

**Scientific-Use-File zur
Zeitverwendungserhebung 2012/13
- Anonymisierungskonzept -**

1 Einführung

Scientific-Use-Files sind standardisierte Einzeldatensätze, die von den Statistischen Ämtern für wissenschaftliche Vorhaben bereitgestellt werden. Im Gegensatz zu Gastwissenschaftlerarbeitsplätzen oder der kontrollierten Datenfernverarbeitung werden Scientific-Use-Files außerhalb der geschützten Räume der amtlichen Statistik genutzt.

§16 Abs. 6 Bundesstatistikgesetz (BStatG) ermöglicht den Statistischen Ämtern des Bundes und der Länder Einzelangaben an Hochschulen und andere Einrichtungen der unabhängigen wissenschaftlichen Forschung zu übermitteln. Voraussetzung hierfür ist jedoch, dass die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können. „Unverhältnismäßig“ bedeutet, dass der Aufwand für eine Reidentifikation höher ist, als der Nutzen, der daraus gezogen wird.

Scientific-Use-Files haben daher stets zwei sehr gegensätzliche Anforderungen zu erfüllen. Auf der einen Seite sollen sie für einen möglichst großen Teil an Forschungsvorhaben ein Höchstmaß an Analysepotenzial bieten, zum anderen müssen sie den strengen Maßgaben der faktischen Anonymität genügen.

Den Anforderungen des BStatG zur Sicherstellung der faktischen Anonymität wird bei einem Scientific-Use-File zum einen dadurch entsprochen, dass mit dem Datennutzer ein Vertrag geschlossen wird, in dem ein Deanonymisierungsverbot festgeschrieben ist und Sanktionen bei Nichteinhaltung dieses Verbotes festgesetzt werden. Zum anderen wird die faktische Anonymität durch datenverändernde Maßnahmen erreicht. Das vorliegende Anonymisierungskonzept beschreibt diese datenverändernden Maßnahmen. Im Focus der Anonymisierung stehen Merkmale, die im Zeitablauf relativ stabil oder anderen Menschen bekannt sind.

2 Datenmaterial

Grundlage für die Erstellung des Scientific Use Files ist die Zeitverwendungserhebung 2012/2013. In den Jahren 2012/2013 hat das Statistische Bundesamt in Zusammenarbeit mit den Statistischen Ämtern der Länder zum dritten Mal nach 1991/1992 und 2001/2002 die Erhebung zur Zeitverwendung privater Haushalte durchgeführt. Die gewonnenen Daten geben Aufschluss darüber, wie Menschen in den unterschiedlichsten Lebenslagen und Haushaltskonstellationen ihre Zeit für verschiedene Lebensbereiche einteilen. Sie liefern beispielsweise Informationen zu Erwerbsarbeit und unbezahlter Arbeit, zur Arbeitsteilung innerhalb von Familien, zu Kinderbetreuung, freiwilligem Engagement sowie zur Zeitverwendung von Kindern, Jugendlichen und älteren Menschen.

In einer schriftlichen Befragung auf freiwilliger Basis gaben etwa 5000 Privathaushalte mit über 11000 Personen zu ihrer Zeitverwendung Auskunft. Zur Ziehung der Stichprobe wurde ein

Quotenverfahren genutzt. Als Quotierungsmerkmale dienten hierbei Bundesland, Haushaltstyp sowie die soziale Stellung des Haupteinkommensbeziehers im Haushalt. Um saisonale Verzerrungen zu vermeiden, fand die Befragung der gezogenen Haushalte kontinuierlich verteilt über den Zeitraum August 2012 bis Juli 2013 statt. Die gewonnenen Daten wurden anhand des Mikrozensus 2012 auf die deutsche Bevölkerung hochgerechnet.

Neben einem Haushaltsfragebogen füllte jedes Haushaltsmitglied ab 10 Jahren einen Personenfragebogen sowie ein Tagebuch aus. Der Haushaltsfragebogen enthielt unter anderem Fragen zur Zusammensetzung des Haushalts, zu Wohnverhältnissen und der Inanspruchnahme von Betreuungsangeboten für Kinder unter 10 Jahren. Der Personenfragebogen deckt Angaben zu Alter und Familienstand, Erwerbstätigkeit, Einkommen, Schulbildung, Freizeitaktivitäten und freiwilligem Engagement der einzelnen Haushaltsmitglieder ab. Zudem beantworteten die Teilnehmenden Fragen zu ihrem subjektiven Zeitempfinden und zu Zeitstress in verschiedenen Lebensbereichen.

Im dritten Erhebungsteil hielten schließlich alle Personen ab 10 Jahren an drei vorgegebenen Tagen – zwei Wochentagen und einem Wochenendtag – ihren Tagesablauf fest. Dazu wurde ein Tagebuch verwendet, in dem die Tage bereits in einzelne Zeilen zu je 10 Minuten strukturiert waren. Die Befragten beschrieben hier in eigenen Worten, welcher Hauptaktivität sie jeweils nachgingen. Weiter gaben sie durch Ankreuzen vorgegebener Kategorien (Partner/Partnerin, Kinder unter 10 Jahren, andere Haushaltsmitglieder, andere bekannte Personen) an, welche Personen dabei beteiligt waren. Gegebenenfalls konnte zur gleichen Zeit eine Nebenaktivität eingetragen werden. Waren die Befragten gerade auf dem Weg von einem Ort zum anderen, war auch das genutzte Verkehrsmittel festzuhalten. Am Abschluss jedes Tages standen Fragen zur subjektiven Einschätzung des konkreten Tagesverlaufs. Die Befragten konnten hier angeben, welche Tätigkeiten ihnen die größte Freude und welche keine Freude gemacht hatten und wofür sie sich mehr Zeit gewünscht hätten.

