
Dokumentation



HOCHSCHUL - INFORMATIONEN - SYSTEM; GOSERIEDE 9, 30159 HANNOVER; Tel: 0511/1220-0

30.Juni 2006

**Dokumentation
des Scientific Use Files
„HIS-Absolventenpanel 1997“**

Gregor Fabian
Karl-Heinz Minks

Inhalt

1 Überblick über die HIS-Absolventenbefragungen.....	2
1.1 Geschichte	2
1.2 Funktion der Absolventenuntersuchungen.....	3
1.3 Aufbau und Inhalt der Absolventenbefragungen	4
1.4 Das HIS-Absolventenpanel 1997	6
1.4.1 Kurzbeschreibung.....	6
1.4.2 Grundgesamtheit und Stichprobe	7
1.4.3 Untersuchungsdesign	8
1.4.4 Reliabilitätsprüfung und Gewichtung	8
1.4.4.1 Anpassung der Stichprobe in der ersten Befragung.....	8
1.4.4.2 Anpassung der Stichprobe über die Erhebungswellen.....	10
1.5 Verlaufsdaten	11
2 Übersicht über die Merkmale des SUF HIS-Absolventenpanel	11
2.1 Bezeichnung der Variablen	11
2.2 Das Kodebuch	12
3 Anonymisierungsmaßnahmen.....	12
3.1 Schlussfolgerungen aus dem Pilotprojekt	12
3.2 Übersicht über die Anonymisierungsmaßnahmen im HIS-Absolventenpanel	14
3.3 Prüfung des Analysepotenzials	14
4 Ausgewählte Veröffentlichungen zu den HIS-Absolventenstudien	16
5 Anlage.....	16
6 Anhang	17

1 Überblick über die HIS-Absolventenbefragungen

Für den reibungslosen technischen Ablauf einer empirischen Analyse ist es unverzichtbar Informationen über das Zustandekommen der Daten zu haben. In diesem Kapitel werden die Geschichte und die Funktion der HIS-Absolventenerhebungen beleuchtet. Von zentraler Bedeutung sind dabei Angaben zum Aufbau und zur Durchführung der Studie. Mittelpunkt des Codebuches sind Informationen zur Struktur und Aufbereitung sowie der Verteilung der kodierten Merkmale.

1.1 Geschichte

Bereits seit den 1970er Jahren führt HIS Befragungen im Bereich der Hochschulforschung durch. Neben Studienberechtigten und Studienanfängern, sind Studierende und Hochschulabsolventen die Untersuchungseinheiten der jeweiligen Erhebungen. Befragungen von Hochschulabsolventen haben bei HIS eine lange Tradition. Sie haben verschiedene Phasen der Hochschulentwicklung durchlebt. 1974 wurde mit der ersten Exmatrikuliertenbefragung begonnen, in die neben den Absolventen auch Studienabbrecher und Hochschulwechsler einbezogen wurden. Ab 1989 gab es mit der Umstellung von Exmatrikulierten- auf reine Absolventenbefragungen eine Zäsur. Diesem Wechsel des Ansatzes lag das Bedürfnis nach fein differenzierten fachspezifischen Daten zum Studienverlauf, zur Studienqualität, zum Berufsübergang und Berufsverbleib von Absolventen zugrunde. Damit erweiterte sich der Nutzerkreis. Aufgrund der feineren Differenzierung und der stärkeren Zuwendung zu Fragen der Qualifikationsanforderungen wurde die Lehre stärker auf die Ergebnisse der Untersuchungen aufmerksam. Beeinflusst durch die ansteigende Nachfrage nach Daten aus verschiedenen Erhebungen, stellt sich auch für HIS die Frage nach der Bereitstellung von Datenfiles für interessierte Wissenschaftler. Seit Jahren existieren verschiedene Auswertungsmöglichkeiten für HIS-Daten. Eine besonders flexible Nutzungsform stellt die Arbeit mit einem Scientific Use File dar. Auf diese Weise kann der Umfang der wissenschaftlichen Nutzung deutlich erhöht werden. Damit könnte das Potenzial der Absolventendaten umfangreicher, differenzierter und effektiver genutzt werden.

1.2 Funktion der Absolventenuntersuchungen

Die HIS-Absolventenbefragungen sollen Informationen zu allen aktuellen und grundsätzlichen Fragen liefern, die im Zusammenhang mit dem Studium, dem Hochschulabschluss, dem Berufsübergang und der beruflichen Integration stehen. Das Instrument besteht aus einem Kern an Fragen, der die Befragungswellen überdauert und einem bei jeder Befragung zusätzlich eingebrachten Teil zu aktuellen hochschulpolitisch bzw. hochschulisch relevanten Fragen. Es gibt vier wesentliche Funktionen der Absolventenbefragungen.

1. Analysen zum Zusammenhang von Studium (Studienverlauf, Studienerfolg und Studienqualität) und Berufserfolg

Hierzu zählen u. a. Analysen über Hintergründe unterschiedlicher Studiendauer, zur Funktion von Aufbaustudiengängen und ihrer Wirkung auf das berufliche Fortkommen, zur Selbständigkeit von Hochschulabsolventen, über die Teilnahme von jungen hoch Qualifizierten an beruflicher Weiterbildung sowie über die Werdegänge von Ingenieuren und Naturwissenschaftlern. Neben einer Reihe von publizierten fächerbezogenen Reports entwickelte sich in den letzten Jahren ein thematischer Schwerpunkt, der für die Hochschulentwicklung und die Qualität der Lehre und des Studiums von zentraler Bedeutung ist: die Auswirkungen der Veränderung hoch qualifizierter Arbeit auf die Qualifikationserfordernisse von Hochschulabsolventen.

