten Beruf vorgenommen. Fälle, bei denen diese Korrektur nicht vorgenommen werden konnte, da sich die Angaben vermutlich nicht auf die Anzahl der Auszubildenden im ausgewählten Beruf, sondern auf die Gesamtzahl aller Auszubildenden (Frage 9) bezogen, wurden auf „keine Angabe“ gesetzt.

- **f45\_k und f46\_k (Beteiligung von haupt- und nebenberuflichen Ausbildern in der Ausbildung):** Sofern angegeben wurde, im Betrieb weder haupt- noch nebenberufliche Ausbilder einzusetzen und die Auszubildenden nicht extern oder überbetrieblich ausgebildet wurden, erfolgte eine telefonische Nachrecherche. Sofern der genannte Umstand auf einen der folgenden beiden Gründe zurückzuführen war, konnten die fehlenden Angaben nacherhoben und im Datensatz korrigiert werden: 1) Insbesondere in Krankenkassen erfolgt die Ausbildung überwiegend zentral in Ausbildungszentren bzw. hauptamtliche Ausbilder aus diesen Zentralen suchen die Auszubildenden von Zeit zu Zeit vor Ort auf. Außerdem rotieren Auszubildende oft zwischen einzelnen Geschäftsstellen. 2) Die Bezeichnung „nebenberuflicher Ausbilder“ wurde als nebenberufliche Tätigkeit missverstanden.
- **f47\_jk (Vergleich der Produktivitätsleistungen nebenberuflich ausbildender und nicht ausbildender Mitarbeiter):** Sofern telefonische Nachrecherchen zu f45 und f46 vorgenommen wurden, wurden auch die Angaben zu f47\_jk ergänzt.
- **f48\_k (Ausbildungsverwaltung im ausgewählten Beruf):** In der telefonischen Nachrecherche zu f45\_k und f46\_k wurde zudem f48\_k nacherhoben, sofern keine Beteiligung von Mitarbeitern in der Ausbildungsverwaltung angegeben war. Gründe für die fehlenden Werte waren Missverständnisse der Bezeichnung „nebenberuflicher Ausbilder“ sowie Betriebsinhaber und Geschäftsführer, welche die Ausbildungsverwaltung durchführen, sich jedoch nicht als Führungskräfte einordneten.
- **f50\_jk (durchschnittliche Bruttolöhne der an der Ausbildung bzw. Ausbildungsverwaltung beteiligten Mitarbeitergruppen):** Sofern telefonische Nachrecherchen zu f45\_k und f46\_k vorgenommen wurden, wurden auch die Angaben zu f50\_jk ergänzt.

Übersicht 1 stellt darüber hinaus Wertebereichskorrekturen von infas zusammen.

#### **Übersicht 1: Korrigierte Wertebereiche und Korrekturwerte originaler Variablen der Kosten-Nutzen-Erhebung aus dem Jahr 2001**

Variablenname	Bezeichnung	Korrigierte Wertebereiche	Korrekturwert
f57_jk	Anzahl der Lehrwerkstatttage im j. Ausbildungsjahr	mehr als 190 Tage (4. Ausbildungsjahr: mehr als 95 Tage)	190 Tage
f32b_jk / f34b_jk	durchschnittliche Abwesenheit pro Teilnehmer	mehr als 190 Tage (4. Ausbildungsjahr: mehr als 95 Tage)	190 Tage
f35_1jk	durchschnittliche Urlaubstage pro Auszubildendem im j. Ausbildungsjahr	weniger als 24 od. mehr als 35 Tage (4. Ausbildungsjahr: weniger als 12 bzw. mehr als 16 Tage).	24 bzw. 35 Tage
f35_2_jk	durchschnittliche andere freie Tage pro Auszubildendem im j. Jahr	mehr als 15 Tage (4. Ausbildungsjahr: mehr als 8 Tage)	15 Tage
f35_3_jk	durchschnittliche Berufsschultage pro Auszubildendem im j. Jahr	weniger als 39 Tage od. mehr als 100 Tage (4. Ausbildungsjahr: weniger als 20 bzw. mehr als 50 Tage)	39 bzw. 100 Tage
f35_4_jk	durchschnittliche Krankheitstage pro Auszubildendem im j. Jahr	mehr als 30 Tage (4.Ausbildungsjahr: mehr als 15 Tage)	30 Tage