Um eine statistische Auswertung der vielfältigen frei eingetragenen Tätigkeiten zu ermöglichen, wurden diese im Statistischen Bundesamt mit Hilfe eines einheitlichen Verzeichnisses von 165 verschiedenen Aktivitäten codiert. Dieses umfasst den persönlichen Bereich, Erwerbstätigkeit, Bildung, Haushaltsführung und Betreuung der Familie, ehrenamtliche und freiwillige Tätigkeiten, soziales Leben, Sport und Hobbys, Mediennutzung und Wegezeiten. Neben der Erfassung der Erhebungsunterlagen und Codierung der Aktivitäten wurden die Daten eingehend geprüft und plausibilisiert.

Das Aktivitätenverzeichnis wurde anhand der folgenden Aspekte erstellt. Zum einen sollten internationale Vergleiche mit Zeitverwendungserhebungen anderer europäischer Länder ermöglicht werden. Daher orientiert sich das Verzeichnis sehr an den aktuellen Empfehlungen des Statistischen Amtes der Europäischen Union, die in den „Harmonised European Time Use Surveys 2008 Guidelines“ festgehalten sind. Zum anderen sollte es so weit wie möglich kompatibel mit dem Verzeichnis der Zeitbudgeterhebung 2001/2002 sein, um Aussagen über die Veränderung der Ergebnisse über die Zeit treffen zu können. Weiter bildet die Liste Tätigkeiten, die in familien-, bildungs- und sozialpolitischer Hinsicht besonders interessant sind, möglichst detailliert ab. Schließlich enthält sie aufgrund des im Tagebuch vorgegebenen 10-Minuten-Taktes keine Aktivitäten, die grundsätzlich weniger als 10 Minuten andauern.¹

¹ Vgl. Maier, Lucia, 2014: Methodik und Durchführung der Zeitverwendungserhebung 2012/2013, In: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, Wiesbaden, S. 672ff.

3 Anonymisierung

Für den Scientific-Use-File der Zeitverwendungserhebung 2012/13 werden Merkmale bereitgestellt, die sich für die Durchführung wissenschaftlicher Vorhaben eignen. Die Anordnung der Datensätze des zu übermittelnden Datenbestands ist systemfrei. Direkte Identifikatoren und Hilfsmerkmale wie beispielsweise der Name und das Geburtsdatum sind im Scientific-Use-File nicht enthalten. Eine Übersicht über die im Scientific-Use-File enthaltenen Merkmale findet sich in den Datensatzbeschreibungen.

Analysen haben gezeigt, dass das Risiko der Reidentifizierung bei Haushalten mit besonders vielen Haushaltsmitgliedern erhöht ist. In Haushalten mit sieben und mehr Haushaltsmitgliedern werden deshalb ausgewählte Personen gelöscht. Bei der Auswahl der zu löschenden Personen wird darauf geachtet, dass der Haushaltstyp unverändert bleibt. Keinesfalls werden Personen gelöscht, die verheiratet sind und deren Ehepartner im Haushalt lebt, das jüngste Kind im Haushalt oder selbständige Personen. Die Datensätze werden aus der Personen-, Zeittakt- und Summendatendatei gelöscht. Die Anzahl der Haushaltsmitglieder und die Anzahl der Kinder unter 10 Jahren werden entsprechend angepasst. Des Weiteren werden Haushalte mit gleichgeschlechtlichen Paaren entfernt, um die faktische Anonymität der Daten sicherzustellen. Die entsprechenden Datensätze werden ebenfalls in der Haushalts-, Personen-, Zeittakt- und Summendatendatei gelöscht.

Da der Auswahlatz der Zeitverwendungserhebung mit ca. 0,01% der Haushalte in Deutschland vergleichsweise klein ist, sind die Datenbestände schon von vornherein sehr gut vor Reidentifikationsversuchen geschützt. Der Datenangreifer weiß auch bei einer eindeutigen Zuordnung nie mit Sicherheit, ob er den richtigen Haushalt zugeordnet hat, wenn die erhobenen Daten die Grundgesamtheit nicht vollständig erfassen. Die von der Stichprobeneigenschaft ausgehende Schutzwirkung besteht jedoch dann nicht mehr, wenn dem Angreifer bekannt ist, dass der von ihm gesuchte Haushalt an der amtlichen Erhebung teilgenommen hat. Das Risiko des spezifischen Wissens ist sehr klein, aber es kann nicht ausgeschlossen werden, dass die Befragten selbst Dritten mitteilen, dass sie an der Zeitverwendungserhebung teilgenommen haben. Dieses spezifische Risiko kann gezielt verringert werden, indem nur eine nach dem Zufallsprinzip gezogene Substichprobe des ursprünglichen Datenfiles weitergegeben wird.²

