

gesis

Leibniz-Institut
für Sozialwissenschaften

Servicedokument

ALLBUS

Allgemeine Bevölkerungsumfrage
der Sozialwissenschaften

Variablennamen in ALLBUS

*Die Namenssystematik für Variablen ab
ALLBUS 2016*

Pascal Siegers

GESIS Datenarchiv für Sozialwissenschaften

GESIS - Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8
50667 Köln

Telefon: +49/(0)221/47694-0
Fax: +49/(0)221/47694-199
info@gesis.org, www.gesis.org

Inhalt

1	Variablennamen im ALLBUS-Programm: Status quo und Änderungsbedarf.....	2
2	Die neue Namenssystematik für den ALLBUS	4
3	Umsetzung des neuen Konzeptes.....	15

1 Variablennamen im ALLBUS-Programm: Status quo und Änderungsbedarf

Seit der Publikation der ersten ALLBUS Erhebung im Februar 1980 wurden in den Datensätzen fortlaufende V-Nummern als Variablennamen vergeben. Die erste Variable hieß in den Datensätzen entsprechend v1 und alle folgenden Variablen wurden aufsteigend benannt. Diese Namenskonvention geht auf Archivstandards der 1980er Jahre zurück. Der Vorteil der aufsteigenden V-Nummern ist die einfache Orientierung innerhalb der einzelnen Querschnittsdatsätze und die direkte Entsprechung der Fragebogendokumentation und des Datensatzes, weil die Variablen in der gleichen Reihenfolge im Datensatz erscheinen, wie in der Fragebogendokumentation bzw. im Variable Report der Kumulationen.

Ein Nachteil des bestehenden Standards ergibt sich, wenn Nutzer Auswertungen für unterschiedliche Erhebungszeitpunkte durchführen wollen, weil inhaltlich identische Variablen in verschiedenen Erhebungswellen und in verschiedenen Versionen der ALLBUS-Kumulation unterschiedliche Namen haben. Die Nutzer müssen ihre Do-Files oder Syntaxen für jeden Datensatz neu anpassen. Zudem entsteht bei der Fortschreibung der ALLBUS-Kumulation erhöhter Aufwand bei der Dokumentation der Variablen.

Gerade die Kumulationen des ALLBUS beinhalten mittlerweile eine große Zahl an Variablen. Da die Variablennamen keine Information über die Struktur des Datensatzes bieten, ist der Datensatz unübersichtlich und die über 3500 Seiten umfassende Dokumentation als Orientierungshilfe nur eingeschränkt verwendbar.

Aus diesem Grund wird mit dem Querschnitt 2016 eine neue Systematik zur Vergabe von Variablennamen im ALLBUS eingeführt.

1.1 Etablierte Verfahren zur Benennung von Variablen in längsschnittlichen Umfrageprogrammen

Bei der Benennung von Variablen haben sich unterschiedliche Verfahren etabliert, die als Vorbilder dienen könnten. In vielen Umfrageprogrammen (zum Beispiel beim General Social Survey oder dem European Social Survey) werden sogenannte sprechende oder mnemotechnische Variablennamen verwendet. Sprechende Variablennamen bestehen aus Buchstabenkürzeln, die die Inhalte der Variable andeuten sollen. Den Nutzern soll es dadurch leichter fallen, sich die Variablennamen zu merken. Allerdings lassen sich die mnemotechnisch gebildeten Namen oft nicht einfach den Inhalten der Variablen zuordnen. Vor allem dann, wenn viele ähnliche Variablen verwendet werden (z.B. in langen Fragebatterien zu Werten oder Persönlichkeit), kommt das mnemotechnische Verfahren an seine Grenzen. Bezogen auf den ALLBUS liegt die große Herausforderung des mnemotechnischen Verfahrens darin, die stark ausdifferenzierte Demographie sinnvoll abzubilden. Ein weiterer (potentieller) Nachteil der sprechenden Variablennamen ist, dass die Dokumentation von ALLBUS zweisprachig erfolgt, damit auch internationale Nutzer die Daten verwenden können. Da die Variablennamen in den beiden Sprachversionen des Datensatzes jedoch identisch sein sollen, wird die Verwendung von sprechenden Variablennamen aus deutschen Begriffen für die internationalen Nutzer zu einem Hindernis. Sprechende Variablennamen beruhen deshalb auch in nationalen Studien in der Regel auf englischen Begriffen.