Nach der Übermittlung der Befragungsdaten an das BIBB wurden weitere intensive inhaltliche Plausibilitätsprüfung vorgenommen (Beicht et al., 2004: 18). Insbesondere wurden bei nahezu allen Variablen weitere Mindest- und Höchstwerte festgelegt, welche jedoch nur von einer sehr kleinen Minderheit der befragten Betriebe unter- bzw. überschritten wurden. Aus Platz- und Übersichtsgründen wird auf eine entsprechende Darstellung verzichtet.

#### 4.3 Datenaufbereitungen im BIBB-FDZ

Die Variablen der Erhebung „Kosten und Nutzen der betrieblichen Berufsausbildung 2001“ 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 werden – soweit möglich – im BIBB-FDZ nachträglich rekonstruiert. Im Zuge der Datenzusammenführung und Datenaufbereitung hat das BIBB-FDZ die folgenden Arbeiten ausgeführt.

- **Deklarierung fehlender Werte**

Fehlende und ungültige Werte wurden auf die Standarddeklarierung fehlender Werte des BIBB-FDZ umkodiert.

**Übersicht 2: Standarddeklaration des BIBB-FDZ für verschiedene Arten fehlender Antwortwerte**

-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 aus dem Jahr 2001 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- 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 wurden (so genannte x-Variablen).

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.



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

- **Synthetische Variablen**

Um 3jährige von 3,5jährigen Ausbildungsberufen unterscheiden zu können, steht die Hilfsvariable lz\_3 zur Verfügung. Dreijährige Ausbildungsberufe sind mit „1“ codiert, 3,5-jährige mit „0“.

- **Datenaufbereitungen Kosten-Fragebogen**

Filtersetzung für fehlende Auszubildende nach Ausbildungsjahren

Sofern in Frage 20 für ein oder mehrere der Ausbildungsjahre (3jährige Ausbildung: max. 2; 4jährige Ausbildung: max. 3) angegeben wurde, dass zum Stichtag keine Auszubildenden vorhanden waren, wurden alle folgenden Kostenangaben für diese(s) Ausbildungsjahr(e) auf -2 umcodiert. Ebenso sind, falls in Frage 11 in dem jeweiligen Jahr keine Auszubildenden die Ausbildung erfolgreich abgeschlossen haben, die Angaben der Frage 13 auf -2 umcodiert.

Umkodierung einzelner ungültiger Werte

In Frage 46a\_2k trat einmal der Wert „999“ auf. Dieser wurde zu „-6“=ungültig umcodiert. In Frage f50\_5k trat einmal der Wert „22“ auf. Dieser wurde ebenfalls auf „-6“=ungültig gesetzt. In Frage f62\_1k trat fünfmal der Wert „-2“ auf, der auf „-6“=ungültig umcodiert wurde. Die gesondert codierte Antwort "gesamte Lehrzeit" in Frage f73\_2k zur durchschnittlichen Wochenzahl, die dreimal angegeben wurde, wurde in „52“ (Wochen) umcodiert. Die gesondert codierte Antwort „Gehalt wird übernommen“ auf die Frage f75\_2k zur Gesamtkostenerstattung für externe Auszubildende, die in einem Fall angegeben wurde, wurde zu „-6“=ungültig umcodiert. Ebenso wurde in Frage f77\_2k die Antwort „Prozentangabe und keine Berechnungsgrundlage angegeben“, die einmal angegeben wurde, zu „-6“=ungültig umcodiert.

- **Datenaufbereitungen Nutzen-Fragebogen**

f7a und f7b: Derzeitige Leistungsfähigkeit des Betriebs

Die Angaben zur derzeitigen Leistungsfähigkeit des Betriebes (f7a\_j\_1) wurden laut Kodierung des Fragebogens in den Interviews fünfstufig erfasst, die Variablen waren jedoch sechsstufig kodiert (sehr schlecht=„6“ statt „5“). Der Wert „6“ trat jedoch nur in zwei Variablen (f7a\_2\_1 und f7a\_6\_1) in je einem Fall auf. Die Werte „6“ wurden zu „5“ umcodiert und die Wertelabels wurden entsprechend der Kodierung des Fragebogens geändert („5“=sehr schlecht).