Die Beschränkung der Zeitverwendungsdaten auf eine 95%-Stichprobe mit 4775 Haushalten stellt damit eine weitere Anonymisierungsmaßnahme dar. Die Unterstichprobe der Haushalte wird zufällig gezogen, nachdem die nicht berücksichtigten Haushalte entfernt worden sind. Ein Datenangreifer verliert bei einer versuchten Identifikation durch die Stichprobe die Kenntnis, ob der gesuchte Haushalt in der Stichprobe enthalten ist. Ordnet er einen Datensatz aus dem Zusatzwissen einem Haushalt zu, so trägt er das Risiko, dass diese Zuordnung nur deshalb zustande kam, weil der richtige Haushalt nicht in der Stichprobe enthalten ist. Die Zuordnung ist für ihn wertlos. Die Stichprobenziehung ist ein entscheidender Beitrag zur Erreichung der faktischen Anonymität.

Im Folgenden werden die Anonymisierungsmaßnahmen beschrieben, die auf die einzelnen Merkmale der Haushalts-, Personen-, Zeittakt- und Summendatensätze angewendet werden.

² Vgl. Müller, Walter/Blien, Uwe/Knoche, Peter/Wirth, Heike u.a., 1991: Die faktische Anonymität von Mikrodaten. In: Statistisches Bundesamt (Hrsg.), Schriftreihe Forum der Bundesstatistik, Band 19. Stuttgart: Metzler-Poeschel, S. 400ff.

3.1 Haushaltsdatensätze

id_hhx Haushaltsnummer

Das Merkmal wird als fortlaufende Nummer der systemfreien Anordnung neu generiert.

hb9x Wohnfläche Wohnung

Die Wohnfläche der Wohnung wird bis 190 qm in 5-qm-Schritten gerundet, ab 190 qm wird in 10 qm-Schritten vergrößert. Wohnflächen von 200 qm bis unter 250 qm werden zu einer Klasse zusammengefasst und ab 250 qm Wohnfläche wird ein Top-Coding durchgeführt.

hb10x Anzahl der Räume

In Haushalten mit neun und mehr Wohn- und Schlafräumen wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Anzahl der Räume „9 und mehr“ beträgt.

hb121bx Auto Anzahl

In Haushalten mit vier und mehr Autos wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Anzahl der Autos „4 und mehr“ beträgt.

hb122bx Computer Anzahl

In Haushalten mit sieben und mehr Computern wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Anzahl der Computer „7 und mehr“ beträgt.

hc141a bis hc1411c Unterstützung des Haushalts durch andere Personen

Bei den Stundenangaben zu den einzelnen Unterstützungsleistungen wird jeweils ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Stundenzahl z.B. „60 und mehr“ beträgt. Daneben werden einige wenige Einzelfälle mit jeweils benachbarten Stundenangaben zusammengefasst.

hd15x Monatliches Haushaltsnettoeinkommen

Bei dem monatlichen Nettoeinkommen des Haushaltes werden die untersten fünf und die obersten drei Einkommensklassen zusammengefasst. In der Folge wird bei den betroffenen Fällen nur ausgewiesen, dass das monatliche Haushaltsnettoeinkommen „unter 900 Euro“ bzw. „7500 Euro und mehr“ beträgt.

Zusätzliche Maßnahmen bei den Haushaltsdatensätzen

Als regionale Information werden die Gebietsstände West (alte Bundesländer) und Ost (neue Bundesländer einschließlich Berlin) weitergegeben. Aufgrund des geringen Stichprobenumfangs der Zeitverwendungserhebung wird das Merkmal Bundesland gelöscht.

Auf Empfehlung der Mitglieder des Auswertungsbeirates der Zeitverwendungserhebung 2012/13 werden im Scientific-Use-File zusätzliche Informationen zur Regionaltypisierung aufgenommen, die der Wissenschaft siedlungsstrukturbezogene Auswertungsmöglichkeiten bieten. Im Hinblick auf die faktische Anonymität wurden im Vorfeld der siedlungsstrukturelle Regionstyp, die

Gemeindegrößenklasse, der siedlungsstrukturelle Kreistyp und der Stadt- und Gemeindetyp umfangreich in Kreuzkombination untereinander und zum Merkmal Gebiet geprüft.

Besonderes Gewicht kommt bei Deanonymisierungsversuchen dem Wissen über den Wohnort zu. Lokale Angaben und Strukturdaten könnten helfen, den Wohnort indirekt zu identifizieren. Die Grundregeln der faktischen Anonymität sind, dass keine einzelne Gemeinde unter 500000 Einwohnern und kein einzelner Gemeindeverband unter 400000 Einwohnern in der Grundgesamtheit zu identifizieren sein darf.³ Dabei ist für die Prüfung, ob die raumbezogenen Merkmale die Anforderungen der faktischen Anonymität erfüllen, nicht die Anzahl der Einheiten in der Stichprobe maßgeblich, sondern in der Grundgesamtheit. Eine Kombination aller erwähnten raumbezogenen Merkmale würde einzelne Gemeinden unterhalb der definierten Grenze identifizieren. Datenverändernde Maßnahmen könnten eingesetzt werden, um diese Grundregeln der faktischen Anonymität sicherzustellen, indem zum Beispiel regionale Angaben einer benachbarten Klasse zugeordnet werden und unter Umständen auch Klassen übersprungen werden. Die räumliche Abgrenzung ließe sich daran messen, dass diese mindestens zwei Gemeinden in der Grundgesamtheit zusammenfasst.