Komplexere Systematiken für die Benennung von Variablen werden im Bereich der sozialwissenschaftlichen Panels (SOEP, NEPS, PAIRFAM, SHARE, GESIS Panel) verwendet. Die Aufschlüsselung der Variablennamen erlaubt in den Panels Rückschlüsse auf Erhebungszeitpunkte und oft auch thematische Module. Hinweise auf den Erhebungszeitpunkt sind für die ALLBUS Daten nicht notwendig, weil

die Daten keine Panelstruktur haben. Was in den Systematiken der Panelerhebungen interessant ist, sind systematische Kennzeichnungen der Erhebungseinheit. Diese zeigen an, ob sich eine Variable auf einen Befragten, den Haushalt oder andere Personen wie z.B. Partner, Eltern oder Kinder beziehen. Diese Systematik ist auch für die ALLBUS Daten relevant, weil in allen Erhebungen detaillierte Informationen über die Familie und den Haushalt der Studienteilnehmer erhoben werden.

2 Die neue Namenssystematik für den ALLBUS

Die Vergabe der neuen Variablennamen erfolgt anhand von sieben Prinzipien.

1. Es werden feste Variablennamen vergeben.

Die neue Namenskonvention sieht feste Variablennamen für Replikationsfragen vor. Jede im ALLBUS verwendete Frage bekommt einen Namen zugewiesen. Dieser Name wird dann in allen Datensätzen identisch verwendet, wenn die Frage identisch (oder annähernd identisch) repliziert wird.

Dadurch soll sowohl die Nutzung als auch die Dokumentation der Daten vereinfacht werden. Für Nutzer verringert sich der Aufwand bei der Anpassung von Syntaxen und Do-Files an neue Erhebungszeitpunkte. In der Dokumentation der Datensätze müssen Verweise zwischen verschiedenen Variablen innerhalb des Datensatzes (zum Beispiel die abgeleiteten Variablen zu Bildungsabschlüssen und Erwerbsstatus) nicht mehr für jede neue Erhebungswelle aktualisiert und geprüft werden.

2. Die Länge der Variablennamen wird auf 10 Zeichen begrenzt.

Obwohl Beschränkungen der Variablennamen in den gängigen Statistikpaketen nicht mehr notwendig sind, erleichtert die Begrenzung die Konvertierung zwischen verschiedenen Datenformaten und damit auch die Langzeitarchivierung der Daten. Darüber hinaus vereinfachen kurze Variablennamen die Datennutzung, weil sie weniger fehleranfällig bei der Eingabe sind. Daher werden alle Variablennamen auf 10 Zeichen begrenzt.

3. Die Variablennamen werden entsprechend der thematischen Struktur des Datensatzes vergeben.

Alle ALLBUS Erhebungen bilden bestimmte thematische Schwerpunkte ab, die von einer ausführlichen Erhebung der Demographie des Haushaltes komplementiert werden. Deshalb spricht vieles dafür, die Namen der Variablen der thematischen Struktur des Datensatzes entsprechend zu vergeben. Variablen des gleichen inhaltlichen Schwerpunktes haben eine hohe Wahrscheinlichkeit, in einer Wiederholungsbefragung gemeinsam repliziert zu werden. So bilden die Variablennamen auch die Struktur der Erhebung ab.

4. Vergabe vierstelliger Stammnamen für die Variablen

In Zukunft wird für fast alle Variablen im ALLBUS ein vierstelliger Stammname vergeben, der in allen Querschnittsdatsätzen und in den Kumulationen verwendet wird. Davon ausgenommen ist nur eine kleine Zahl an technischen und inhaltlichen Variablen, für die sprechende Variablennamen verwendet werden (siehe Tabelle 3). Die ersten beiden Stellen bestehen aus Buchstaben, die die inhaltliche Kategorie und die Unterkategorie abbilden der eine Variable zugeordnet wird. Die dritte und vierte Stelle werden als Kennziffer für die Nummer der Variable innerhalb der inhaltlichen Unterkategorie aufsteigend verwendet.