Auch die Variablen f7b\_j\_1 enthielten die sechsstufige Kodierung der Wertelabels. Der Wert „6“=sehr schlecht trat jedoch in keinem Fall auf. Daher mussten nur die Wertelabels an die Kodierung des Fragebogens angepasst werden.

#### f7a und f7b: Erwartung der zukünftigen Leistungsfähigkeit des Betriebs

Bei den dreistufig kodierten („1“=besser, „2“=gleich bleiben, „3“=sinken) Variablen zur erwarteten Veränderung der Leistungsfähigkeit der Betriebe in zwei bis drei Jahren (f7a\_j\_2 und f7b\_j\_2) kamen in Variable f7a\_2\_2 in je einem Fall die nicht belabelten Werte „4“ und „6“ vor, in den Variablen f7b\_2\_2 sowie f7b\_6\_2 je einmal der Wert „4“. Die genannten Werte wurden zu „-6“=ungültig umkodiert.

#### **4.4 Imputation fehlender Werte**

Eine Imputation fehlender Werte kann beim vorliegenden Datensatz auf Eigenarten des Untersuchungsgegenstands zurückzuführen sein, 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 alle nicht besetzten 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 Datenimputation, die sich teilweise beträchtlich im methodischen Vorgehen und dem Ergebnis unterscheiden<sup>11</sup>.

#### 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. Um diese gegebenenfalls von den originalen Variablen zu unterscheiden, haben sie im Variablennamen den Zusatz „\_i“. Die Auswahl basiert auf bereits durchgeführten Forschungsarbeiten mit den Kosten-Nutzen-Daten.

<sup>11</sup> Einen kurzen Überblick über gängige Imputationsverfahren gibt das BIBB-FDZ in der Dokumentation der Imputation der Lohnvariable der BIBB-BAuA-Erhebung des Jahres 2006.

**Übersicht 3: Originalvariablenamen und korrespondierende Variablenamen bei imputierten Werten bzw. synthetischen Variablen**

originaler Variablen-name	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
f21_1k	Ausbildungsvergütung ersten Lehrjahr im	1469 (941)	f21_1k_i	2496 (910)
f21_2k	Ausbildungsvergütung zweiten Lehrjahr im	1624 (1051)	f21_2k_i	2496 (1031)
f21_3k	Ausbildungsvergütung dritten Lehrjahr im	1530 (1161)	f21_3k_i	2496 (1153)
f21_4k	Ausbildungsvergütung vierten Lehrjahr im	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)
-	Durchschnitts-Bruttokosten (voll) in Euro – Modellrechnung	-	x1	2496 (19178)
-	Durchschnitts-Äquivalenzerträge (voll) in Euro – Modellrechnung	-	x2	2496 (7891)
-	Durchschnitts-Nettokosten (voll) in Euro – Modellrechnung (= x1 – x2) <sup>12</sup>	-	x3	2496 (11287)

<sup>12</sup> Die Differenzbildung muss dafür auf der einzelbetrieblichen Ebene erfolgen.

originaler Variablen-name	Bezeichnung	Anzahl gültiger Werte (Mittelwert)	Variablenname inklusive imputierter Werte	Anzahl gültiger Werte (Mittelwert)
-	Durchschnitts-Bruttokosten (teil) in Euro – Modellrechnung	-	x4	2496 (11202)
-	Durchschnitts-Nettokosten (teil) in Euro – Modellrechnung (= x4 – x2) <sup>11</sup>	-	x5	2496 (3311)
-	Durchschnitts-Substitutionserträge <sup>13</sup> (voll) in Euro – Modellrechnung	-	x6	2496 (15790)

Anmerkung: die ausgewiesenen Mittelwerte basieren auf ungewichteten Angaben

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 vorgeschlagene Vorgehen gebildet worden, aber mit dem FDZ-Auswertungsdatsatz und unter Verwendung der synthetischen Variablen werden nicht exakt die Eckwerte der Publikation von Beicht et al. (2004) erreicht. Die Gründe hierfür sind vielfältig (etwa: unterschiedliche Imputationsverfahren) und in der entsprechenden Publikation nachzulesen (ebd., Teil II). Der entscheidende Grund für die Abweichungen ist die Hochrechnung einmal nach Auszubildengewichten (Beicht et al.) und Betriebsgewichten (FDZ-Datensatz).