2. Ermittlung von fächerbezogenen repräsentativen Daten, im Sinne des Benchmarking für Hochschulen

Die Datenberichte erlauben jeder Hochschule bzw. jedem Fachbereich, sich mit den bundesweit repräsentativen Ergebnissen zu vergleichen, sofern bei lokalen Absolventenbefragungen gleiche bzw. vergleichbare Instrumente bzw. Fragen zum Einsatz kommen. Eine Reihe von Hochschulen nutzt diese Möglichkeit bereits. Da für die Fragebogen kein Copyright existiert, kann sich jede Hochschule und jedes Forschungsinstitut der in der Projektgruppe entwickelten Instrumente frei bedienen.

3. Unterstützung der Evaluation an Hochschulen

Schon vor der Verwendung von Begriffen wie Qualitätssicherung und Evaluation im hochschulischen Bereich hatte die Arbeit in den Absolventenprojekten - auf Nachfrage von Hochschulen, Prüfungsverwaltungen bzw. Fachbereichen - unmittelbar evaluative Aspekte. In zahlreichen Beispielen wurden Fragen zur Ausbildung von Hochschulabsolventen mit dem Ziel verfolgt, zur Aufklärung über die Effizienz des Studiums und den Berufserfolg von Absolventen vor Ort beizutragen.

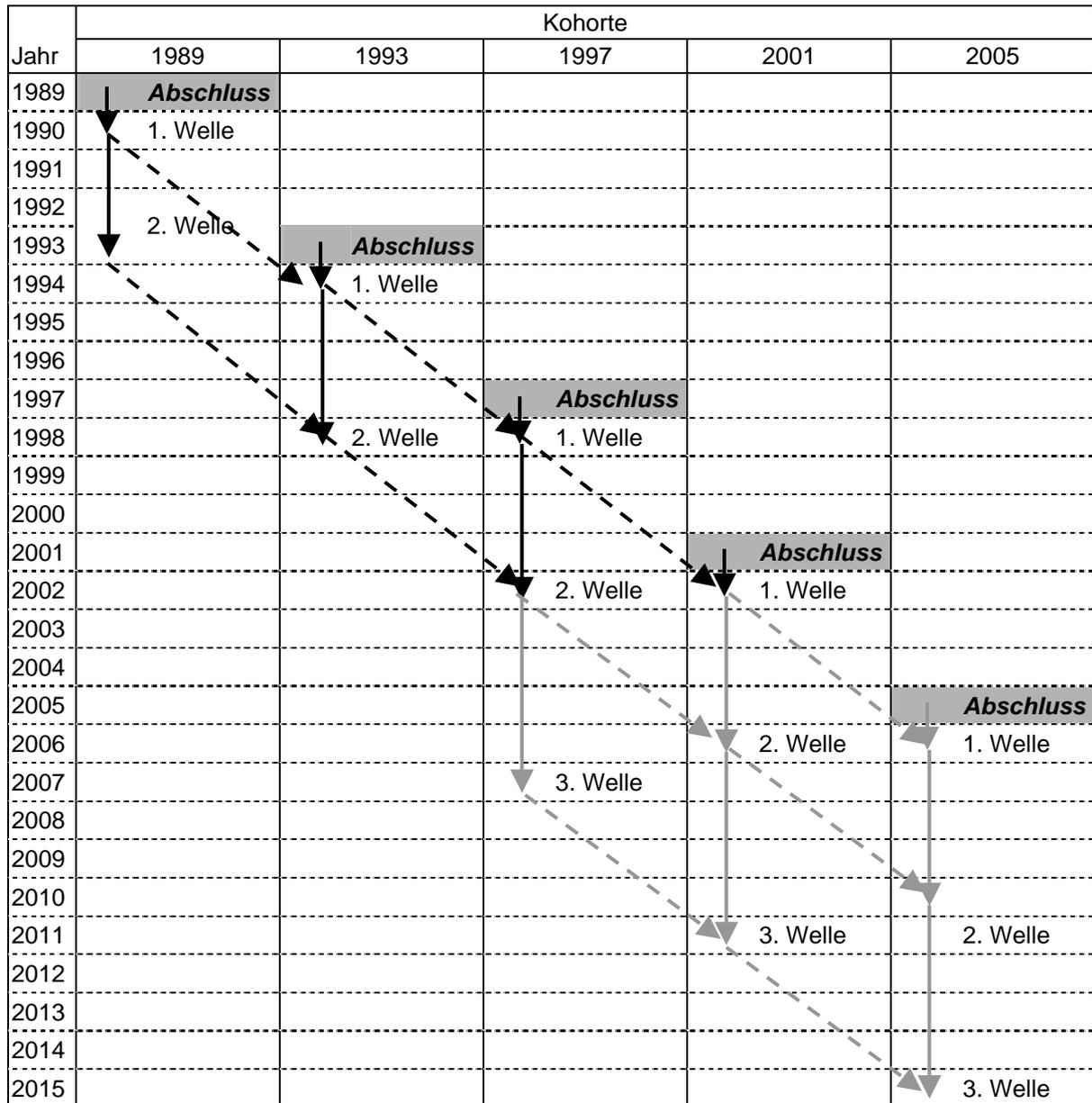
4. Länderspezifische und regionale Untersuchungen

Soweit es die Stichprobe zulässt, können auch regionale bzw. länderbezogene Analysen durchgeführt werden. Die Möglichkeit, dies flächendeckend für alle Länder zu tun, war trotz der relativ großen Stichprobe bis zum Absolventenpanel 2005 nicht gegeben. Dies gilt insbesondere für kleinere Länder und Regionen. Für das Absolventenpanel 2005 wurde nun ein Verfahren zur Stichprobenziehung entwickelt, das es ermöglicht, für möglichst viele Länder spezifische Analysen durchzuführen. Auf der Basis der Befragung des Absolventenjahrgangs 2001 konnten solche länderspezifischen Untersuchungen bereits für Brandenburg und Nordrhein-Westfalen durchgeführt werden.

