

Scientific Use File und Analysen auf Basis der Europäischen Erhebung zur betrieblichen Weiterbildung (CVTS2)

Maurice Brandt / Dr. Rainer Lenz* / Dr. Hans-Peter Hafner** / Daniel Schmidt****

** Statistisches Bundesamt, Forschungsdatenzentrum, Wiesbaden*

*** Hessisches Statistisches Landesamt, Forschungsdatenzentrum, Wiesbaden*

**** Statistisches Bundesamt, Gruppe „Bildung, Forschung und Entwicklung, Kultur, Rechtspflege“, Wiesbaden*

1 Einleitung

Mit dem demografischen Wandel und der zunehmenden Alterung der Bevölkerung ist eine Veränderung der bisherigen Rentenpraxis in der Bundesrepublik Deutschland unumgänglich. Um eine weitere Belastung Jüngerer und Rentenkürzungen zu vermeiden, wird eine Erhöhung des Renteneintrittsalters und somit eine Verlängerung der Lebensarbeitszeit notwendig. Durch die Forderung nach einem längeren Verbleib im Erwerbsleben werden die Ansprüche an ältere Arbeitnehmer steigen. Um den Anforderungen dieses immer dynamischeren Arbeitsmarktes gerecht zu werden, wird lebenslanges Lernen unerlässlich, und die berufliche Weiterbildung gewinnt für den Verbleib im Erwerbsleben an Bedeutung. Mit der Transformation zur sogenannten Wissensgesellschaft wird in Zukunft ein großer Bedarf an gut ausgebildeten und hoch qualifizierten Personen bestehen. In diesem Zusammenhang wird nicht nur für die Arbeitnehmer und Unternehmen das Thema der beruflichen Weiterbildung wichtiger, sondern auch in der Politik und Wissenschaft wächst die Nachfrage nach Daten zur beruflichen Weiterbildung.

Mit dem als Scientific Use File bezeichneten Datensatz der Zweiten Europäischen Erhebung zur beruflichen Weiterbildung (CVTS2, „Second Continuing Vocational Training Survey“), der Angaben von etwa 3.200 deutschen Unternehmen enthält, wurde dem jüngst geäußerten Wunsch der Wissenschaft nach Daten zur beruflichen Weiterbildung zeitnah Folge geleistet. Seit Ende 2005 können Wissenschaftler Daten der amtlichen Statistik zur beruflichen Weiterbildung in Unternehmen für eigene Analysen nutzen. In einem gemeinsamen Projekt haben die Statistischen Ämter des Bundes und der Länder¹ Einzeldaten der Zweiten Europäischen Erhebung zur beruflichen Weiterbildung aus dem Jahre 2000 mit Berichtsjahr 1999 so anonymisiert, dass sie einerseits den strengen Anforderungen des gesetzlichen Datenschutzes genügen und andererseits ausreichend Potenzial für wissenschaftliche Analysen bieten.

Nachfolgend soll die Datenbereitstellung für die deutsche Erhebung zur beruflichen Weiterbildung in Unternehmen in Form eines Scientific Use Files beschrieben und ein Überblick darüber gegeben werden, wie die deutschen Unternehmen mit ihrem Weiterbildungsangebot aufgestellt sind.

¹ Beim dritten Workshop des Forschungsdatenzentrums der Rentenversicherung hat Frau *Helga Christians* vom Forschungsdatenzentrum der Statistischen Landesämter einen Vortrag zu den Aufgaben und Zielen, dem Datenangebot und den Zugangswegen zu den Forschungszentren der Statistischen Ämter des Bundes und der Länder gehalten.

Die Erstellung von Scientific Use Files, wie z. B. zu den Daten des Mikrozensus, der Einkommens- und Verbrauchsstichprobe und zur Zeitbudgeterhebung, wird vom Bundesministerium für Bildung und Forschung im Rahmen seiner Förderung der informationellen Infrastruktur zwischen Wissenschaft und Statistik in Deutschland unterstützt.

Eine erste Anwendung findet der Scientific Use File zur CVTS2 in dem vom Rat für Sozial- und Wirtschaftsdaten ausgerufenen Expertisen-Wettbewerb zum Thema „Bildung im Erwerbsleben“.²

2 Vorbemerkung

Im Jahr 1987 wurde mit § 16 Abs. 6 des Bundesstatistikgesetzes³ der Wissenschaft ein privilegierter Zugang zu Mikrodaten der amtlichen Statistik eingeräumt. Hiernach ist die Übermittlung von Einzeldaten an die Wissenschaft erlaubt, sofern diese nur mit unverhältnismäßig hohem Aufwand an Zeit, Kosten und Arbeitskraft reidentifiziert werden können (faktische Anonymität). „Unverhältnismäßig“ bedeutet hier, dass der Aufwand einer Reidentifikation deren Nutzen übersteigt. Dies impliziert, dass eine Deanonymisierung von Einzelangaben in einem faktisch anonymen Datensatz nicht mit absoluter Sicherheit ausgeschlossen werden muss, es für einen potenziellen Datenangreifer aber unattraktiv wäre, eine Deanonymisierung zu versuchen. Im Folgenden wird eine faktisch anonymisierte Datei für die Wissenschaft (ein so genannter Scientific Use File), generiert aus den Daten der Zweiten Europäischen Erhebung zur beruflichen Weiterbildung in Unternehmen (CVTS2) aus dem Jahr 2000 mit Berichtsjahr 1999, vorgestellt. Dieses Produkt ist in einem Kooperationsprojekt zwischen dem Hessischen Statistischen Landesamt und dem Statistischen Bundesamt entstanden.

3 Basismaterial und Analysen

3.1 Europäische Erhebung

In der Zweiten Europäischen Erhebung zur betrieblichen Weiterbildung wurden insgesamt 76.000 Unternehmen in 26 europäischen Ländern zu ihren Weiterbildungsmaßnahmen im Jahr 1999 befragt. Ziel war es, vergleichbare Informationen ausschließlich über die betriebliche Weiterbildung zu erhalten. Die betriebliche Weiterbildung hebt sich von der beruflichen Erstausbildung ab und unterscheidet sich von anderen beruflichen Weiterbildungen dadurch, dass sie ganz oder teilweise von den Unternehmen finanziert wird (**siehe Schaubild 1**).

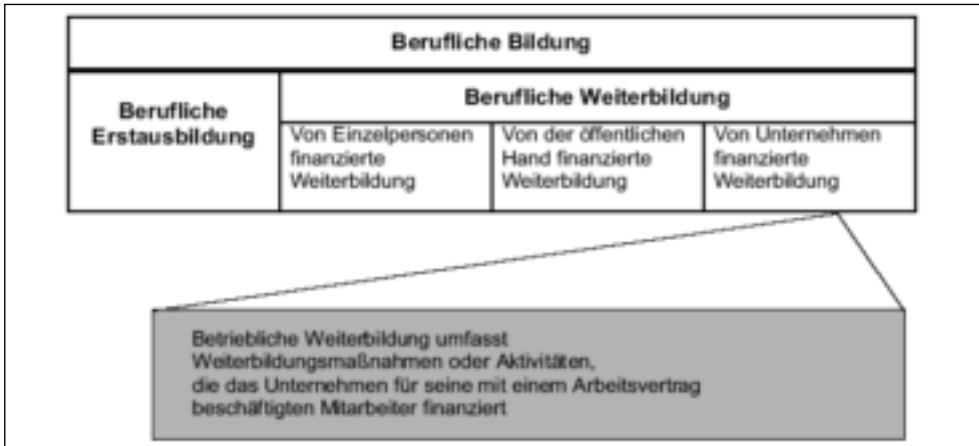
3.2 Deutschland im europäischen Vergleich

Ein wichtiger Indikator, um die unternommenen Anstrengungen in der betrieblichen Weiterbildung darzustellen, ist der Anteil von Kursstunden an den geleisteten Gesamtarbeitsstunden aller Unternehmen.