<sup>13</sup> Die Definition von Substitutionserträgen befindet sich in Beicht et al. (2004) auf den Seiten 198 ff.

- 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.

#### 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 über eine Gewichtung (Beicht et al., 2004: 19). 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 Grundgesamtheit aller Ausbildungsbetriebe in Deutschland). Die erste Form der Hochrechnung auf Basis der Auszubildendenzahlen diente 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 den Auszubildendenzahlen der Grundgesamtheit 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 Gewichts- und Hochrechnungsfaktoren berechnet: Zum einen wurden die Verteilungen der Auszubildenden in der Stichprobe an die Verteilungen der Auszubildenden in der Grundgesamtheit der 52 abgedeckten Berufe angepasst, zum anderen wurde eine Anpassung an die Grundgesamtheit aller 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 gleichzeitige 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 (vgl. Beicht et al., 2004: 19). 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 im FDZ-Datensatz Nutzen und Nettokosten der Berufsausbildung**

Kennwert	<i>alte Bundesländer</i>		<i>neue Bundesländer</i>	
	<b>b_gew</b>	<b>b_hoch</b>	<b>b_gew</b>	<b>b_hoch</b>
Mittelwert	1,001	196	0,982	186
Standardabweichung	0,946	185	0,806	153
Median	0,592	116	0,666	126
10-Perzentil	0,077	15	0,183	35
25-Perzentil	0,240	47	0,352	67
75-Perzentil	1,469	288	1,651	313
90-Perzentil	2,316	454	1,841	349

## 5. Zusammenführung der Datensätze

Wie eingangs erwähnt musste das volle Variablenspektrum der Erhebung Nutzen und Nettokosten der Berufsausbildung aus dem Jahr 2001 für die FDZ-Aufbereitung der Daten aus Teildatensätzen erschlossen werden. Darüber hinaus war bereits die erste Datenlieferung vom Erhebungsinstitut infas in drei Einzeldateien aufgeteilt. Zum einen war das ein Datensatz, der die Variablen des Kostenteils sowie die Gewichtungs- und Hochrechnungsfaktoren für die Betriebe enthält. Die zweite Teillieferung von infas ist mit der ersten insofern identisch, als dass ebenfalls der Kostenteil im Datensatz abgebildet wurde.

Im Unterschied zur ersten Teillieferung handelt es sich aber um einen Datensatz, „in dem die Ergebnisse auf alle in der Befragung ermittelten Auszubildenden umgesetzt sind“ (infas, 2001: 53). Da es sich in diesem Fall nicht mehr um Ursprungswerte der Befragung handelt, wurde dieser zweite Teildatensatz in den FDZ-Aufbereitungsarbeiten nicht berücksichtigt. Ein weiterer Grund ist, dass sich der dritte Teildatensatz von infas auf die Angaben der Betriebe über den Nutzen der Berufsausbildung bezieht. Eine Korrespondenz der Kosten- und Nutzteile ist demnach nur für Datensätze mit Betriebsgewichten gegeben, während Auszubildendengewichte nur für den Kostenteil verfügbar wären.

Die Variablennamen orientieren sich an den Fragennummern in den jeweiligen Fragebögen. Da im Kosten- und Nutzteile jeweils alle Fragen von eins an fortlaufend numeriert sind, enthalten zur Unterscheidung die Variablen des Kostenteils den Zusatz *varname\_k*. Neben den Variablen des Fragebogens und den bereits angesprochenen zusätzlichen Variablen aus den Abschnitten 4.3 und 4.4 enthält der Datensatz die folgenden zusätzlichen Merkmale.