Angesichts des technischen Fortschrittes in der Datenverarbeitung und der leichten Zugänglichkeit von Zusatzwissen, wie den raumbezogenen Merkmalen, ist der Schutz um das Wissen des Wohnortes sehr sorgfältig zu überprüfen. Der Aufwand an Zeit, Kosten und Arbeitskraft, mit welchem Deanonymisierungsversuche gegenwärtig durchgeführt werden können, ist im Vergleich zu vor zwanzig Jahren extrem gesunken. Der Datenbestand der Zeitverwendungserhebung 2012/13 wurde einer Fallzahlenanalyse unterzogen. Aufgrund des geringen Stichprobenumfangs führt die Kombination der raumbezogenen Merkmale vielfach zu nur wenigen Haushalten. Das räumliche Abgrenzungskriterium von mindestens 2 Gemeinden in der Grundgesamtheit würde einen unzureichenden Schutz bieten, wenn die einzelnen Gemeinden klein sind und eine geringe Einwohnerzahl aufweisen. Durch derart vielfältige raumbezogenen Merkmale würde sich für einen Datenangreifer die Wahrscheinlichkeit enorm erhöhen, einen Haushalt, der an der Erhebung teilgenommen hat, aufzufinden und auch die Sicherheit, dass es sich um den gesuchten Haushalt handelt. Umgekehrt haben zu grobe raumbezogene Merkmale, die hinsichtlich des Datenschutzes unbedenklich sind, oft keinen Aussagewert mehr. Im Ergebnis werden im Scientific-Use-File nur die Gemeindegrößenklasse und der siedlungsstrukturelle Kreistyp weitergegeben.

kreistypx Kreistyp

Aus der Kombination Gebietsstand „West“, der Gemeindegrößenklasse „500.000 und mehr“ und dem Kreistyp „städtischer Kreis“ ließen sich Haushalte eines Bundeslands identifizieren. Bei dem Merkmal Kreistyp wird deshalb eine Zufallsüberlagerung vorgenommen, die die Unsicherheit eines Datenangreifers über die Richtigkeit von Zuordnungsversuchen erhöht.

3.2 Personendatensätze

alterx Alter

Das Alter von Personen ab 80 bis unter 85 Jahren wird zu einer Altersklasse zusammengefasst und bei Personen, die 85 Jahre und älter sind, wird ein Top-Coding durchgeführt. In der Folge wird

^{3 3} Vgl. Müller, Walter u.a., 1991, S. 443ff.

bei den betreffenden Altersgruppen nur ausgewiesen, dass diese „80 bis unter 85 Jahre“ bzw. „85 Jahre und älter“ sind.

pa2x Familienstand

Die Familienstände „in eingetragener Lebenspartnerschaft getrennt lebend“ bzw. „eingetragener Lebenspartner verstorben“ bzw. „eingetragene Lebenspartnerschaft aufgehoben“ werden in „verheiratet und dauernd getrennt lebend“ bzw. „verwitwet“ bzw. „geschieden“ umcodiert.

ha6x Stellung im Haushalt

Der Einzelwert „Großvater/Großmutter“ des Merkmals Stellung im Haushalt wird in „anders verwandt/verschwägert“ geändert.

he182x Betreute Zeit durch Tagesmutter/Tagesvater

Das Merkmal betreute Zeit durch Tagesmutter/Tagesvater wird bis 30 Stunden mit einer Klassenbreite von fünf Stunden klassifiziert. Außerdem werden bei dem Merkmal Einzelwerte umcodiert und ab 36 Stunden wird ein Top-Coding durchgeführt.

he183x Betreute Zeit durch Krippe

Das Merkmal betreute Zeit durch Krippe wird mit einer Klassenbreite von fünf Stunden klassifiziert und ab 41 Stunden wird ein Top-Coding durchgeführt.

he184x Betreute Zeit durch Kindertagesstätte

Das Merkmal betreute Zeit durch Kindertagesstätte wird mit einer Klassenbreite von fünf Stunden klassifiziert und ab 46 Stunden wird ein Top-Coding durchgeführt.

he185x Betreute Zeit durch vorschulische Einrichtungen

Bei dem Merkmal betreute Zeit durch vorschulische Einrichtungen werden Einzelwerte verändert und die Klassen „1-10 Stunden“, „21-25 Stunden“ und „31-40 Stunden“ gebildet.

he186x Betreute Zeit durch Hort

Das Merkmal betreute Zeit durch Hort wird mit einer Klassenbreite von fünf Stunden klassifiziert und ab 36 Stunden wird ein Top-Coding durchgeführt.

he18_kibetrx Kinderbetreuung

Das Merkmal Kinderbetreuung wird im mittleren Wertebereich mit einer Klassenbreite von fünf Stunden und im oberen und unteren Wertebereich mit einer Klassenbreite von zehn Stunden klassifiziert. Ab 41 Stunden wird ein Top-Coding durchgeführt. Das Merkmal Kinderbetreuung fasst die Merkmale p_ef13 bis p_ef17 zusammen. Die vorangehend beschriebenen datenbearbeiteten Maßnahmen der Merkmale p_ef13 bis p_ef17 werden bei der Anonymisierung des Merkmals p_ef18 Kinderbetreuung berücksichtigt.