² Weitere Informationen hierzu finden sich im Internet unter <http://www.ratswd.de/wettbew.htm>.

³ Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3322).

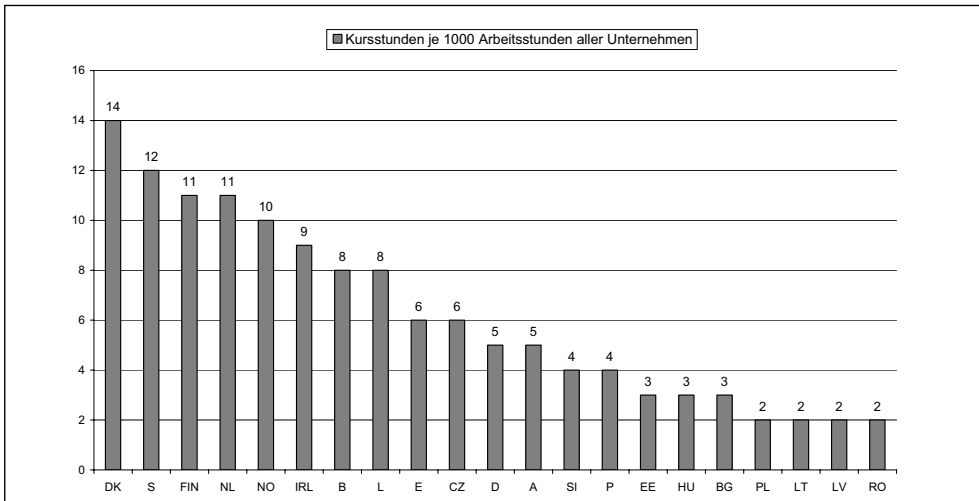
Schaubild 1:



Quelle: Eurostat (2002): Europäische Sozialstatistik, S. 9.

Mit 5 Kursstunden je 1.000 Arbeitsstunden liegt Deutschland bei diesem Vergleich im Mittelfeld der befragten europäischen Länder, während die skandinavischen Länder den höchsten Anteil von Kursstunden je 1.000 Arbeitstunden aufweisen (siehe Schaubild 2).

Schaubild 2:

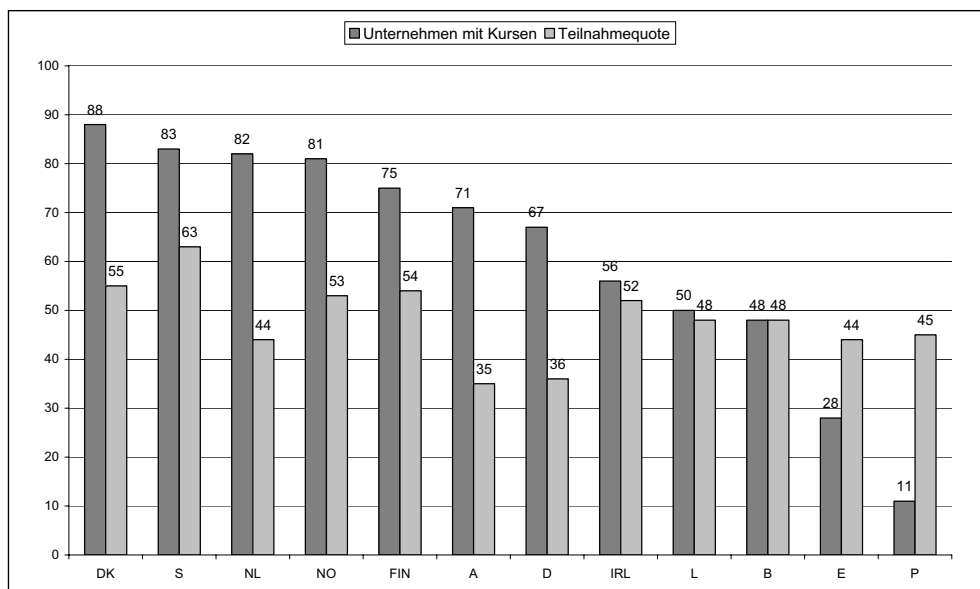


Quelle: Statistik kurz gefasst (1/2003), S. 1.

Abkürzungen: B: Belgien, DK: Dänemark, D: Deutschland, E: Spanien, IRL: Irland, L: Luxemburg, NL: Niederlande, A: Österreich, P: Portugal, FIN: Finnland, S: Schweden, NO: Norwegen, BG: Bulgarien, CZ: Tschechische Republik, EE: Estland, HU: Ungarn, LV: Lettland, LT: Litauen, PL: Polen, RO: Rumänien, SI: Slowenien.

Wird der Anteil der Unternehmen betrachtet, die Kurse anbieten, so liegt Deutschland in dieser Kategorie ebenfalls im europäischen Mittelfeld (**siehe Schaubild 3**). Bei der Teilnahme der Beschäftigten an den angebotenen Kursen hat Deutschland, neben Österreich, mit 36% eine der geringsten Teilnahmequoten. Verglichen mit Dänemark, Schweden oder Norwegen ist der Anteil der deutschen Unternehmen, die Weiterbildungsmaßnahmen anbieten, geringer. Die Teilnahmequoten der skandinavischen Länder sind, verglichen mit Deutschland, ebenfalls höher und in Schweden mit 63% am höchsten. In Portugal ist der Anteil an Unternehmen mit Kursangeboten mit 11% eher gering, die Teilnahmequote beträgt jedoch 45%. Das heißt, wenn por-

Schaubild 3:



Quelle: Statistik kurz gefasst (3/2002), S. 3.

tugiesische Unternehmen betriebliche Weiterbildungskurse anbieten, dann nimmt fast die Hälfte der Beschäftigten daran teil.

3.3 Deutsche Daten der Europäischen Erhebung

In der Erhebung liegen Angaben von 3.184 deutschen Unternehmen mit mehr als 10 Beschäftigten aus den Wirtschaftszweigen der Abschnitte C-K und O der NACE Rev. 1 zur Teilnahme von Beschäftigten an Maßnahmen zur beruflichen Weiterbildung im Jahr 1999 vor.