**Tabelle 5: Zusätzliche Variablen im FDZ-Datensatz Nutzen und Nettokosten der Berufsausbildung aus dem Jahr 2000**

Variablen-name	Bezeichnung	numerische Vercodung bzw. range	Ausprägungen	Bemerkung
beschg	Beschäftigte inklusive Azubis	2 ; 108.144	kontinuierliches Merkmal	Betriebsgröße
berufart	Berufsart	1 2	gewerblich-technisch kaufmännisch-verwaltend	
westost	West-Ost-Kennzeichen für Standort des Betriebs	1 2	Westdeutschland Ostdeutschland	
bgk	Betriebsgrößenklasse	1 2 3	unter 50 Beschäftigte 50 bis unter 250 Beschäftigte 250 Beschäftigte und mehr	
abereich	Ausbildungsbereich	1 2 3 4 5	Industrie und Handel Handwerk Landwirtschaft Freie Berufe Öffentlicher Dienst	
b_art3	Berufsart	1 2 3	gewerblich technisch kaufmännisch	
zuststell	zuständige Stelle	1 ; 74	aus Platzgründen nicht abgedruckt	enthält die Kammerbezirke aufgeteilt nach Handwerks,-Industrie- und Handels- sowie weiterer Kammern (Ärzte,- Steuerberaterkammer etc.)
bland	Bundesland (Standort des Betriebs)	1 ; 16	aus Platzgründen nicht abgedruckt	
wzw	Rekodierung Wirtschaftszweig	1 ; 17	aus Platzgründen nicht abgedruckt	Rekodierung von f2_1k0 in 17 Wirtschaftszweige



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

## 6. Hinweise zur Datennutzung

Das Kapitel beschreibt Besonderheiten der Daten, welche bei der Datenaufbereitung im BIBB-FDZ identifiziert wurden bzw. auf welche im Methodenbericht von infas (2001) eingegangen wird. Aufgrund der Größe und Komplexität des Datensatzes beschränkt sich die Aufstellung auf zentrale Auffälligkeiten.

In dem aus Einzelbestandteilen zusammengeführten Gesamtdatensatz waren bei zahlreichen Variablen keine nach der Fragebogendeklaration differenzierten Missing Values vorhanden und stattdessen lediglich Systemmissings enthalten. Nach dem nachträglichen Anlegen der Filtermissings wurden die Systemmissings auf -9=keine Angabe umkodiert.

### Besonderheiten der Variablen des Nutzen-Fragebogens

Bei folgenden Variablen sind außer nachträglich angelegten Filtern alle im Fragebogen differenzierten Missing Values (keine Angabe, weiß nicht, verweigert, Gruppe ist nicht beteiligt) im Wert „0“ enthalten:

- **f23\_1 bis f23\_5:** Zeitaufwand beteiligter Mitarbeitergruppen an Bewerbungsgesprächen
- **f27\_1 und f27\_2:** Durchschnittliche Dauer von Weiterbildungsveranstaltungen und Kosten hierfür pro Fachkraft
- **f29\_1 und f29\_2:** Lohn Auf- bzw. Abschläge extern eingestellter Fachkräfte

Somit kann nicht unterschieden werden zwischen Nichtvorkommen und fehlenden Werten. Bei den 22 Fällen, die den Wert -9= keine Angabe aufweisen, handelt es sich um diejenigen Betriebe, welche ausschließlich die Fragen zu den Ausbildungskosten beantwortet haben (vgl. Kap. 3.4).

### Besonderheiten der Variablen des Kosten-Fragebogens

Bei folgenden Variablen sind alle im Fragebogen differenzierten Missing Values (keine Angabe, weiß nicht, verweigert, Gruppe ist nicht beteiligt) im Wert „0“ enthalten:

- **f26\_1k bis f26\_5\_2k:** Gewährte tarifliche oder freiwillige Sozialleistungen
- **f48\_j\_1k und f48\_j\_2k:** Anzahl hauptberuflich und nebenberuflich mit der Ausbildungsverwaltung beschäftigte Mitarbeitergruppen
- **f49k:** Höhe der Sachkosten für die Ausbildungsverwaltung

In diesen Variablen kann somit nicht unterschieden werden zwischen Nichtvorkommen und fehlenden Werten.