he187x Betreute Zeit durch Verwandte, Freunde, Nachbarn

Bei der betreuten Zeit durch Verwandte, Freunde, Nachbarn von 26 Stunden und mehr wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass

die betreute Zeit durch Verwandte, Freunde, Nachbarn „26 Stunden und mehr“ beträgt. Des Weiteren werden bei dem Merkmal Einzelwerte verändert.

he19x Betreuung Tage pro Woche

Bei der Betreuung eines Kindes unter zehn Jahren von sechs und sieben Tagen pro Woche wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Betreuung „6 Tage und mehr“ beträgt.

he202x Unterrichtsstunden pro Woche (HFB)

Das Merkmal Unterrichtsstunden pro Woche von Kindern unter zehn Jahren wird im oberen und unteren Wertebereich mit einer Klassenbreite von fünf Stunden klassifiziert. Als Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Unterrichtsstunden pro Woche „15 - 20 Stunden“, „31 - 35 Stunden“ bzw. „36 - 40 Stunden“ betragen.

he212x AG-Stunden pro Woche (HFB)

Bei Kindern unter zehn Jahren mit sechs und mehr AG-Stunden pro Woche wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die AG-Stunden pro Woche „6 Stunden und mehr“ betragen.

he22x Zeitstunden insgesamt in der Schule pro Woche (HFB)

Bei den wöchentlichen Zeitstunden in der Schule von Kindern unter zehn Jahren werden im oberen Wertebereich Klassen gebildet. Als Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Zeitstunden in der Schule pro Woche „41 - 44 Stunden“ bzw. „45 - 50 Stunden“ betragen. Außerdem werden bei dem Merkmal Einzelwerte verändert.

he241x Stunden pro Woche: Sport und Bewegung außerhalb der Betreuungseinrichtung/Schule

Bei Kindern unter zehn Jahren, die wöchentlich sieben und mehr Stunden außerschulische Freizeitangebote im Bereich Sport und Bewegung nutzen, wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass diese „7 Stunden und mehr“ pro Woche Sport- und Bewegungsangebote außerhalb der Betreuungseinrichtung/Schule nutzen.

he24_sprachfx Stunden pro Woche: Sprachförderung (Deutsch, andere Sprache) außerhalb der Betreuungseinrichtung/Schule

Aufgrund der geringen Fallzahl von Kindern unter zehn Jahren, die Sprachförderung in Deutsch oder Sprachförderung in einer anderen Sprache außerhalb der Betreuungseinrichtung/Schule erhalten, und dem damit verbundenen Reidentifikationsrisikos werden die zwei Variablen zu einer Variablen Sprachförderung (Deutsch, andere Sprache) außerhalb der Betreuungseinrichtung/Schule zusammengefasst.

Bei Kindern unter zehn Jahren mit drei und mehr Stunden außerschulische Sprachförderung wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Sprachförderung außerhalb der Betreuungseinrichtung/Schule „3 Stunden und mehr“ beträgt.

he245x Stunden pro Woche: Singen, Musikinstrument spielen außerhalb der Betreuungseinrichtung/Schule

Bei Kindern unter zehn Jahren, die wöchentlich vier und mehr Stunden außerschulische Freizeitangebote im Bereich Singen und Musikinstrument spielen nutzen, wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass diese „4 Stunden und mehr“ pro Woche Angebote im Bereich Singen und Musikinstrument spielen außerhalb der Betreuungseinrichtung/Schule nutzen.

he246x Stunden pro Woche: Malen, Basteln außerhalb der Betreuungseinrichtung/Schule

Bei Kindern unter zehn Jahren, die wöchentlich fünf und mehr Stunden außerschulische Freizeitangebote im Bereich Malen, Basteln nutzen, wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass diese „5 Stunden und mehr“ pro Woche Angebote im Bereich Malen, Basteln außerhalb der Betreuungseinrichtung/Schule nutzen.

he247x Stunden pro Woche: Tanzen, Theater spielen außerhalb der Betreuungseinrichtung/Schule

Bei Kindern unter zehn Jahren, die wöchentlich drei und mehr Stunden außerschulische Freizeitangebote im Bereich Tanzen, Theater spielen nutzen, wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass diese „3 Stunden und mehr“ pro Woche Angebote im Bereich Tanzen, Theater spielen außerhalb der Betreuungseinrichtung/Schule nutzen.

pc5x Berufliche Stellung in der Haupterwerbstätigkeit

Gering besetzten Merkmalsausprägungen der beruflichen Stellung werden zusammengefasst. Mithelfende Familienangehörige werden mit den Selbstständigen, Landwirten gruppiert, der freiwillige Wehrdienst-/Bundesfreiwilligendienst und das freiwillige Jahr werden mit den Angestellten zusammengefasst.

pc8x Wochenstunden in der Haupterwerbstätigkeit

und

pc9x Wunscharbeitszeit in der Haupterwerbstätigkeit

Bei der Wochenarbeitszeit und der Wunscharbeitszeit in der Haupterwerbstätigkeit wird bei 70 Stunden und mehr ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Wochenstunden in der Haupterwerbstätigkeit bzw. die Wunscharbeitszeit in der Haupterwerbstätigkeit „70 Stunden oder mehr“ betragen. Des Weiteren werden bei den beiden Merkmalen Einzelwerte derart umcodiert, dass sich die Relation zwischen der Wochen- und Wunscharbeitszeit nicht ändert. Ein Beispiel: Arbeitet eine Person 47 Stunden pro Woche und gibt die Person als Wunscharbeitszeit auch 47 Stunden an und ist dieser Wert bei dem letztgenannten Merkmal ein Einzelwert, werden beide Merkmale entsprechend geändert.