Die Daten enthalten Informationen zum Angebot an verschiedenen Formen beruflicher Weiterbildung, zu Teilnehmern an Lehrveranstaltungen, Teilnahmezeiten in Lehrveranstaltungen und Kosten für Lehrveranstaltungen sowie qualitative Angaben zur Weiterbildungskonzeption und zum Stellenwert der Weiterbildung im Unternehmen. Bei der Anonymisierung ist es unter an-

derem gelungen, eine wissenschaftliche Behandlung relevanter Fragestellungen nach Wirtschaftsbereichen und Beschäftigtengrößenklassen zu ermöglichen.⁴

In 67% der befragten Unternehmen nahmen Beschäftigte im Jahr 1999 an Lehrveranstaltungen zur Weiterbildung teil. Im Kredit- und Versicherungsgewerbe sowie in der Datenverarbeitung bot jedes der befragten Unternehmen solche Veranstaltungen an, hingegen waren es im Textil- und Bekleidungs-gewerbe nur 42%, im Bergbau 46% und im Verkehr 50% (**siehe Schaubild 4**).

36,2% der Beschäftigten der befragten Unternehmen nahmen an Lehrveranstaltungen teil (bezogen auf die Unternehmen, die solche Veranstaltungen anbieten). Die Teilnahmequote der Männer lag mit 38,2% deutlich über derjenigen der Frauen (32,7%). Aber auch hier gab es sehr starke Differenzen zwischen den Branchen (**siehe Schaubild 5**). Im Bereich Kultur, Sport, Unterhaltung lag die Teilnahmequote insgesamt nur bei 13,5% (da hier aber nur wenige Unternehmen geantwortet haben, ist dieser Wert nicht unbedingt repräsentativ), in der Nachrichtenübermittlung bei 78,6%. In der Metallindustrie waren die Teilnahmequoten von Frauen und Männern fast umgekehrt wie bei allen Unternehmen (Frauen: 37,5%, Männer: 34,6%) und auch in einigen anderen Branchen wie z. B. im Baugewerbe und im Bereich Rechts-, Steuer-, Unternehmensberatung und Markt- und Meinungsforschung lag die Teilnahmequote der Frauen über derjenigen der Männer, während bei Zeitarbeit, Detekteien, Schutzdiensten, Reinigung und sonstigen Dienstleistungen für Unternehmen der Anteil der Männer, die an Lehrveranstaltungen teilnahmen, mehr als doppelt so hoch war wie derjenige der Frauen (28,9% zu 12,8%).

Beim Anteil der Teilnahmestunden an Lehrveranstaltungen nach Themen lag die EDV mit 20,7% klar vorne, gefolgt von Technik und Produktion mit 11,1%, Verkaufstraining mit 11,0% und Management, Verwaltung u. Ä. mit 10,4%.

Umweltschutz, Gesundheit und Sicherheit hatten insgesamt nur einen geringen Anteil von 3,8%; im Wirtschaftsbereich Bergbau und Gewinnung von Steinen und Erden nahmen diese Themen jedoch 18,4% der Weiterbildungszeit ein. Wenig überraschend ist, dass der Anteil an EDV-Weiterbildung im Bereich Datenverarbeitung mit 65,9% am höchsten lag; und auch dass 42,2% der Zeit für Weiterbildungskurse im Versicherungsgewerbe auf Verkaufstraining entfallen, ist nicht unbedingt unerwartet.

Nach den klassischen Lehrveranstaltungen waren Informationsveranstaltungen (in 61,0% der Unternehmen) und Unterweisung durch Vorgesetzte oder Lernen durch normale Arbeitsmittel (53,7%) weitverbreitete Methoden der beruflichen Weiterbildung. Hingegen spielten sowohl selbstgesteuertes Lernen (13,9%), Lern- und Qualitätszirkel (11,5%) als auch Austauschprogramme und Jobrotation (4,2%) nur eine geringe Rolle (**Schaubild 7**).

⁴ Weitere Informationen zu dem Basismaterial sowie eine vollständige Merkmalsliste finden sich in Egner, U.: Berufliche Weiterbildung in Unternehmen (CVTS2), Projektbericht 2002.

Schaubild 4:

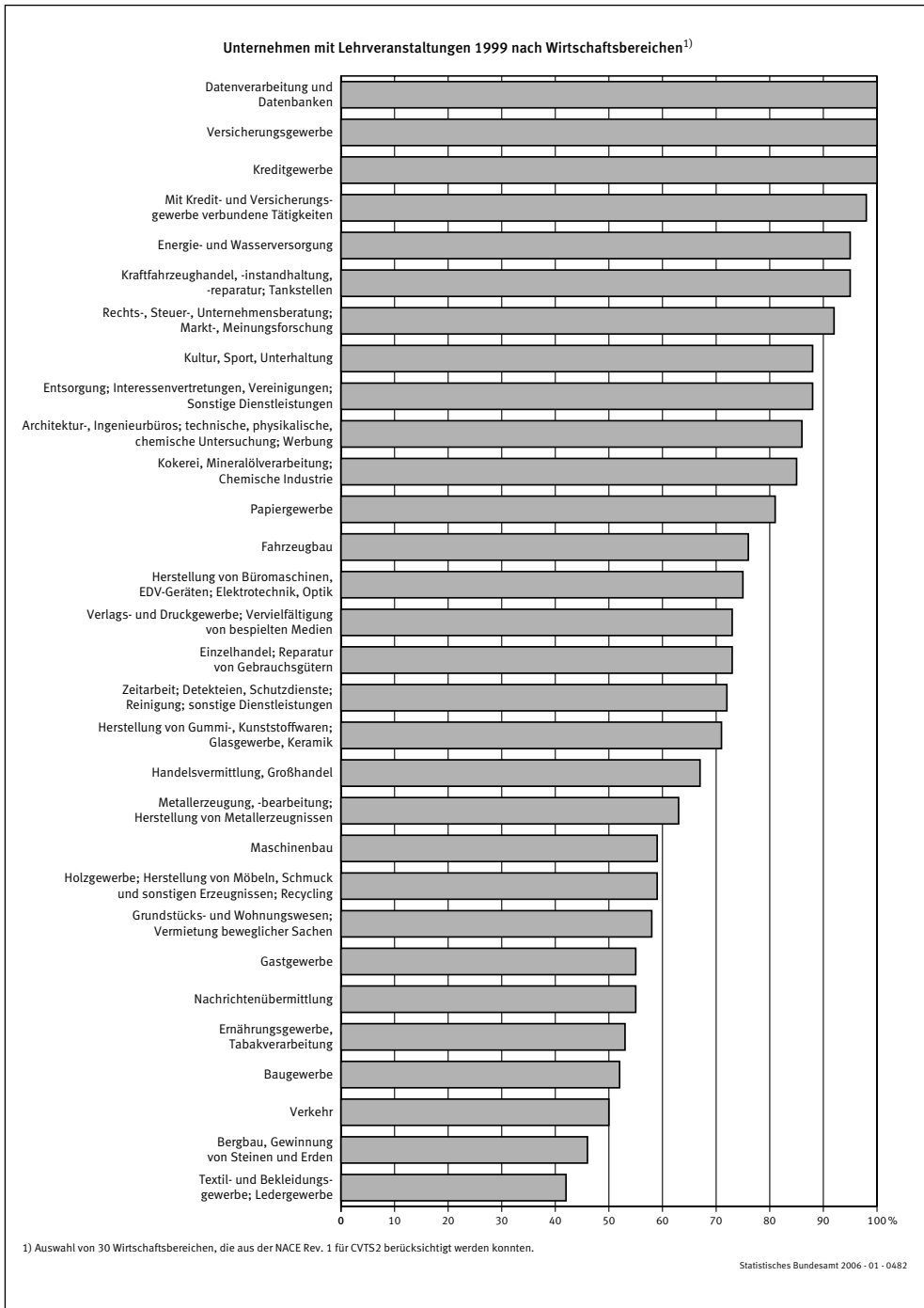


Schaubild 5:

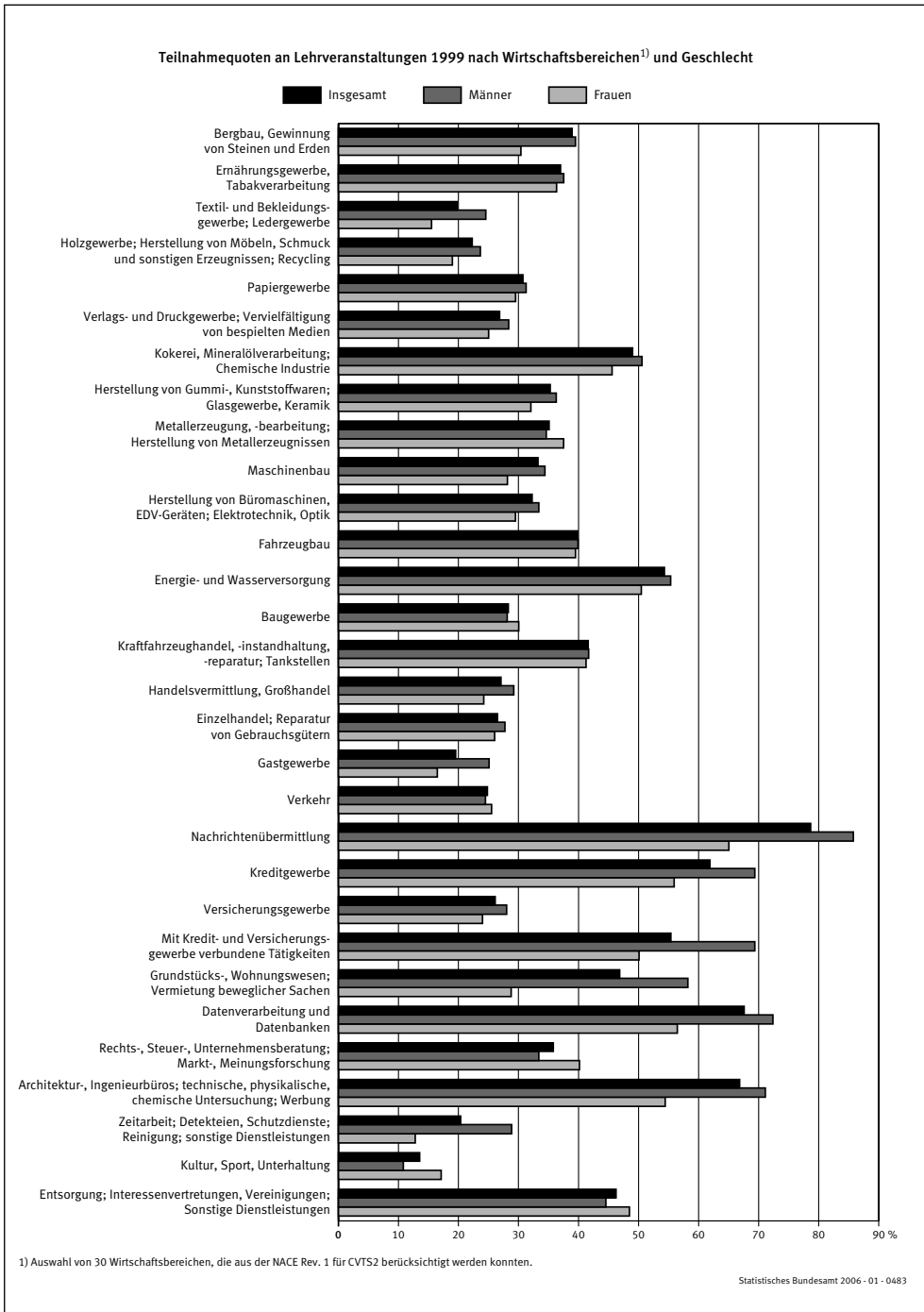


Schaubild 6:

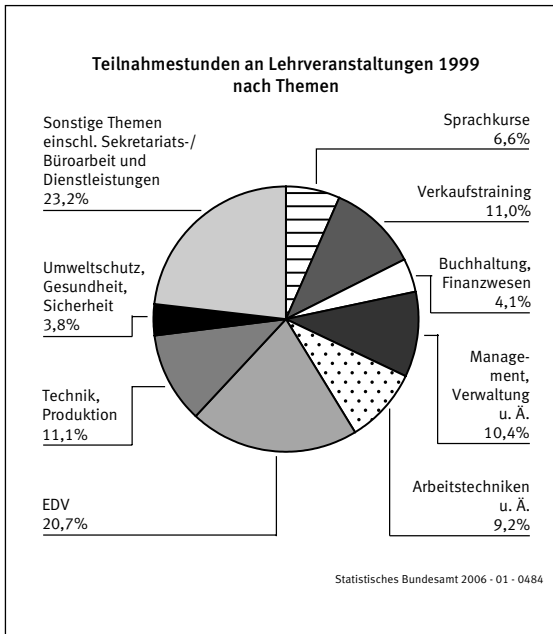
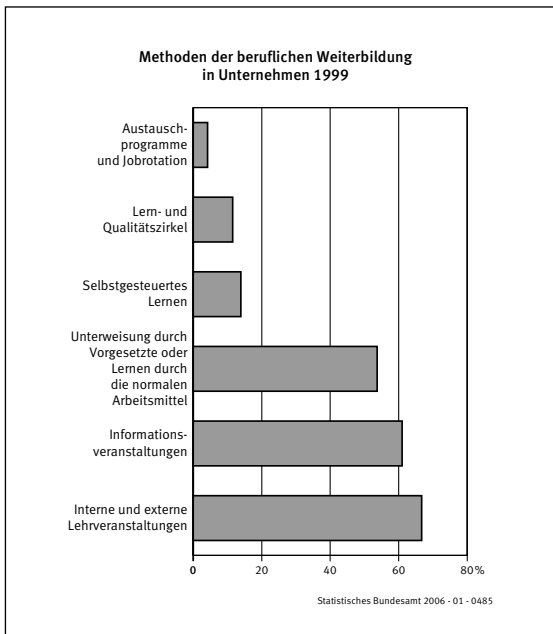


Schaubild 7:

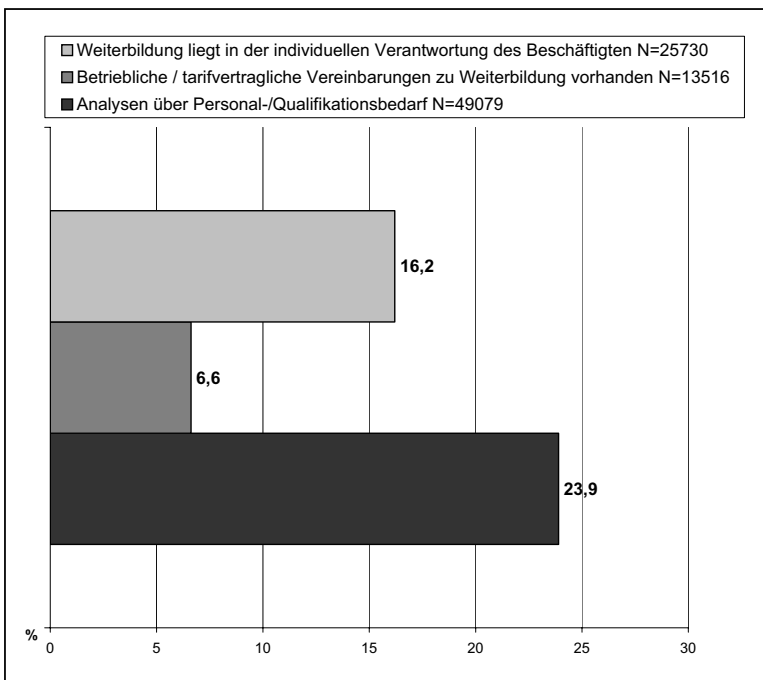


3.4 Potenzial der betrieblichen Weiterbildung in Deutschland

Wird die Bereitschaft zur betrieblichen Weiterbildung in Deutschland genauer betrachtet, so wird ersichtlich, dass etwa ein Viertel (ca. 24%) der deutschen Unternehmen Analysen über ihren Personal- und Qualifikationsbedarf durchführen (**siehe Schaubild 8**). Verbindliche Zusagen zur betrieblichen Weiterbildung in Form von betrieblichen und tarifvertraglichen Vereinbarungen sind nur bei etwa 7% der deutschen Unternehmen zu finden. Dem entgegen steht die Aussage, dass die Weiterbildung eher nicht in der individuellen Verantwortung des Beschäftigten gesehen wird, denn nur etwa 16% der befragten Unternehmen ohne Weiterbildungsplan gaben an, dass die Weiterbildung allein Sache der Beschäftigten sei. Einerseits erkennen die Unternehmen, dass die Weiterbildung nicht nur dem persönlichen Bereich der Mitarbeiter überlassen werden kann, andererseits werden kaum verbindliche Weiterbildungskurse angeboten.

Die eingangs erwähnte Gruppe der älteren Arbeitnehmer, für die Weiterbildung einen wichtigen Faktor für den Verbleib im Erwerbsleben darstellt, wird im Folgenden Gegenstand der Betrachtung sein. Dabei sollen die Beschäftigten ab 50 Jahre und Beschäftigte, deren Arbeitsplatz gefährdet ist, genauer untersucht werden. Die Beschäftigten bis 25 Jahre werden als Referenzgruppe für den Vergleich hinzugezogen.

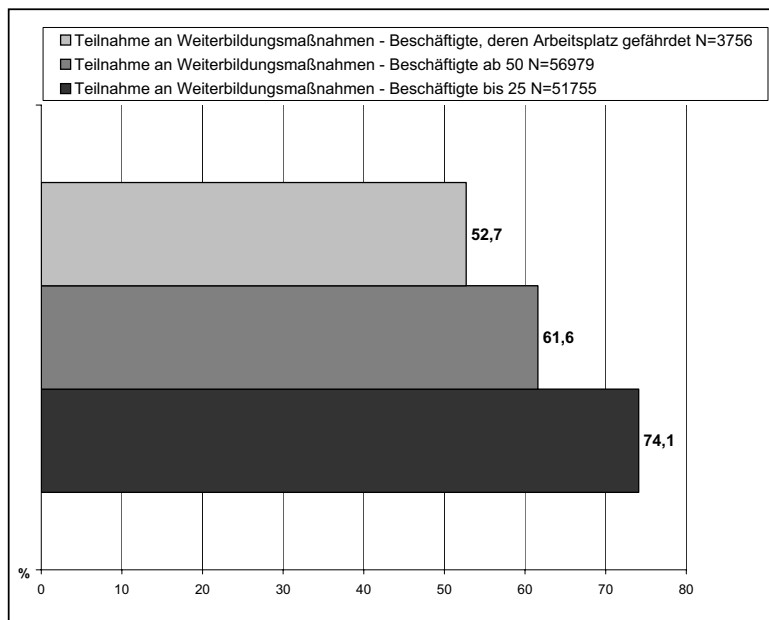
Schaubild 8:



Quelle: Statistisches Bundesamt, CVTS2, eigene Berechnung.

Aus **Schaubild 9** ist ersichtlich, dass sich die Teilnahme an Weiterbildungsmaßnahmen zwischen jüngeren und älteren Arbeitnehmern voneinander unterscheidet. Wenn Unternehmen Weiterbildungen anbieten, dann nehmen jüngere Beschäftigte mit 74% häufiger an Weiterbildungsveranstaltungen teil als Beschäftigte ab 50 Jahre. Arbeitnehmer, deren Arbeitsplatz gefährdet ist, besuchen mit 53% weitaus seltener Weiterbildungskurse. Ein speziell auf diese Ziel-

Schaubild 9:

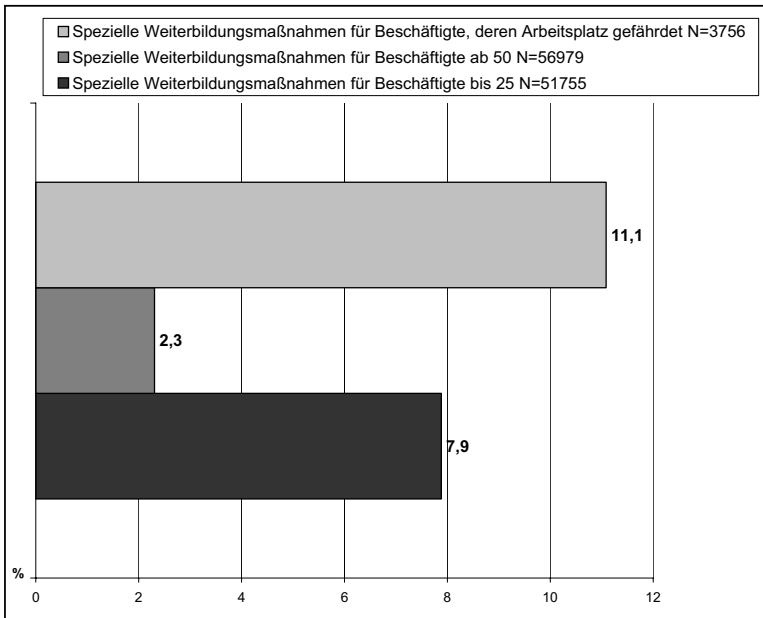


Quelle: Statistisches Bundesamt, CVTS2, eigene Berechnung.

gruppen ausgerichtetes Weiterbildungsangebot existiert in den meisten Unternehmen nicht. Von den deutschen Unternehmen, die überhaupt Weiterbildungsangebote bereitstellen, bieten nur etwa 2% spezielle Weiterbildungsmaßnahmen für Beschäftigte ab 50 Jahren an. Für jüngere Arbeitnehmer bis 25 Jahre werden in ca. 8% der Unternehmen und für Beschäftigte, deren Arbeitsplatz gefährdet ist, werden immerhin in ca. 11% der Unternehmen, die generell weiterbilden, spezielle Weiterbildungsmaßnahmen durchgeführt (**siehe Schaubild 10**).