Bei den **Fragen 45b** (f45bk) und **45c** (f45c\_jk) trifft dies möglicherweise auf alle Missings außer den nachträglich angelegten Filtern von Frage 44 zu. Dies kann nicht definitiv bestimmt werden, da von dem Filter rund 2.300 Fälle betroffen sind und in der Variable zur ersten Teilfrage von Frage 45 (f45ak), in der differenzierte Missings vorhanden sind, lediglich zwei Fälle mit der Angabe „weiß nicht“ enthalten sind.

Die Zahl der in den **Frage 77** gefilterten Fälle (2256) entspricht nicht der Zahl der Fälle, die in Frage 76 „nein“, „weiß nicht“ oder „verweigert“ angegeben haben (2243 Fälle). Dies ist darauf zurückzuführen, dass bei diesem Filter auch die Fälle mit „keine Angabe“ mit in den Filter aufgenommen wurden. Das waren ursprünglich 13 Fälle. Durch die Umcodierung der 22 Systemmissings ebenfalls zu „keine Angabe“ erhöhte sich die Zahl der „keine Angabe“ Fälle auf 35, so dass keine Übereinstimmung mit der Zahl der gefilterten Fälle in Frage 77 mehr gegeben ist.

In den Variablen zu **Frage 23**, f23\_1k und f23\_2k sind die Tarifbereiche und Tarifbezirke nicht wie im Fragebogen vorgesehen namentlich als Volltext erfasst, sondern bei Angehörigkeit des befragten Betriebes zu einem Tarifbereich bzw. Tarifbezirk mit „1“ (gehört dazu) codiert. Die tatsächlichen Tarifbezirke müssen also von Anwendern über die regionale und Wirtschaftszweigangabe im Bedarfsfall eigenständig rekonstruiert werden.

Die Volltextangaben zu **Frage 26** sind in dem FDZ-Datensatz nicht mehr enthalten, ebenso nicht die Variablen zu den Angaben zu Bau- und Ausstattungsinvestitionen für die Ausbildung im Betrieb insgesamt (Fragen 78 bis 83).

Die Variable zu **Frage 43** (durchschnittliche Wochenarbeitszeit eines Vollzeitbeschäftigten) sowie die Variablen zu **Frage 50** (durchschnittliche Bruttolöhne in der Ausbildung bzw. Ausbildungsverwaltung beschäftigter Mitarbeitergruppen) sind sowohl inklusive fehlender Werte (f43k bzw. f50\_jk) als auch imputiert (f43kosten bzw. f50\_kostenj) in dem FDZ-Datensatz enthalten. Bei den Imputationen dieser Variablen handelt es sich um Mittelwert-Imputationen. Ebenfalls sowohl mit Missings als auch imputiert enthalten sind die Variablen zu **Frage 21** (durchschnittliche Brutto-Ausbildungsvergütung pro Auszubildendem). Die im BIBB-FDZ vorgenommene Imputation dieser Variablen ist in Kapitel 4.3 dargestellt. Die imputierten Variablen zu Frage 21 sind an der Endung „\_i“ des Variablennamens zu erkennen.

## **7. FDZ-Angebot zum Datensatz Nutzen und Nettokosten der Berufsausbildung aus dem Jahr 2000**

Als Arbeitshilfen werden über das BIBB-FDZ die Fragebögen (pdf-Files) sowie Spieldaten zu 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 des hier dokumentierten Kosten-Nutzen-Datensatzes aus dem Jahr 2000 aufgelistet (7.3).

## 7.1 Arbeitshilfen

- **Fragebögen**

Über die Internetseiten des BIBB-FDZ ([www.bibb-fdz.de](http://www.bibb-fdz.de)) können die Fragebögen der Kosten-Nutzen-Erhebung 2001 abgerufen werden. Beim Kosten-Fragebogen handelt es sich um einen face-to-face eingesetzten Fragebogen, der Nutzen-Fragebogen ist als schriftlicher Selbstausfüller angelegt.