1.3 Aufbau und Inhalt der Absolventenbefragungen

Die Befragungen sind als Panel angelegt. Die erste Welle wird je nach Prüfungsdatum zwischen 6 und 18 Monaten nach Examen durchgeführt. Die zweite Befragung fand bei den Absolventinnen und Absolventen 1989, für die keine obligatorische zweite Ausbildungs- oder praktische Phase vorgesehen ist, drei bis dreieinhalb Jahre nach Studienabschluss, bei den übrigen, vor allem staatlich und kirchlich geprüften Graduierten vier bis viereinhalb Jahre nach dem Examen statt. Bei den bislang realisierten (Kohorten 1993 und 1997) bzw. geplanten (Kohorte 2001) zweiten Befragungen umfasst das Beobachtungsfenster einen Zeitraum von ca. fünf Jahren nach der Abschlussprüfung. Für die Absolventen des Prüfungsjahres 1997 ist zum ersten Mal eine dritte Befragung, 10 Jahre nach Beendigung des Studiums geplant. Die Erhebungen der unterschiedlichen Jahrgänge (1989, 1993, 1997, 2001, 2005) sind so angelegt, dass die Kohorten miteinander vergleichbar sind. Dadurch sind Aussagen über Veränderungen in den Erwerbskarrieren und der Lebensführung verschiedener Absolventenkohorten an deutschen Hochschulen möglich.

Abbildung 1: Design der HIS-Absolventenpanels



—▶ Längsschnitt realisiert
 - -▶ Trendanalyse realisiert

—▶ Längsschnitt geplant
 - -▶ Trendanalyse geplant

1.4 Das HIS-Absolventenpanel 1997

1.4.1 Kurzbeschreibung

Inhaltliche Schwerpunkte

1. Befragung

- Studienverlauf
- Evaluation des Studiums, Kompetenzen und Kenntnisse der Hochschulausbildung
- weitere Ausbildungsphasen (Studium, Promotion, zweite Ausbildungsphase)
- Start und Verlauf von Berufs- und Erwerbskarriere
- Selbständigkeit
- soziales Umfeld, soziale Herkunft

2. Befragung

- Übergangsverläufe in das Erwerbsleben
- berufliche Fort- und Weiterbildung
- Wege der Zukunftssicherung
- Partnerschaft und Familie

3. Befragung (geplant)

- weiterer Berufsverlauf
- berufliche Integration / berufliche Desintegration
- Selbständigkeit
- Vereinbarkeit von Familie und Beruf

Untersuchungseinheit

- Hochschulabsolventen deutscher Universitäten nach Beendigung ihres Erststudiums
- Prüfungszeitraum: Prüfungsjahr 1997 (September 1996 – September 1997)

Grundgesamtheit

- 191.948 Absolventen des Erststudiums an deutschen Hochschulen

Stichprobe und Rücklauf

- Netto-Stichprobe 1. Welle (1998): 9583 Personen
- 2. Welle (2003): 6220
- im Scientific Use File sind 6216 Personen enthalten
- die Ausschöpfungsquote beträgt damit ca. 3,6 Prozent aller Absolventen

1.4.2 Grundgesamtheit und Stichprobe

Die Grundgesamtheit, über die Aussagen getroffen werden soll, besteht aus allen Hochschulabsolventen, die im Prüfungsjahr 1997 ihren ersten berufsqualifizierenden Studienabschluss an einer Hochschule im Gebiet der Bundesrepublik Deutschland erreicht haben. Nicht in die Stichprobe einbezogen sind die Bundeswehrhochschulen, die Verwaltungsfachhochschulen, die Hochschulen für Berufstätige und die Berufsakademien. Da HIS vorab keine Adressen der Absolventen zur Verfügung stehen, hat der Versand der Befragungsunterlagen durch die Prüfungsverwaltungen zu erfolgen. Die sehr heterogenen Strukturen der Prüfungsverwaltungen erlauben es nicht, vor Ort Zufallsstichproben der jeweiligen Personen zu ziehen. Aus diesem Grund ist eine Klumpenstichprobe – alle Absolventen der gezogenen Fachrichtung einer Prüfungsverwaltung – erforderlich. Die Grundgesamtheit für das Prüfungsjahr 1997 stellen 191.948 Personen dar. Um bei einer einfachen Klumpenauswahl das größere Verzerrungsrisiko zu minimieren, ist ein relativ aufwändiges Verfahren der Stichprobenziehung erforderlich.

Die Klumpen werden dabei durch folgende Merkmale definiert:

- Studienbereich (nach der Definition des Statistischen Bundesamtes)
- Abschlussart (Diplom an Universitäten und vergleichbare Abschlüsse (Magister, Staatsexamen außer Lehramt, künstlerisches Examen), Diplom an Fachhochschulen, Staatsexamen für ein Lehramt (ohne Differenzierung nach Fachrichtung))
- Hochschule

Außerdem wird die Verteilung der Absolventen nach Hochschulart und Studienort (alte bzw. neue Bundesländer) berücksichtigt. Bei Lehramtsstudiengängen ist die Kontrolle des Studienbereichs nicht möglich, da deren Absolventen in der Regel mehrere Fächer studiert haben und nicht in allen Bundesländern eine formale Festlegung auf ein Hauptfach vorgesehen ist. Um Aussagen über die bildungs- und arbeitsmarktpolitisch interessanten Gruppen der Absolventen der neuen bzw. alten Bundesländer machen zu können, wird eine disproportional geschichtete Klumpenstichprobe gezogen, in der die Absolventen aus den neuen Bundesländern überproportional vertreten sind. Bei Aussagen über die Gesamtheit der Absolventen wird dieser Überrepräsentanz durch geeignete Gewichtungsfaktoren (vgl. Kapitel 2.4.4) Rechnung getragen.