Zusammenfassend ist festzustellen, dass Deutschland sich beim Angebot der betrieblichen Weiterbildung allenfalls im Mittelfeld befindet und bezüglich der Teilnahmequote mit Österreich das Schlusslicht der befragten europäischen Länder bildet. Es ist noch viel Potenzial in Sachen betrieblicher Weiterbildung vorhanden, das gilt zum einen für die Analysen des Qualifikationsbedarfs der Unternehmen und zum anderen für das Weiterbildungsangebot für bestimmte Gruppen von Mitarbeitern, insbesondere für ältere Arbeitnehmer. Auf die Bedürfnisse von älteren Beschäftigten wird in Form von speziellen Weiterbildungsmaßnahmen kaum eingegangen und es wird wenig unternommen, um die Qualifikation von älteren Mitarbeitern und somit den längeren Verbleib im Erwerbsleben zu fördern. Positiv hervorzuheben ist, dass die Teilnahme-

Schaubild 10:



Quelle: Statistisches Bundesamt, CVTS2, eigene Berechnung

quoten bei den Unternehmen, die die jeweilige Zielgruppe von Mitarbeitern beschäftigen und Weiterbildungsmaßnahmen anbieten, recht hoch sind.

4 Analyse des Gefährdungspotenzials

Obleich durch die Tatsache, dass es sich bei CVTS2 um eine Stichprobenerhebung handelt, bereits eine Schutzwirkung erreicht wird, sind bei der Erstellung eines Scientific Use Files, der Wissenschaftlern den Datenzugang außerhalb der geschützten Räume der amtlichen Statistik ermöglicht, weitere Schutzmaßnahmen erforderlich. Dies ist darüber hinaus wichtig, um das Vertrauen der (teilnehmenden) Unternehmen in die amtliche Statistik zu erhalten und sie zu einer Teilnahme an zukünftigen Erhebungen motivieren zu können.

Zusatzwissen über Weiterbildungsmaßnahmen von Unternehmen liegt einem potenziellen Datenangreifer nicht in systematischer Form vor, sodass für die diesbezüglichen Merkmale höchstens Einzelangriffe, d. h. Individualrecherchen über einzelne Unternehmen, denkbar wären. Dazu müsste man aber über ein Unternehmen schon sehr viel wissen und würde dann kaum mehr zusätzliche Informationen gewinnen. Das einzig wirklich kritische Überschneidungsmerkmal zu kommerziellen Datenbanken ist die Anzahl der Beschäftigten. Da es aber bei diesem Merkmal erhebliche Differenzen zwischen verschiedenen Quellen gibt und in den unteren und mittleren Beschäftigtengrößenklassen sehr viele Unternehmen enthalten sind, deren Beschäftigtenzahlen relativ eng beieinander liegen, sind in erster Linie die ganz großen Unternehmen gefährdet.

5 Anonymisierungsmaßnahmen

Bei den nachfolgend beschriebenen Maßnahmen wurde ein besonderes Gewicht auf die Vergrößerung kategorialer Merkmale gelegt, die ein potenzieller Datenangreifer als Überschneidungsmerkmale mit den vertraulichen Daten der CVTS2 verwenden könnte.

Regionale Gliederung

Es wird keine Regionalinformation weitergegeben. Regionale Angaben sind besonders geeignet für Reidentifikationen. Der Erhalt solcher Merkmale in einem Scientific Use File stellt daher für die Anonymisierung ein schwieriges Unterfangen dar. Hinzu kommt, dass aufgrund der geringen Fallzahl in den neuen Bundesländern (585 Unternehmen) für die meisten Fragestellungen keine belastbaren Ergebnisse getrennt nach alten und neuen Bundesländern zu erzielen wären, sodass der Wegfall des Regionalmerkmals keine wesentliche Einschränkung des Analysepotenzials bedeutet.

Wirtschaftszweigklassifikation

Ausgangspunkt sind die 30 Wirtschaftsbereiche der Klassifikation NACE 30, nach denen die Stichprobenauswahl für die Erhebung erfolgte.

Auf Basis der Hochrechnungsfaktoren für die Grundgesamtheit haben sich einige Wirtschaftsbereiche als besonders gefährdet herausgestellt. Diese mussten daher mit anderen Wirtschaftsbereichen wie folgt zusammengelegt werden:

1. Mit dem Kredit- und Versicherungsgewerbe verbundene Tätigkeiten und Versicherungsgewerbe,
2. Bergbau und Gewinnung von Steinen und Erden sowie Kokerei, Mineralölverarbeitung und Chemische Industrie,
3. Nachrichtenübermittlung und Verkehr,
4. Kraftfahrzeughandel, -instandhaltung und -reparatur sowie Einzelhandel und der Reparatur von Gebrauchsgütern,
5. Holzgewerbe, Herstellung von Möbeln, Schmuck, Musikinstrumenten, Spielwaren, Sportgeräten und sonstigen Erzeugnissen und Papiergewerbe.

Für die zusammengelegten Wirtschaftsbereiche wurden die Hochrechnungsfaktoren neu berechnet. Sie ergeben sich als Quotient aus der Anzahl der Unternehmen des zusammengefassten Bereichs in der Grundgesamtheit und der Anzahl der Unternehmen des zusammengefassten Bereichs in der Erhebung.

Im Dienstleistungsbereich war die Anzahl der Unternehmen in der Stichprobe durchgängig sehr gering. Hier bestand daher kein Reidentifikationsrisiko, sodass die Entscheidung über eine für Analysen sicher erforderliche Zusammenfassung den Wissenschaftlern überlassen werden kann.