- **Spieldaten**

Auf den Internetseiten des BIBB-FDZ können für Stata und SPSS formal anonymisierte Spieldatensätze der Kosten-Nutzen Befragung 2000 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 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 Zusatzvariablen der Kosten-Nutzen-Erhebung für das Jahr 2000 (zuststel sowie x1 bis x6) sind - in inhaltlich verfälschter Form - in den Spieldaten vorhanden.

## 7.2 Datenzugangswegen

Die Daten der Befragung „Kosten und Nutzen der betrieblichen Berufsausbildung 2000“ sind – wie alle Betriebsdaten – grundsätzlich nur über Datenfernverarbeitung und im Gastwissenschaftlermodell verfügbar.

Sofern Anwender Zusatzvariablen auswerten möchten, bitten wir dies in dem dafür vorgesehenen Nutzungsantrag zu begründen. Im Anhang A2 sind diese Variablen an der kursiven Schrift erkennbar.

### 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 Literaturdatenbank 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.

*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. Berichte zur Beruflichen Bildung, Heft 264/2004

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

*Dionisius, R., S. Muehlemann, et al. (2009):* Costs and benefits of apprenticeship training. A comparison of Germany and Switzerland. Applied economics quarterly **55**(1): 7-37.

*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

Schweri, J. (2008) Kosten und Nutzen der betrieblichen Berufsbildung - Schweizer Resultate im Vergleich mit Deutschland. In: Piening, D. /Rauner, F. (Hrsg.): Innovative Berufsbildung: auf die Attraktivität für Jugendliche und Unternehmen kommt es an! 72 - 89

*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

*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.

## 8 Ausblick

Das vorliegende Daten- und Methodenhandbuch stellt die Arbeitsergebnisse des BIBB-FDZ bei der Datenrekonstruktion der Kosten-Nutzen-Erhebung zusammen. Es ist nicht auszuschließen, dass die weitere Arbeit mit den Daten noch die eine oder andere Besonderheit des FDZ-Datensatzes offen legt. Solche neuen Informationen werden im Internet zeitnah eingestellt und das vorliegende Handbuch in regelmäßigen Abständen entsprechend aktualisiert. Die Datendokumentation selber ist relativ ausführlich, um bei neueren Datenangeboten der Kosten- und Nutzenerhebungen lediglich gezielt auf Unterschiede gegenüber den vorliegenden Ausführungen eingehen zu können<sup>14</sup>.

Im Jahr 2010 wird die Kosten-Nutzen-Erhebung aus dem Jahr 2007 über das BIBB-FDZ der nicht-kommerziellen Forschung zur Verfügung stehen. Die vorliegende Rekonstruktion der 2000er-Daten wird dann auch Zeitvergleiche und damit Analysen zu möglichen Veränderungen der Kosten-Nutzen-Relationen bei der betrieblichen Berufsausbildung in Deutschland ermöglichen. Abhängig von etwaigen internationalen Auswertungsinteressen ist auch die Erstellung einer (kürzeren) englischsprachigen Datendokumentation denkbar.

---

<sup>14</sup> Insbesondere die Ausführungen zu Zielen und Zweck der Untersuchung sowie zum Erhebungsdesign dürften zukünftig weitaus weniger ausführlich zu dokumentieren sein.

**Literatur:**

Bardeleben, R. von/ Beicht, U./ Feher, 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./ 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**

Bundesinstitut für Berufsbildung (2002): Abschlussbericht zum Forschungsprojekt 2.9005. Nutzen und Nettokosten der Berufsausbildung für Betriebe, Bonn

infas (2001): Nutzen und Nettokosten der Berufsausbildung. Methodenbericht zu einer repräsentativen Betriebsbefragung. internes Dokument, Bonn

Sachverständigenkommission „Kosten und Finanzierung der beruflichen Bildung (1974): Kosten und Finanzierung der außerschulischen beruflichen Bildung, Bielefeld