Um zu gewährleisten, dass die Verteilung in der Stichprobe derjenigen der Grundgesamtheit entspricht, werden mehrere tausend Stichproben zufällig gezogen und eine Stichprobe mit geringen Abweichungen zur Grundgesamtheit ausgewählt. Alle Klumpen haben dabei die

gleiche Auswahlwahrscheinlichkeit. Alle Absolventen der ausgewählten Klumpen wurden in die Untersuchung einbezogen (einstufige Auswahl).

Um differenzierte Auswertungen vornehmen zu können, war angesichts der bei standardisierten schriftlichen Befragungen üblichen Ausschöpfungsquote und der zu erwartenden Panelmortalität eine Brutto-Stichprobe von ca. 35.000 Absolventen erforderlich.

1.4.3 Untersuchungsdesign

Die Befragungen erfolgten schriftlich und postalisch. In den einzelnen Panelwellen werden zum Teil dieselben Merkmale (z. B. Qualifikationsanforderungen der ausgeübten Berufstätigkeit) erhoben, so dass zeitdiskret erhobene Längsschnittdaten vorliegen, die die Analyse von Veränderungen zwischen den Panelwellen erlauben. Darüber hinaus werden der Berufsverlauf sowie familiäre Prozesse retrospektiv mit einem Event-History-Design erfasst, das umfassende Informationen über die Geschichte und das Timing von Ereignissen zwischen den Befragungen und damit kontinuierliche Verlaufsdaten liefert. Diese Mischung von mehrmaliger und retrospektiver Erhebung hat den Vorteil, dass Längsschnittdaten für Merkmale vorliegen, die nicht retrospektiv erfasst werden können, etwa Einstellungen, aber z. B. auch Qualifikationsanforderungen, die für frühere Beschäftigungen nur schwer rekonstruierbar sind. Mit diesem komplexen Design existiert eine in Deutschland einmalige Datenbasis über den Berufseintritt, den Berufsverlauf und die Berufstätigkeit von Hochschulabsolventen, die sich für vielfältige Analysen nutzen lässt.

Das Konzept der Absolventenuntersuchung sieht wie bei den Vorgängerbefragungen vor, dass in der ersten Befragungswelle der Versand der Befragungsunterlagen und Erinnerungsschreiben durch die Prüfungsämter erfolgt. Bei den folgenden Panelwellen können die Befragungsunterlagen an die Anschriften versandt werden, die die Teilnehmer im Zuge der ersten Befragung mitgeteilt haben.

1.4.4 Reliabilitätsprüfung und Gewichtung

Die Repräsentativität von Paneldaten muss für jeden Befragungszeitpunkt neu kontrolliert werden. Durch das Stichprobendesign bedingte Disproportionalitäten, sind mit Hilfe entsprechender Gewichtungsfaktoren auszugleichen.

1.4.4.1 Anpassung der Stichprobe in der ersten Befragung

Das Untersuchungsdesign sah vor, die Absolventen so zu erheben, dass repräsentative Aussagen für Studienbereiche, Abschlussarten und in der Unterscheidung nach alten bzw.

neuen Bundesländern¹ möglich sind. Durch die Konzeption der Stichprobe und unterschiedlich hohe Rücklaufquoten der einzelnen Hochschulen ist eine derartige Abbildung nicht immer gewährleistet. Um die Aussagen aus den Analysen trotzdem auf die Grundgesamtheit übertragen zu können, sind Gewichte berechnet worden. Sie stehen getrennt gerechnet für die alten und neuen Bundesländer mit dem Merkmal „X1gewoswe“ und für das gesamte Bundesgebiet mit dem Merkmal „X1gewinsg“ zur Verfügung.

Die Gewichte sind durch den Vergleich der Verteilungen aus Stichprobe und Grundgesamtheit entstanden. Dabei wurden alle Befragten aus der ersten Welle berücksichtigt, die valide Werte auf den entsprechenden Merkmalen aufweisen. Für die Berechnung der Verhältnisse in der Grundgesamtheit wurden die amtlichen Daten des Statistischen Bundesamtes aus dem Prüfungsjahr 1997 unter Nutzung der HIS-Datenbank „Information, Controlling, Entscheidung“ (ICE) herangezogen. Ein Beispiel soll das Zustandekommen der einzelnen Gewichte verdeutlichen. Im Studienbereich Wirtschaftswissenschaften schlossen im Prüfungsjahr 1997 in den neuen Bundesländern 1415 Personen mit einem Universitätsdiplom ab. Das entspricht 8,66 Prozent aller Absolventen. In der Stichprobe der 1. Welle finden sich 195 Absolventen aus diesem Studienbereich. Bezogen auf die Stichprobe sind das 16,18 Prozent aller Absolventen. Damit sind die Wirtschaftswissenschaftler mit Universitätsdiplom in der Stichprobe überrepräsentiert. Das Verhältnis wird entsprechend durch ein Gewicht mit dem Faktor 0,55 angepasst. Für Wirtschaftswissenschaftler, die an einer Fachhochschule absolvieren, ergibt sich folgendes Verhältnis: Hier sind die Absolventen mit 4,65 Prozent der Befragten gegenüber den Absolventen der Grundgesamt unterrepräsentiert (7,35 Prozent). Die entsprechenden Fälle werden deshalb mit dem Faktor 1,58 gewichtet.