pc11x Wirtschaftszweig

Der Einzelwert der Merkmalsausprägung „Konsulat, Botschaft, inter- und supranationale Organisation“ wird in einen anderen Wirtschaftszweig umcodiert.

pc122x Kilometer zur Arbeitsstätte in der Haupterwerbstätigkeit

Der Arbeitsweg in Kilometern wird ab einer Entfernung von mehr als 30 km in 10-km-Schritten gerundet, Entfernungen von 100 bis unter 150 km und 150 bis unter 200 km werden zu Klassen zusammengefasst. Ferner wird ab einer Entfernung von 200 km ein Top-Coding durchgeführt.

pc13x Arbeitsweg in Minuten in der Haupterwerbstätigkeit

Der Arbeitsweg in Minuten wird bis 80 Minuten in 5-Minuten-Schritten gerundet, bei einem Arbeitsweg bis 120 Minuten wird in 10-Minuten-Schritten vergrößert. Ab 120 Minuten wird ein Top-Coding durchgeführt.

pd192x Anzahl der Nebenerwerbstätigkeiten

Bei Erwerbstätigen mit zwei und mehr Nebenerwerbstätigkeiten wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Anzahl der Nebenerwerbstätigkeiten „2 oder mehr“ beträgt.

pd21x Tage pro Monat, an denen Nebenerwerbstätigkeit ausgeübt wird

und

pd22x Stunden pro Tag, an denen Nebenerwerbstätigkeit ausgeübt wird

Bei diesen beiden Merkmalen werden Einzelwerte umcodiert.

pe23x Monatliches Nettoeinkommen aus Haupterwerbstätigkeit und Nebenerwerbstätigkeit(en)

Bei dem monatlichen Nettoeinkommen aus Erwerbstätigkeit werden die untersten zwei Einkommensklassen und im oberen Bereich die obersten neun Einkommensklassen zu zwei Einkommensklassen zusammengefasst. In der Folge wird bei den betroffenen Fällen nur ausgewiesen, dass das monatliche Nettoeinkommen aus Haupt- und Nebenerwerb „unter 300 Euro“, „4000 bis unter 5000 Euro“ bzw. „5000 Euro und mehr“ beträgt.

pf25x Art der Schule

Gering besetzten Merkmalsausprägungen der Schulart werden zusammengefasst. Die berufliche Schule, die zur mittleren Reife führt, wird mit dem Berufsvorbereitungsjahr, Berufsgrundbildungsjahr zusammengefasst, die Schule des Gesundheitswesens mit der Fachschule, Fachakademie.

pf26x Unterrichtsstunden pro Woche (PFB)

Bei Schülern, die unter acht Unterrichtsstunden pro Woche haben, wird ein Bottom-Coding durchgeführt, der obere Wertebereich von 41 bis 45 Unterrichtsstunden wird zu einer Klasse zusammengefasst. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Unterrichtsstunden pro Woche „unter 8 Stunden“ bzw. „41 bis 45 Stunden“ betragen. Ferner werden bei dem Merkmal Einzelwerte umcodiert.

pf272x AG-Stunden pro Woche (PFB)

Bei Schülern mit sieben und mehr AG-Stunden pro Woche wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die AG-Stunden pro Woche „7 Stunden und mehr“ betragen.

pf28x Zeitstunden insgesamt in der Schule pro Woche (PFB)

Bei Schülern, die wöchentlich weniger als acht Zeitstunden in der Schule verbringen, wird ein Bottom-Coding, vorgenommen. Bei Schülern, die 45 Zeitstunden oder mehr in der Schule pro Woche verbringen, wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Personen nur ausgewiesen, dass diese wöchentlich „unter 8 Stunden“ oder „45 Stunden und mehr“ in der Schule verbringen. Des Weiteren werden bei dem Merkmal Einzelwerte umcodiert.

pf292x Sprachförderung in Deutsch außerhalb der Schule

und

pf293x Sprachförderung andere Sprache außerhalb der Schule

Bei Personen mit drei und mehr Stunden Sprachförderung in Deutsch bzw. in einer anderen Sprache wird ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Sprachförderung in Deutsch bzw. in einer anderen Sprache außerhalb der Schule „3 Stunden und mehr“ beträgt.

pf294x Nachhilfe, Förderkurse außerhalb der Schule

Bei Schülern mit fünf und mehr Stunden Nachhilfe, Förderkurse wird ebenfalls ein Top-Coding durchgeführt. In der Folge wird bei den betreffenden Fällen nur ausgewiesen, dass die Nachhilfe, Förderkurse außerhalb der Schule „5 Stunden und mehr“ beträgt.

pf31x Höchster beruflicher Ausbildungs- oder Hochschulabschluss

Die gering besetzten Merkmalsausprägungen Berufsvorbereitungsjahr und Fachakademie (nur Bayern) werden mit anderen Ausprägungen des höchsten beruflichen Ausbildungs- oder Hochschulabschlusses zusammengefasst. Das Berufsvorbereitungsjahr wird mit der Anlernausbildung und dem beruflichen Praktikum zusammengefasst, die Fachakademie (nur Bayern) mit dem Meister, Techniker oder gleichwertigen Fachschulabschluss.