Der erste Buchstabe des Variablennamens zeigt das Thema an, dem eine Frage zugeordnet wird. Insgesamt werden 16 Oberthemen definiert, die in etwa den Schwerpunktmodulen zuzüglich der Demographie entsprechen (siehe Tabelle 1). Variablen zur Bildungs- und Erwerbsbiographie werden der Demographie zugeordnet.

Die gewählten Buchstaben orientieren sich an den Themen der Kategorien. Die Anlehnung an die Themen wiederum erfolgt in englischer Sprache, damit die Variablennamen für deutsche und internatio-

nale Nutzer gleichermaßen eingängig sind. Werden neue thematische Module für den ALLBUS definiert, kann die Zahl der thematischen Kategorien erweitert werden.

Tabelle 1: Übersicht über inhaltliche Kategorien und ihre Buchstabenkennungen

Nr.	Kategorie	Buchstabenkennung
1	Abweichendes Verhalten und Sanktion	c (crime)
2	Arbeitsorientierungen	j (job)
3	Demographie	d (demography)
4	Ethnozentrismus und Minoritäten	m (minorities/migration)
5	Familie	f (family)
6	Gesundheit	h (health)
7	Wertorientierungen	v (values)
8	Lebensstil und Persönlichkeit	l (lifestyle)
9	Politik	p (politics)
10	Religion und Weltanschauung	r (religion)
11	Soziale Netzwerke und soziales Kapital	s (social)
12	Soziale Ungleichheit	i (inequality)
13	Wohnen und Umwelt	a (accomodation)
14	Wirtschaft (Einstellungen/Wahrnehmungen)	e (economy)
15	Regionaldaten	g (geography)
16	Technische Daten, Para- und Interviewerdaten	x

Innerhalb der thematischen Kategorien wird ein zweiter Buchstabe für die Zuweisung thematischer Unterkategorien verwendet. Zum Beispiel wird im Bereich Politik zwischen den Unterkategorien Partizipation (p für participation), Einstellungen (a für attitudes), Vertrauen (t für trust) und Nationalismus (n für nationalism) unterschieden. Tabelle 2 gibt einen Überblick über die Kategorien und Unterkategorien und den daraus resultierenden Buchstabenkennungen für die ersten zwei Stellen des vierstelligen Stammmamens.

Tabelle 2: Übersicht über inhaltliche Unterkategorien und dazugehörige Buchstabenkennungen

Kategorie	Unterkategorie	Buchstabenkennung	
Abweichendes Verhalten		c (crime)	
	Einstellungen zu abweichendem Verhalten	ca (attitudes)	
	Opfererfahrung	ce (experience)	
	Kriminalitätsfurcht	cf (fear)	
	Wahrgenommene Entdeckungswahrscheinlichkeit	cp (perception)	
	Selbstberichtetes abweichendes Verhalten	cs (self-report)	
Arbeitsorientierungen		j (job)	
	Arbeitsorientierungen	ja (attitudes)	
Demographie		d (demography)	
	Bildungsabschlüsse und -biographie	de (education)	
	Familienstand / Ehebiographie	df (family)	
	Wohnbiographie	dg (geography)	
	Haushalt	dh (household)	
	Einkommen	di (income)	
	Kinder	dk (kids)	
	Migration	dm (migration)	
	Staatsbürgerschaft	dn (nationality)	
	Partnerschaft	dp (partner)	
	Arbeit und Beruf	dw (work)	
	Ethnozentrismus und Minoritäten		m (minorities/migration)
		Einstellungen zu Migranten	ma (attitudes)
Kontakte zu Migranten		mc (contact)	
Diskriminierung von Migranten		md (discrimination)	
Emotionen zu verschiedenen ethnischen Gruppen		me (emotion)	
Einstellungen zu verschiedenen ethnischen Gruppen		mg (groups)	
Antisemitismus		mj (jews)	
Einstellungen zu Einwanderung		mi (immigration)	
Islamophobie		mm (muslims)	
Einbürgerung		mn (naturalization)	
Wahrnehmungen von Migranten		mp (perception)	