Für Berechnungen auf Bundesebene werden die Daten mit einem errechneten Ausgleichsfaktor gewichtet, der die Verhältnisse der Teilgebiete der Grundgesamtheit berücksichtigt. So können die bei der Stichprobenziehung bewusst eingesetzten Disproportionalitäten ausgeglichen werden. Berücksichtigt man diese Gewichtungsfaktoren in der Auswertung, entsprechen die Anteile der Hochschulabsolventen in der Stichprobe, unterschieden nach Studienbereich und Abschlussart, denen der Grundgesamtheit.

¹ In der Absolventenerhebung 1997 wird Berlin zu den alten Bundesländern gerechnet.

1.4.4.2 Anpassung der Stichprobe über die Erhebungswellen

Werden Personen über verschiedene Erhebungen wiederholt befragt, kommt es aus unterschiedlichen Gründen zu Ausfällen im Rücklauf. Dieses Phänomen wird als Panelmortalität bezeichnet. Der Ausfall einzelner Personen aus dem Panel hat nicht zwangsläufig Verzerrungen in den Analysen zur Folge. Aussagen können dann auf die Grundgesamtheit übertragen werden, wenn sichergestellt ist, dass die Ausfälle unsystematisch sind. Eine bestimmte Gruppe darf nicht signifikant häufiger oder seltener an der Folgebefragung teilnehmen. Durch so genannte Ausfallanalysen können Veränderungen festgestellt und gegebenenfalls ausgeglichen werden. Im Zuge der zweiten Befragung der 1997er Absolventen in den Jahren 2002 und 2003 sind alle relevanten Merkmale in univariaten und bivariaten Verteilungen geprüft worden. Einbezogen wurden die Merkmale Fachgliederung (C1fagrup) in Kombination mit Geschlecht, unterschieden nach neuen bzw. alten Bundesländern. Weiterhin wurden die Bundesländer der Hochschulorte, das Land der aktuellen Beschäftigung, deren Wirtschaftsbereiche und verschiedene Tätigkeitsepisoden allgemein und zu bestimmten Zeitpunkten geprüft. Hier wurde besonderes Augenmerk auf im Sample eher seltene Tätigkeitsformen, etwa Arbeitslosigkeit, Übergangsjobs oder Honorartätigkeit gelegt. Außerdem wurden Unterschiede in der Studiendauer, der Examensnote und bezüglich des Auslandsstudiums gesucht. In keinem Falle konnten signifikante Abweichungen festgestellt werden. Gruppen, die erfahrungsgemäß häufiger aus Panelbefragungen herausfallen, etwa Arbeitslose oder besonders mobile Personen, sind sowohl in der ersten Welle als auch in der zweiten Befragung zu gleichen Teilen vertreten. Diese Analysen wurden an Merkmalen vorgenommen, die Überschneidungsbereiche in beiden Wellen aufweisen, so dass eine Längsschnittgewichtung für die Befragten der zweiten Welle der Absolventen 1997 nicht notwendig war.

In einem weiteren Arbeitsschritt wurden Erwerbsepisoden aus beiden Wellen verglichen. Dazu wurde die Schnittmenge der Angaben über Erwerbstätigkeit in beiden Welle herangezogen. Sicherlich sind hier die besonderen Umstände der Befragten zu berücksichtigen, die Beantwortung einer Frage zu verschiedenen Zeitpunkten kann unterschiedlich ausfallen. Trotzdem ist festzuhalten, dass bei einem überwältigenden Anteil von Befragten keine Unterschiede in den Aussagen zu Erwerbstätigkeit auftraten. Lediglich bei einem unerheblichen Anteil der Befragten mussten einige wenige Angaben zur Erwerbskarriere angepasst werden.²

²Die Unterschiede ergeben sich zum Teil aus fehlerhaften Angaben der Zeiträume von Erwerbsepisoden. So kommt es vor, dass Episoden um 12 Monate verschoben werden.

1.5 Verlaufsdaten

Das Mikrodatenfile kann um die Berufsepisoden der Befragten ergänzt werden. Dafür existiert ein so genannter Verlaufsdatensatz. Er enthält Angaben über Zeitraum und Inhalt der Tätigkeiten seit Ende des Studiums. Die Verlaufsdaten sind aus den Tätigkeitsmerkmalen der Frage 4.1 in Fragebogen 97.1 und dem Kalendarium aus der zweiten Befragung (Frage 1.6) entstanden. Die Überschneidungen der Tätigkeitszeiträume wurden bewusst konzipiert und dienen unter anderem der Reliabilitätsprüfung der erhobenen Daten. Bei widersprüchlichen Angaben in beiden Wellen, wurden Einzelfallprüfungen vorgenommen und die Merkmale gegebenenfalls angepasst.

2 Übersicht über die Merkmale des SUF HIS-Absolventenpanel

Die Merkmale des SUF wurden in zwei Wellen erhoben. Für einen besseren Überblick wurde eine Struktur entwickelt, die in verschiedenen Detailstufen, Auskunft über den Inhalt der Befragungen gibt. Sie ist so gewählt, dass weitere Wellen problemlos eingebunden werden können.