## Anhang

### A1: Übersicht der berücksichtigten Ausbildungsberufe in der Kosten-Nutzen-Erhebung aus dem Jahr 2001

#### Industrie und Handel

- Bankkaufmann/frau
- Bauzeichner/-in
- Bürokaufmann/frau
- Chemielaborant/-in
- Chemikant /-in
- Drucker/-in
- Energieelektroniker/-in - Betriebstechnik
- Fachinformatiker/-in
- Fachkraft für Lagerwirtschaft
- Florist /-in
- Hotelfachmann/frau
- Industriekaufmann/frau
- Industriemechaniker/-in – Betriebstechnik
- Industriemechaniker/-in – Maschinen- u. Systemtechnik
- 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
- Reiseverkehrskaufmann/frau
- Versicherungskaufmann/frau
- Werkzeugmechaniker/-in – Stanz- und Umformtechnik

#### Handwerk

- Augenoptiker
- Bäcker/-in
- Bürokaufmann/frau
- Elektroinstallateur/-in
- Fachverkäufer/-in im Nahrungsmittelhandwerk
- Fleischer/-in
- Friseur/-in
- Gas- und Wasserinstallateur/-in
- Informationselektroniker/-in
- Kraftfahrzeugmechaniker/-in
- Maler/-in und Lackierer/-in
- Maurer/-in
- Metallbauer/-in
- Tischler/-in
- Zahntechniker/-in
- Zentralheizungs- u. Lüftungsbauer/-in

#### Landwirtschaft

- Gärtner/-in
- Landwirt/-in

#### Freie Berufe

- Arzthelfer/-in
- Rechtsanwaltsfachangestellte/-r
- Steuerfachangestellte/-r
- Zahnarzthelfer/-in

#### Öffentlicher Dienst

- Fachangestellte/-r für Bürokommunikation
- Sozialversicherungsfachangestellte/-r
- Vermessungstechniker/-in
- Verwaltungsfachangestellte/-r

**A2: Zusatzvariablen der Kosten-Nutzen-Erhebung aus dem Jahr 2001**

<b>Variablen- name</b>	<b>Bedeutung</b>
ld	Identifikationsnummer
bkz	Berufskennziffer
s_ausb	Stichprobe: Ausbildungsbereich numerisch
<i>zuststel</i>	<i>zuständige Stelle</i>
bland	Bundesland
westost	Standort West-/Ostdeutschland
wzw	17 Wirtschaftszweige aus Frage 1 Kostendatensatz
bgroesse	4 Betriebsgrößenklassen
b_gew	Betriebsgewichtungsfaktor
b_hoch	Betriebshochrechnungsfaktor
bgk	3 Betriebsgrößenklassen
abereich	Ausbildungsbereich
beruf	Ausbildungsberuf
berufart	Berufsart
lz_3	3-jährige Ausbildungsberufe ja/nein
b_art3	Berufsart gewerblich/ technisch/ kaufmännisch
x1	<i>Durchschnitts-Bruttokosten (voll) in €</i>
x2	<i>Durchschnitts-Äquivalenzerträge in € -</i>
x3	<i>Durchschnitts-Nettokosten (voll) in € -</i>
x4	<i>Durchschnitts-Bruttokosten (teil) in € -</i>
x5	<i>Durchschnitts-Nettokosten (teil) in € -</i>
X6	<i>Durchschnitts-Substitutionserträge in € -</i>
ausbstd_azubi	Ausbilderstunden (gesamt) pro Auszubildenden in einer Standardwoche

*Anmerkung:* kursiv gedruckt sind Zusatzvariablen, die nur mit einer gesonderten Begründung für Auswertungen zur Verfügung stehen.



## Impressum

---

BIBB-FDZ Daten- und Methodenberichte  
Nr.4/2009  
Autoren: Holger Alda, Stefan Koscheck

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

Redaktion: Holger Alda  
Redaktionsassistentz: Tanja Stierner

Downloads unter:  
[www.bibb-fdz.de](http://www.bibb-fdz.de)

Tel.: +49-228-107-2041  
Fax: +49-228-107-2020  
E-Mail: [fdz@bibb.de](mailto:fdz@bibb.de)

ISSN-Nr.: 2190-300X  
urn:nbn:de:0035-0792-4

---

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](http://www.bibb.de/cc-lizenz).