Zusätzliche Maßnahmen bei den Personendatensätzen

Bei einigen Datensätzen müssen zusätzliche Maßnahmen durchgeführt werden. Durch die Kombination ausgewählter Merkmale kann ein erhöhtes Risiko für die Identifikation eines Haushaltes ausgehen. Beispiele hierfür sind z.B. eine große Anzahl an Haushaltsmitgliedern und das Alter der Haushaltsmitglieder, Haushalte mit auffällig älteren Kindern, Haushalte mit besonders vielen Kindern unter 10 Jahren oder Haushalte, in denen nur Großeltern mit Enkeln leben. Betrifft die Merkmalskonstellationen nur wenige Haushalte, werden die Datensätze geringfügig verändert.

Bei besonders alten Menschen, die in großen Haushalten leben, besteht beispielsweise ein erhöhtes Reidentifikationsrisiko. Um die faktische Anonymität der Daten zu erreichen, wird in diesen Haushalten das Alter dieser Personen durch ein zufällig gewähltes Alter innerhalb einer Altersklasse ersetzt. Auch wird das Alter bei Familien mit auffällig älteren Kindern modifiziert. Leben viele Kinder unter 10 Jahren im Haushalt, wird das Alter mehrerer Personen im Haushalt geändert. Außerdem wird das Alter von Personen in besonders großen Haushalten, in denen ein Haushaltsmitglied Leistungen aus der Pflegeversicherung erhält, verändert. In markanten

Haushalten mit Enkeln werden ebenfalls geringfügige datenverändernde Maßnahmen durchgeführt.

3.3 Zeittakt- und Summendatensätze

tag_folgtx Datum des Tagebuchtages der Person folgt dem vorherigen Tagebuchtag der Person

In wenigen Fällen wird das Merkmal Datum des Tagebuchtages der Person folgt dem vorherigen Tagebuchtag der Person verändert, da Kombinationen mit anderen Merkmalen regional unterschiedliche Feiertage und damit die Bundeslandzugehörigkeit eines Haushaltes offenlegen könnten.

Zusätzliche Maßnahmen bei den Zeittakt- und Summendatensätzen

Das Merkmal Ausfülldatum wird gelöscht. Anstelle wird der Kalendermonat, das Quartal und das Jahr des Ausfülldatums im Scientific-Use-File weitergegeben.

4 Hinweise zum Vergleich mit früheren Jahren⁴

Das Stichprobenkonzept und Erhebungskonzept der Erhebung 2012/2013 wurde für zeitliche Vergleiche weitestgehend auf die Erhebung 2001/2002 abgestellt. Die nachfolgend beschriebene Unterschiede beruhen auf geänderten Empfehlungen zur Harmonisierung der europäischen Zeitverwendungserhebungen (HETUS), vor allem aber auf anderen Schwerpunkten in familien-, bildungs- und sozialpolitischen Fragestellungen.

Unterschiedliche Erhebungskonzepte gibt es bei der Selbsteinschätzung der sozialen Stellung der Befragten. Während 2012/2013 den Nichterwerbstätigen über die Selbsteinschätzung die soziale Stellung in der gewünschten Untergliederung eindeutig zugeordnet werden kann, ist das in der Erhebung 2001/2002 nicht möglich, da personenbezogene Angaben zur Selbsteinschätzung der sozialen Stellung nicht erfasst wurden. Entsprechende Abgrenzungen lassen sich zwar aus den Erhebungsmerkmalen der Zeitverwendungserhebung 2001/2002 konstituieren, sind aber mit 2012/2013 nicht vergleichbar.

Die Erwerbsbeteiligung ist in beiden Erhebungen an das Erwerbskonzept des Mikrozensus angelehnt. Erwerbstätig sind Personen ab 15 Jahren, deren Erwerbstätigkeit normalerweise mindestens eine Stunde pro Woche beträgt. Methodische Unterschiede gibt es aber bei der Abfrage zur Erwerbstätigkeit. In der Erhebung 2001/2002 wurden beispielweise für geringfügig Erwerbstätige eine Vielzahl an Merkmalen zur Erwerbstätigkeit nicht erhoben, die in der Zeitverwendungserhebung 2012/2013 infolge einer anderen Abgrenzung von Haupt- und Nebentätigkeit vorliegen.

Der Haushaltsfragebogen 2001/2002 enthält eine Beziehungsmatrix, mit deren Hilfe sich alle Beziehungen der Haushaltsmitglieder untereinander bestimmen lassen. In schriftlichen Befragungen bereitete den Auskunftsgibenden das Ausfüllen dieser komplexen Beziehungsmatrix große Schwierigkeiten. In der Erhebung 2012/2013 wurde deshalb nur die Beziehung zum Haupteinkommensbezieher erfragt.