2.1 Bezeichnung der Variablen

Die Variablennamen beinhalten den jeweiligen Bereich, die Welle, einen Variablenstamm und einen lautmalerischen Zusatz, aus dem sich der Inhalt des Merkmals ergeben soll. Die Bereiche A bis K verweisen auf verschiedene Schwerpunkte der Befragung. So ist einem inhaltlichen Bereich zuzuordnen. Die Vorgehensweise bei der Bezeichnung und deren Bedeutung lässt sich an einem Beispiel veranschaulichen. Das Merkmal „A1sitedv“ enthält Angaben über Beurteilung einzelner Aspekte des abgeschlossenen Studiums. A bezeichnet den Bereich „Studienverlauf“, der Wert 1 verweist auf die Welle in der das Merkmal abgefragt wurde. Mit dem Variablenstamm *sit* ist auf den Fragekomplex „Situation im Hochschulstudium“ verwiesen und *edv* fragt konkret nach den Zugangsmöglichkeiten zu EDV-Diensten an der Hochschule. Durch die gewählte Bezeichnung der im SUF enthaltenen Variablen sollte auch bei kurzer Einarbeitungszeit eine ausreichende Kenntnis des Datenfiles möglich sein.

Abbildung 2: Die vier Ebenen der Variablenbezeichnung im Scientific Use File

Ebene und Welle	Welle	Stamm	Variablenzusatz
A1sitedv	A1sitedv	A1sitedv	A1sitedv
Kompetenzen, Kenntnisse, Studienevaluation	in der 1. Welle erhoben	Fr.1.16: Beurteilung einzelner Aspekte des abgeschlossenen Studiums	Zugang zu EDV- Diensten

2.2 Das Kodebuch

Das Kodebuch ist in 4 Ebenen unterteilt, die inhaltlich und technisch miteinander verbunden sind. Die erste Ebene fasst die inhaltlichen Bereiche der Befragungen zusammen, die Ebene der Variablenstämme gibt einen Überblick über verschiedene Fragen einzelner Bereiche. Auf Variablenebene finden sich neben den Fragetexten, die Variablenbezeichnungen sowie kurze technische Erklärungen zu den einzelnen Merkmalen. Außerdem sind die Kunstvariablen angefügt. Häufigkeitsverteilungen und Merkmalsausprägungen der einzelnen Variablen finden sich in einer vierten Ebene des Kodebuchs. Die Bildungsvorschriften für die Kunstvariablen sind ebenfalls dort zu finden.

3 Anonymisierungsmaßnahmen

Das Pilotprojekt „HIS-Absolventenpanel 1997“ diente vor allem dazu, spezifische Gefahrenpotentiale für die Kohorte der Hochschulabsolventen, aber auch der Studierenden und Studienberechtigten, abzugrenzen. Durch die Merkmale „Hochschule“, „Studienfach“, „Abschlussart“ und „Abschlussjahr“ sind einzelne Gruppen von Hochschulabsolventen einfach von anderen Gruppen zu unterscheiden. Sie wurden deshalb in der Risikoabwägung besonders berücksichtigt.

3.1 Schlussfolgerungen aus dem Pilotprojekt

Durch die Simulation der verschiedenen Angriffszenarien wird deutlich, dass die Wahrscheinlichkeit für die erfolgreiche Zuordnung von personenbezogenen Daten aus Identifikationsfiles mit einzelnen Fällen aus dem Mikrodatenfile gering ist. Außerdem zeigt sich, dass der Umfang potenzieller Identifikationsfiles durch die notwendige Verfügbarkeit von Überschneidungsmerkmalen, bei gleichzeitig eindeutiger Abgrenzung der Zielpopulation, stark limitiert wird. So ist es nicht besonders wahrscheinlich, dass die Verwaltungen der Hochschulen Zusatzwissen für eine mögliche Reidentifikation zur Verfügung stellen. Im Einzelfall ist aber eine eindeutige Zuordnung nach Studienfach und Hochschule möglich. Die

Verfügung über Datenbanken bzw. Absolventenbücher ist aufgrund entstehender Kosten bzw. des von vornherein eingeschränkten Zuganges ebenfalls nicht ohne erheblichen Aufwand vorstellbar. Sollten solche Quellen trotzdem als Zusatzwissen zur Verfügung stehen, beschränkt sich das Wissen auf einzelne Studiengänge an bestimmten Hochschulen. Dies gilt in noch stärkerem Maße für die meisten Absolventenstudien. Hier besteht zudem nur dann eine direkte Gefahr, wenn auf personenbezogene Daten tatsächlich auch zugegriffen werden kann. Frei bzw. einfach zugängliche Quellen lassen zwar einen beachtlichen Umfang an Zusatzwissen vermuten. Die simulierten Angriffszenarien zeigen jedoch, dass weder für Einzelpersonen noch für Gruppen von Absolventen ein erfolgsversprechendes Maß an Zusatzwissen bereitgestellt werden kann. Trotzdem ist nicht auszuschließen, dass, etwa bei Teilnahmekennntnis bestimmter Personen oder durch den Abgleich mehrerer durch Zusatzwissen aufgefundener Personen, eine eindeutige Zuordnung - auch mit relativ geringem Aufwand - möglich ist. Der wesentliche Risikofaktor für die faktische Anonymität von Absolventendaten besteht in der gleichzeitigen regionalen und sachlichen Tiefengliederung. Durch die Vergrößerung des Merkmals Hochschule, durch die regionale Unterscheidung nach Alten und Neuen Bundesländern sowie die moderate Anpassung des Merkmals Studienfach wird dieser Risikofaktor minimiert. Weitere Regionalangaben, etwa zum Arbeitsort, stellen kein erhöhtes Reidentifikationsrisiko dar, wenn durch sie kein direkter Rückschluss auf die besuchte Hochschule möglich ist. Daten, die ohne kleinräumige Regionalinformationen übermittelt werden, sind in aller Regel bereits durch die Entfernung personenbezogener Angaben und der von einer Stichprobe ausgehenden Schutzwirkung faktisch anonym. Auf die Ziehung einer Substichprobe als zusätzliche Schutzmaßnahme, die besonders bei Paneldaten Einschränkungen des Analysepotenzials zur Folge hat, kann durch die Umsetzung der oben vorgeschlagenen Maßnahmen verzichtet werden (Wirth 2006:9).