Eine Übersicht der in der anonymisierten Datei vorhandenen Wirtschaftsbereiche (im Wesentlichen die sogenannte europäische NACE 30-Klassifikation) samt Fallzahlen ist in nachfolgender **Tabelle 1** gegeben:

Tabelle 1: Verteilung der an der Zweiten Europäischen Erhebung zur beruflichen Weiterbildung (CVTS2) beteiligten Unternehmen auf Wirtschaftsbereiche

Wirtschaftsgliederung (H. v. = Herstellung von)	NACE 30- Angabe ¹⁾	Anzahl	Grund- gesamtheit
Ernährungsgewerbe und Tabakverarbeitung	02	184	16.886
Textil- und Bekleidungsgewerbe; Ledergewerbe	03	193	3.549
Verlags- und Druckgewerbe; Vervielfältigung von bespielten Ton-, Bild- und Datenträgern	06	125	5.737
H. v. Gummi- und Kunststoffwaren; Glasgewerbe; Keramik; Verarbeitung von Steinen und Erden	08	148	8.207
Metallerzeugung und -bearbeitung; H. v. Metallerzeugnissen	09	188	16.107
Maschinenbau	10	96	10.449
H. v. Büromaschinen, Datenverarbeitungsgeräten und -einrichtungen; Elektrotechnik; Feinmechanik und Optik	11	77	10.658
Fahrzeugbau	12	163	1.806
Energie- und Wasserversorgung	13	228	1.363
Baugewerbe	14	202	61.083
Handelsvermittlung und Großhandel	16	115	20.727
Gastgewerbe	18	125	16.543
Kreditgewerbe	21	198	2.788
Grundstücks- und Wohnungswesen; Vermietung beweglicher Sachen; Forschung und Entwicklung	24	22	27.940
Datenverarbeitung und Datenbanken	25	22	4.169
Rechts-, Steuer- und Unternehmensberatung; Markt- und Meinungsforschung; Beteiligungsgesellschaften	26	25	15.167
Architektur- und Ingenieurbüros; technische, physika- lische und chemische Untersuchung; Werbung	27	26	9.559
Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften; Detekteien und Schutzdienste; Reinigung von Gebäuden, Inventar und Verkehrs- mitteln; Erbringung von sonstigen Dienstleistungen überwiegend für Unternehmen	28	17	7.939

Wirtschaftsgliederung (H. v. = Herstellung von)	NACE 30- Angabe ¹⁾	Anzahl	Grund- gesamtheit
Kultur, Sport und Unterhaltung	29	16	8.650
Abwasser- und Abfallbeseitigung, sonstige Entsorgung; Interessenvertretungen sowie kirchliche und sonstige Religionseinrichtungen; Erbringung von sonstigen Dienstleistungen	30	30	15.765
Bergbau, Gewinnung von Steinen und Erden; Kokerei, Mineralölverarbeitung, Herstellung und Verarbeitung von Spalt- und Brutstoffen, Chemische Industrie	01 und 07	254	2.746
Holzgewerbe, H. v. Möbeln, Schmuck, Musikinstru- menten, Sportgeräten, Spielwaren und sonstigen Erzeugnissen; Recycling; Papiergewerbe	04 und 05	253	9.710
Kraftfahrzeughandel; Instandhaltung und Reparatur von Kraftfahrzeugen; Tankstellen; Einzelhandel (ohne Handel mit Kraftfahrzeugen und ohne Tankstellen); Reparatur von Gebrauchsgütern	15 und 17	221	48.343
Landverkehr; Transport in Rohrfernleitungen; Schiff- und Luftfahrt; Tätigkeiten für den Verkehr; Verkehrs- vermittlung; Nachrichtenübermittlung	19 und 20	204	16.138
Versicherungsgewerbe; mit dem Kredit- und Versiche- rungsgewerbe verbundene Tätigkeiten	22 und 23	52	905

1) Auswahl von 30 Wirtschaftsbereichen, die aus der NACE Rev. 1 für CVTS2 berücksichtigt werden konnten.

Beschäftigte des Unternehmens

Für Unternehmen mit mehr als 2.000 Beschäftigten am 31.12.1999 wurde die absolute Anzahl der Beschäftigten nicht ausgewiesen, sondern nur, dass das Unternehmen über 2.000 Beschäftigte hatte. Ferner wurden bei diesen Unternehmen für männliche und weibliche Beschäftigte nur die Anteile an den Gesamtbeschäftigten und anstelle der Beschäftigtenzahl Ende 1998 die Veränderung von 1998 zu 1999 in Prozent angegeben. Bei den Merkmalen, die abhängig von der Anzahl der Beschäftigten sind (geleistete Arbeitsstunden, Personalaufwendungen, Kosten, Teilnahmestunden, Teilnehmer), wurden die Originalwerte auf Pro-Kopf-Werte (Division durch die Anzahl der Beschäftigten Ende 1999) bzw. bei den Kosten und Teilnahmestunden auch auf Werte pro Teilnehmer umgerechnet. Von diesen Änderungen waren 83 Unternehmen betroffen, also knapp 3% aller Fälle. Die Grenze wurde bei 2.000 Beschäftigten gewählt, da bei den Unternehmen mit maximal 2.000 Beschäftigten der Abstand zur Beschäftigtenzahl des jeweils nächstgrößeren oder -kleineren Unternehmens immer weniger als 3% beträgt. Bei einem solch geringen Unterschied ist das Risiko einer möglichen Fehlidentifikation sehr hoch. Außerdem gab es oft deutliche Unterschiede zu vergleichbaren Angaben in kommerziellen Datenbanken, nicht zuletzt da in CVTS2 die Auszubildenden, Praktikanten und Trainees mit Ausbildungsvertrag nicht zu den Beschäftigten zählen.

Weitere Anonymisierungsmaßnahmen

Einige Merkmale wurden aus dem Datensatz entfernt oder modifiziert. Im Einzelnen betraf dies die folgenden Merkmale:

Jahresdurchschnitt Beschäftigte 1999: Diese Angabe war nur von Unternehmen mit starken saisonbedingten Schwankungen der Beschäftigten vorzunehmen. Lediglich für 76 Unternehmen waren hier Werte angegeben. Es wird stattdessen ein Merkmal „saison“ ausgewiesen mit $\text{saison} = 1$, falls es saisonbedingte Schwankungen der Beschäftigung gab und $\text{saison} = 0$ in allen übrigen Fällen.

Anteil der indirekten Kosten an den Personalaufwendungen insgesamt und Zahl der ganz oder teilweise mit Lehrveranstaltungen beschäftigten Personen: Diese beiden Merkmale wurden aus dem Datensatz entfernt, da nur vereinzelt gesicherte Angaben vorgelegen haben.

Fonds, an die Beiträge für die Weiterbildung gezahlt werden: Diese Merkmale fallen weg, da es bei regionalen Fonds nur 77 Ja-Antworten gab, bei nationalen Fonds 7 und bei sonstigen Fonds 47.

Einnahmen, Einnahmequellen, Saldo: Diese Merkmale fallen weg, da es im Datensatz nur 57 Unternehmen gibt, die Einnahmen aus Lehrveranstaltungen haben. Sollte ein potenzieller Datenangreifer Kenntnisse über diese Merkmale haben, so können in Kombination mit den anderen Überschneidungsmerkmalen (Wirtschaftsbereich und Beschäftigtenangabe) eindeutige Fälle entstehen. Es wurde daher ein neues Merkmal berechnet, welches den Wert 1 annimmt, wenn ein Unternehmen Einnahmen aus Lehrveranstaltungen hatte und 0 in allen übrigen Fällen.

Insgesamt verbleiben nach Anwendung der Anonymisierungsmaßnahmen etwa 180 Merkmale im Datensatz.