⁴ In diesem Abschnitt werden Hinweise zur Vergleichbarkeit der Scientific-Use-Files von 2012/2013 und 2001/2002 beschrieben. Zur Vergleichbarkeit der Zeitverwendungsdaten von 2001/2002 und 1991/1992 siehe Holz, Erlend, 2005: Kleines Vademekum zu den Scientific Use File des Zeitbudgeterhebung (ZBE) 2001/02 und 1991/92, Statistisches Bundesamt, Bonn, S. 5ff.

Die Variablen für Kinder unter 10 Jahren sind in der Zeitverwendungserhebung 2012/2013 mit „trifft nicht zu“ und die Variablen der Personen ohne ausgefüllten Fragebogen mit „keine Angabe“ codiert. Im Scientific-Use-File 2001/2002 sind die Variablen für Kinder unter 10 Jahren mit „Kind unter 10 Jahren“ codiert.

Die Datensätze der Haushalts- und Personendatei enthalten zusätzliche Standardtypisierungen, die für einheitliche Definitionen für Haushalts- und Personentypen über verschiedene Themengebiete hinweg verwendet werden können. Aufgrund der verschiedenen Erhebungskonzepte sind die Typisierungen 2001/2002 und 2012/2013 nur eingeschränkt miteinander vergleichbar. Neben den bisher erörterten Unterschieden bei der Erwerbstätigkeit und dem sozialen Status wurde die Haushaltstypisierung anders abgegrenzt. In 2001/2002 werden als Kinder ledige Personen ohne Altersbegrenzung betrachtet, die Kind der Bezugsperson des Haushaltes und/oder des Partners sind. Bei den Standardtypisierungen zu den Haushaltstypen 2012/2013 gibt es hingegen eine Altersbegrenzung. In einem Alleinerziehenden-Haushalt oder Paar-Haushalt mit Kind gibt es mindestens ein Kind/Enkelkind unter 18 Jahren, weitere Kinder sind ledig und unter 27 Jahren. Einbezogen sind dabei grundsätzlich alle Paare, unabhängig davon, ob diese verheiratet sind oder nicht.

Anders als bei der Zeitbudgeterhebung 2001/2002 wird bei der Verschlüsselung zwischen Tätigkeiten im Kontext von Schule und Hochschule getrennt. Der schulische Bereich wurde 2012/2013 um einzelne Codes für verschiedene Unterrichtsfächer und Arbeitsgemeinschaften ergänzt. Dafür wurden verschiedene Formen der Qualifikation und Weiterbildung außerhalb der Arbeitszeit in einem Code zusammengefasst. Die Kategorien „Haushaltsführung und Betreuung der Familie“, „Ehrenamtliche Tätigkeit/Freiwilligenarbeit/Unterstützung für andere Haushalte/-Teilnahme an Versammlungen“ und „Sport/Hobbys/Spiele“ wurden im Vergleich zur letzten Erhebung ebenfalls etwas gestrafft. Insgesamt bedeutet dies eine Reduktion von vormals 272 auf 165 dreistellige Aktivitätscodes. Für die Zeitvergleiche der Aktivitäten wird ein Umsteigeschlüssel für die Aktivitätscodes zur Verfügung gestellt.

Ein weiterer konzeptioneller Unterschied betrifft die Angabe, mit wem die Zeit verbracht wurde. Im Gegensatz zu 2001/2002 beschränkt sich in der aktuellen Erhebung die Information, dass ein Kind unter 10 Jahren mit dabei war, auf Kinder, die im selben Haushalt wohnen. 2001/2002 war es gleichgültig, ob die Kinder unter 10 Jahren in dem Haushalt der Person, die das Tagebuch ausfüllt, leben oder nicht. Entsprechendes gilt für den Ehe-/Lebenspartner. In der Zeitverwendungserhebung 2012/2013 wurden Partner und Kinder, mit denen die Person nicht zusammen wohnt, unter andere bekannte Personen eingetragen. Die Variable zum Ort der Aktivität, die 2001/2002 aus den Angaben zur Aktivität in Verbindung mit den Wegezeiten angelegt wurde, steht in der Erhebung 2012/2013 nicht zur Verfügung.

5 Fazit

Mit den faktisch anonymisierten Daten der Zeitverwendungserhebung 2012/2013 ist es gelungen, die Reihe der faktisch anonymisierten Daten aus der Zeitbudgeterhebung 1991/1992 und 2001/2002, die der Wissenschaft zu Analyse Zwecken zur Verfügung gestellt werden können, fortzusetzen. Auch wenn bei der Anonymisierung größtmöglicher Wert auf den Erhalt des Analysepotenzials gelegt wurde, sind nicht alle Fragestellungen der Wissenschaft exakt mit den Daten analysierbar. Für diese Fälle sei auf die alternativen Zugangswege zu Mikrodaten, die von den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder angeboten werden, verwiesen.

Literatur

Holz, Erlend, 2005: Kleines Vademekum zu den Scientific Use File der Zeitbudgeterhebung (ZBE) 2001/02 und 1991/92, Statistisches Bundesamt, Bonn

Maier, Lucia, 2014: Methodik und Durchführung der Zeitverwendungserhebung 2012/2013, In: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, Wiesbaden

Müller, Walter/Blien, Uwe/Knoche, Peter/Wirth, Heike u.a., 1991: Die faktische Anonymität von Mikrodaten. In: Statistisches Bundesamt (Hrsg.), Schriftreihe Forum der Bundesstatistik, Band 19. Stuttgart: Metzler-Poeschel