Die Gefahr, dass eine Reidentifikation mit geringem Aufwand möglich ist, lässt sich dann praktisch ausschließen, wenn Fälle im Datensatz nicht mehr auf den Hochschulort bezogen werden können. Die eindeutige Zuordnung eines Datensatzes zu einer bestimmten Hochschule ist auch dann möglich, wenn die Merkmalskombination aus den Überschneidungsmerkmalen im Mikrofile einzigartig ist. Diese Fälle müssen in der Stichprobe so verändert werden, dass sie in der Grundgesamtheit keine Einzelfälle darstellen bzw. nur an mehr als einer Hochschule in den alten bzw. neuen Ländern vorkommen.

Faktische Anonymität für das Scientific Use File „HIS-Absolventenpanel“ ist also dann hergestellt, wenn Merkmale, die einen Hochschulbezug des Befragten möglich machen, entsprechend anonymisiert werden. Die Anonymisierungsmaßnahmen sollen die Daten im

SUF so anpassen, dass die Analysequalität gegenüber den Originaldaten möglichst unverändert bleibt.

Die Szenarien zeigen auch, dass die Quellen für Zusatzwissen sich mit der Zeit verändern können. Deshalb ist bei jeder Erstellung eines Scientific Use Files zu prüfen, ob und inwieweit Veränderungen in der Verfügbarkeit von Zusatzwissen Einfluss auf Reidentifikationsrisiken haben können.

3.2 Übersicht über die Anonymisierungsmaßnahmen im HIS-Absolventenpanel

Vergrößerung	
Hochschulen der Abschlüsse und des Studienverlaufes	→ dichotomisiert zu „neue/alte Bundesländer“ Die Zusammenfassung betrifft ersten, zweiten und dritten Abschluss, sowie fünf mögliche Studienepisoden
Studienfächer	→ aggregiert zu Studienbereichen Das betrifft 16 Merkmale in Studienverlauf und Abschluss
Postleitzahlen der Arbeitsorte	→ zusammengefasst zu Raumordnungsregionen Das betrifft 9 Job-Episoden in der zweiten Befragung
Merkmale entfernt	
Entfernung Hochschulort-Arbeitsort	→ entfernt
Sonstiges	
4 Fälle	In der Grundgesamtheit eindeutig zuordenbare Datensätze wurden entfernt
Einzelfälle	Der eindeutige Hochschulbezug wurde unterdrückt

3.3 Prüfung des Analysepotenzials

Die zur Erstellung eines SUF notwendigen Anonymisierungsmaßnahmen haben immer Auswirkungen auf das Analysepotenzial des Datenfiles. Dies trifft insbesondere auf Datensätze mit wirtschaftsstatistischen Angaben zu. Für den Nachweis des Analysepotenzials von SUF-Daten werden deshalb normalerweise umfassende Vergleiche zwischen Originalfile und Scientific Use File vorgenommen. Im vorliegenden Fall erübrigen sich diese Maßnahmen. Das erklärt sich vor allem aus dem vergleichsweise geringen Anteil an vorgenommenen Anpassungen, die zudem nur wenige Merkmale betreffen. Die Unterdrückung des Hochschulbezuges macht Reidentifikationsversuche derart aufwändig und unwahrscheinlich, dass auf die Anpassung von Extremwerten weitestgehend verzichtet

werden konnte. Lediglich die Regionalangaben zu den Arbeitsorten mussten aggregiert werden.

Die Veränderungen für das Merkmal „Abschlussart“ werden so vorgenommen, dass die Voraussetzungen zur Repräsentativität der Stichprobe erhalten bleiben.

Zur Erhaltung des Reidentifikationsschutzes wurden die Studienbereiche 74 und 75 zusammengelegt. Analysen, die eine Differenzierung dieser Merkmale erfordern, können mit den SUF-Daten nicht durchgeführt werden³. Analysen, die sich auf einzelne Hochschulen stützen, sind mit dem Scientific Use File „HIS-Absolventenpanel 1997“ ebenfalls nicht durchführbar. Eine weitere, eher hypothetische Einschränkung ergibt sich bei Berücksichtigung des Studienganges, sofern er auf Ebene der Studienfächer in die Analysen aufgenommen werden soll. Die für die Bundesrepublik repräsentativ erhobenen Studienbereiche sollte in nahezu allen Fällen für die Analyse genügen.

Auch wenn vier Fälle aus dem SUF entfernt und einzelne Datensätze angepasst werden mussten, haben die Veränderungen keine Auswirkungen auf das Analysepotenzial.

Die Plausibilisierung des SUF erfolgte durch den Vergleich des aufbereiteten Datensatzes mit Hilfe der Veröffentlichungen zu den Absolventenprojekten. Als Referenz diente dabei die Veröffentlichung von Kerst und Kollegen (2005a). Bedingt durch die notwendigen Anpassungen, kann es in einzelnen Zellen zu geringfügigen Abweichungen kommen. Zur Prüfung des fehlerfreien Ablaufes können deshalb auch die Verteilungen im Anhang genutzt werden⁴. Hier sind Abweichungen gegenüber dem Originalfile markiert.

³ Allerdings sind diese Studienbereiche in bisherigen HIS-Analysen nie einzeln ausgewertet worden.

⁴ Zur Prüfung des selbst erstellten Datenfiles eignet sich das auch Kodebuch.