6 Beurteilung der Schutzwirkung

Zur Messung der Schutzwirkung wurde im Statistischen Bundesamt die Simulationssoftware Destatis-Anonymeter zur Durchführung sogenannter Massenfischzüge entwickelt.⁵ Bei einem Massenfischzug versucht ein Datenangreifer, möglichst viele Einheiten einer externen Datenbank den Zieldaten (vertrauliche, anonymisierte Daten) zuzuordnen. Unter Verwendung der kommerziell erhältlichen MARKUS-Datenbank als mögliche externe Datenbank eines potenziellen Datenangreifers wurden Simulationen durchgeführt. Dabei wurde insbesondere die im vorherigen Abschnitt reidentifizierende Wirkung von Informationen zum Standort der Unternehmen (Regionalinformationen) und die Notwendigkeit von Wirtschaftszweigszusammenfassungen deutlich. Zur Beurteilung der Schutzwirkung der Anonymisierungsmaßnahmen wurde die durch die Simulation erhaltene Verteilung des Reidentifikationsrisikos auf die oben beschriebenen Wirtschaftszweige und sieben Beschäftigtengrößenklassen⁶ unter Berücksichti-

⁵ Zur theoretischen Beschreibung und Funktionalität der Software siehe *Lenz, R.*: "Measuring the disclosure protection of micro aggregated business microdata – An analysis taking the example of German Structure of Costs Survey", erscheint in: *Journal of Official Statistics*, Schweden.

⁶ Folgende Kategorien wurden verwendet: 10-19 Beschäftigte, 20-49, 50-99, 100-249, 250-499, 500-999, 1.000 und mehr Beschäftigte.

gung der zugrunde liegenden Stichprobenauswahl, die bedeutet, dass ein potenzieller Datenangreifer in vielen Fällen Kenntnis über die Teilnahme des gesuchten Unternehmens an der Erhebung hat, herangezogen. Die mit den probeweise anonymisierten Daten durchgeführten Datenangriffs-Simulationen haben insgesamt gezeigt, dass hier eine Reidentifikation von Einheiten nur mit unverhältnismäßig großem Aufwand möglich und mit einer sehr großen Unsicherheit für den Datenangreifer behaftet ist. Die Daten können damit in Form eines sogenannten Scientific Use Files an die Wissenschaft weitergegeben werden.

7 Schlussbemerkung

Die Wissenschaft genießt seit Inkrafttreten des geltenden BStatG von 1987 eine Sonderstellung unter den Datennutzern der Bundesstatistik, das sogenannte Wissenschaftsprivileg. Es besteht darin, dass für die Durchführung wissenschaftlicher Vorhaben faktisch anonymisierte Mikrodaten vom Statistischen Bundesamt und den statistischen Ämtern der Länder an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden dürfen. Der große Vorzug einer solchen Datennutzung für Wissenschaftler liegt darin, dass ihnen die Einzeldaten übermittelt werden und diese von ihnen am eigenen Arbeitsplatz ausgewertet werden können.⁷ Ein weiterer Vorteil eines Scientific Use Files besteht darin, dass sichergestellt ist, dass verschiedene Nutzer mit demselben Datenmaterial arbeiten können. Dadurch können sich zwischen den Wissenschaftlern Forschungsnetzwerke leichter aufbauen und das Prinzip der Überprüfbarkeit von wissenschaftlichen Ergebnissen wird gestärkt.

Derzeit wird im Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder an einem sogenannten Campus File, einem für die Lehre breit einsetzbaren und für jedermann verfügbaren Datensatz zur CVTS2 gearbeitet, der seit dem Frühjahr 2007 verfügbar ist.

Der hier vorgestellte Scientific Use File ist über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder für die Wissenschaft erhältlich. Die hierzu notwendigen Anträge stehen als Download im Internet unter www.forschungsdatenzentren.de zur Verfügung.

Literaturverzeichnis

Egner, U. (2002): Berufliche Weiterbildung in Unternehmen (CVTS2), Projektbericht 2002.

Eurostat (2002): Europäische Sozialstatistik. Erhebung über die betriebliche Weiterbildung (CVTS2).

Lenz, R. (2006): „Measuring the disclosure protection of micro aggregated business microdata – An analysis taking the example of German Structure of Costs Survey“, erscheint in: *Journal of Official Statistics*, Schweden, Vol 22 (3).

⁷ Zum Vgl. siehe *Zühlke, Zwick, Scharnhorst und Wende*: The research data centres of the Federal Statistical Office and the Statistical Offices of the Länder. In: *Schmollers Jahrbuch*, Vol 124 (2004: 567-578).

Nestler, K. & Kailis, E. (2003): Arbeitszeitaufwand für betriebliche Weiterbildung in Europa. In: Statistik kurz gefasst - Bevölkerung und Soziale Bedingungen, 1/2003, 1-7.

Nestler, K. & Kailis, E. (2002): Betriebliche Weiterbildung in der Europäischen Union und Norwegen (-CVTS2-). In: Statistik kurz gefasst - Bevölkerung und Soziale Bedingungen, 3/2002, 1-7.

Zühlke, Zwick, Scharnhorst & Wende (2004): The research data centres of the Federal Statistical Office and the Statistical Offices of the Länder. In: Schmollers Jahrbuch, Vol 124, 567-578.

Maurice Brandt studierte Soziologie, Psychologie, Politikwissenschaften und BWL an der Friedrich-Wilhelms-Universität in Bonn und arbeitet als wissenschaftlicher Mitarbeiter am Forschungsdatenzentrum des Statistischen Bundesamtes in Wiesbaden. Arbeitsschwerpunkt ist derzeit das Projekt „Wirtschaftsstatistische Paneldaten und Faktische Anonymisierung“.

Dr. rer. nat. Rainer Lenz studierte Mathematik an der Technischen Universität Kaiserslautern. Sein Arbeitsschwerpunkt ist die Entwicklung von Anonymisierungsstrategien für Einzeldaten der amtlichen Statistik und die Leitung des Projektes „Wirtschaftsstatistische Paneldaten und Faktische Anonymisierung“ im Forschungsdatenzentrum des Statistischen Bundesamtes in Wiesbaden.

Dr. Hans-Peter Hafner promovierte in Mathematik an der Johannes Gutenberg - Universität in Mainz und arbeitet als wissenschaftlicher Mitarbeiter am regionalen Standort Wiesbaden des Forschungsdatenzentrums der Statistischen Landesämter. Fachlicher Schwerpunkt ist die Aufbereitung und Anonymisierung von Daten zu Verdienststruktur, Arbeitskosten und betrieblicher Weiterbildung.

Diplom-Kaufmann Daniel Schmidt ist wissenschaftlicher Mitarbeiter in der Gruppe VIB - Bildung, Forschung und Entwicklung, Kultur, Rechtspflege des Statistischen Bundesamtes. Sein Arbeitsschwerpunkt ist derzeit die Dritte Europäische Erhebung zur beruflichen Weiterbildung (CVTS3 - 3rd Continuing Vocational Training Survey).

Das **Forschungsdatenzentrum des Statistischen Bundesamtes** bietet zusammen mit dem **Forschungsdatenzentrum der Statistischen Landesämter** ein breites Angebot von ausgewählten Mikrodaten und ermöglicht den Zugang zu den Daten der amtlichen Statistik für die unabhängige wissenschaftliche Forschung. Weitere Informationen zu den beiden Forschungsdatenzentren, zum Datenangebot und zu den Nutzungsmöglichkeiten sind unter www.forschungsdatenzentrum.de erhältlich.