4 Ausgewählte Veröffentlichungen zu den HIS-Absolventenstudien

Briedis, Kolja (2006): Die HIS-Absolventenstudien. Vorstellung der bundesweit repräsentativen Absolventenpanels. Gehalten auf der Tagung "Potentiale von Absolventenstudien für die Hochschulentwicklung" in Kassel.
http://www.his.de/Abt2/Berufseintritt/absolventenprojekt/vortrag/Vortrag_Kassel_Abs.pdf
zuletzt recherchiert 30.06.2006

Kerst, Christian; Minks, Karl-Heinz (2005a): Fünf Jahre nach dem Studienabschluss Berufsverlauf und aktuelle Situation von Hochschulabsolventinnen und -absolventen des Prüfungsjahrgangs 1997, HIS-Hochschulplanung Band 173, Hannover

Kerst, Christian; Minks, Karl-Heinz (2005b): Selbständigkeit und Unternehmensgründung von Hochschulabsolventen fünf Jahre nach dem Studium. Eine Auswertung der HIS-Absolventenbefragungen 2002/2003, HIS, Hannover

weitere Veröffentlichungen: <http://www.his.de/Abt2/Berufseintritt/absolventenprojekt/>

5 Anlage

Den Datenfiles liegen neben der Dokumentation, der Fragebogen 97.1, der Fragebogen 97.2, der Variablenbogen 97.1 der Variablenbogen 97.2 und das Kodebuch der beiden Wellen bei.

6 Anhang

Verteilung ausgewählter Merkmale

Tabelle: Derzeitige Tätigkeiten nach abgeschlossener Ausbildung (in Prozent, Mehrfachnennungen)

Blfagrup	erwerbstätig	jobben	Volontariat	Ref. etc.	Praktikum	Ausb.	Studium	Promotion	Habil.	Suche	Erw.	Elternzeit	Hausmann	Umschulung	Fortbild.	arbeitslos	and.n.erw.	son.
1. Agrar, Ernaeh	89	0	0	1	0	0	1	0	0	4	9	5	0	1	1	0	0	
2. Arch., Raumpl	76	5	0	0	0	0	4	0	0	6	10	10	0	0	11	1	2	
3. Bauing., Verm	91	1	0	1	0	0	4	0	0	3	4	2	0	0	2	0	0	
4. E-Technik FH	97	0	0	1	0	0	3	1	0	4	2	0	0	1	2	0	0	
5. Masch.bau FH	96	0	0	0	0	0	4	1	0	4	3	1	0	1	1	0	1	
6. Wirt.-Ing. FH	90	0	0	0	0	0	2	0	0	3	5	0	0	0	3	0	0	
7. Informat. etc	98	0	0	0	0	0	1	5	0	1	0	1	0	0	1	0	0	
8. WiWi FH	88	1	0	0	0	0	2	1	0	3	11	2	0	3	3	0	1	
9. Sozialwesen F	79	1	0	1	0	0	8	0	0	10	12	11	0	10	3	1	3	
10. sonstiges FH	77	2	0	0	0	0	0	1	0	10	14	9	0	2	10	0	2	
11. Agrar, Ernae	73	3	0	3	0	0	8	10	0	8	12	2	0	1	4	4	1	
12. Arch., Raump	76	3	0	6	0	0	3	3	0	7	11	2	0	2	8	0	2	
13. Bauing., Ver	89	0	0	1	0	0	3	5	0	10	6	2	0	2	5	0	0	
14. E-Technik Un	98	0	0	0	0	0	3	9	0	3	1	0	0	0	1	1	0	
15. Masch.bau Un	94	1	0	0	0	0	5	15	0	3	2	1	0	1	1	0	0	
16. Wirt-Ing. Un	95	0	0	0	0	1	1	5	2	8	3	1	0	1	3	0	1	
17. Physik Uni	91	2	0	0	0	1	2	11	3	6	1	0	0	1	3	1	1	
18. Biologie Uni	79	1	0	3	0	0	3	17	3	7	16	8	0	3	3	3	2	
19. Chemie Uni	89	1	0	1	0	3	1	6	3	9	2	2	0	1	3	1	0	
20. Pharma, Chem	82	4	0	0	0	0	4	4	0	2	20	9	0	2	0	0	4	
21. Mathematik U	90	1	0	0	0	0	3	7	5	4	6	3	0	0	3	0	0	
22. Informatik U	92	1	0	0	0	0	0	8	2	4	7	2	0	1	1	1	0	
23. Humanmed. St	86	0	0	0	0	5	1	12	2	4	11	4	0	15	4	1	1	
24. Psychologie	86	1	0	2	2	2	1	10	8	6	11	4	0	18	1	1	0	
25. Pädagogik U	81	2	0	1	0	0	5	7	0	8	10	6	1	7	2	1	3	
26. Sprach., Kul	85	0	0	12	0	0	5	8	0	4	7	4	0	1	1	0	0	
27. Rechtwiss. S	88	1	1	2	0	0	4	10	0	7	5	1	0	2	4	0	1	
28. WiWi Uni	90	1	0	0	0	0	2	4	0	5	8	2	0	1	2	0	0	
29. La Primar.,	83	1	0	1	1	0	4	2	0	3	14	6	0	2	2	0	0	
30. LA Real., Se	88	2	0	0	0	0	1	1	0	4	8	1	1	0	1	0	0	
31. La Gym., Ber	86	1	0	3	0	0	5	8	2	3	10	3	0	1	2	1	0	
32. Magister	77	5	1	1	0	0	3	15	2	13	7	9	0	1	6	1	2	
33. sonstiges Un	83	2	0	1	1	0	3	12	0	8	8	5	0	2	7	1	2	
Total	87	1	0	1	0	0	3	6	1	6	8	4	0	3	3	1	1	

Fett markierte Werte stellen Abweichungen zum Originalfile dar.