Kategorie	Unterkategorie	Buchstabenkennung
Ethnozentrismus und Minoritäten (fortgesetzt)	Soziale Distanz gegenüber Migranten	ms (social distance)
Familie		f (family)
	Erziehungsziele	fe (education)
	Aufteilung Hausarbeit	fh (housework)
	Familiale Normen	fn (norms)
	Geschlechterrollen	fr (roles)
Gesundheit		h (health)
	Gesundheitsspez. Einstellungen	ha (attitudes)
	Gesundheitsspez. Verhalten	hb (behavior)
	Krankheiten	hd (disease)
	Gesundheitliche Belastungen	he (exposure)
	Gesundheitlicher Zustand	hs (status)
	Arztbesuche/Krankenhaus	ht (therapy)
	Arbeitsbedingungen	hw (work)
Wertorientierungen		v (values)
	Postmaterialismus	va (attitudes)
	Einstellung zu Abtreibung	vm (morality)
	Individuelle Wertorientierungen (Klages)	vi (individual values)
Lebensstil und Persönlichkeit		l (life style)
	Freizeitaktivitäten	la (activities)
	Lebensziele	lg (goals)
	Wichtigkeit Lebensbereiche	li (importance)
	Mediennutzung	lm (media)
	Persönlichkeit	lp (personality)
	Lebenszufriedenheit	ls (satisfaction)
	Transnationalismus	lt (transnationalism)
Politik		p (politics)
	Politische Einstellungen	pa (attitudes)
	Politische Konflikte	pc (conflicts)
	Demokratieverständnis	pd (democracy)
	Politische Selbstwirksamkeit	pe (efficacy)
	Einstellungen zu Behörden	pg (government)

Kategorie	Unterkategorie	Buchstabenkennung
Politik (fortgesetzt)	Politische Sachfragen	pi (issues)
	Nationalismus	pn (nationalism)
	Politische Partizipation	pp (participation)
	Einstellungen zur Wiedervereinigung	pr (reunification)
	Zufriedenheit mit Politik	ps (satisfaction))
	Vertrauen in Politik & Institutionen	pt (trust)
	Wahlen	pv (vote)
Religion und Weltanschauung		r (religion)
	Einstellungen zur Religion	ra (attitudes)
	Glaube	rb (beliefs)
	Aktuelle und frühere Konfessionsmitgliedschaft	rd (denomination)
	Religiöse Erfahrung	re (experience)
	Alternative Glaubensformen	rh (heterodoxy)
	Religiöse Praxis	rp (practice)
	Religiöse Riten	rr (rites)
	Soziale Distanz gegenüber religiösen Gruppen	rs (social distance)
	Religiöse Sozialisation	ru (upbringing)
Soziale Netzwerke und soziales Kapital		s (social)
	Mitgliedschaft in Organisationen	sm (membership)
	Soziale Netzwerke	sn (network)
	Soziales Vertrauen	st (trust)
Soziale Ungleichheit		i (inequality)
	Deprivation	id (deprivation)
	Soziale Mobilität	im (mobility)
	Soziale Sicherung	is (social security)
	Einstellungen zum Wohlfahrtsstaat	iw (welfare state)
Wohnen und Umwelt		a (accomodation)
	Belastungen im Wohnumfeld	ae (exposure)
	Qualität des Wohnumfelds	aq (quality)
Wirtschaft (Einstellungen/Wahrnehmungen)		e (economy)
	Wahrnehmung der Wirtschaftslage	ep (perception)

Regionaldaten		g (geography)
	Dauer am Wohnort	gd (duration)
	Wohnortsqualität	gq (quality)
	Regionale Attribute	gr (region)
	Größe Wohnort	gs (size)
Technische Daten, Para- und Interview- werdaten		x
	Angaben zum Wohnhaus	xh (house)
	Daten zum Interviewer	xi (interviewer)
	Angaben zum Interview	xs (survey)
	Angaben zum Zeitpunkt des Interviews	xt (time)
	Angaben zur (Wieder-)Befragbarkeit	xr (respondent)
	Gewichte	wght (weight)
	Split-Variablen	splt (split)

Die dritte und vierte Stelle des Stammmens werden für die aufsteigende Nummerierung der Variablen innerhalb der thematischen Unterkategorien verwendet. Für die einstelligen Ziffern wird eine Null auf die erste Stelle eingefügt. Exemplarische Variablennamen aus der Fragebatterie zum Vertrauen in politische Institutionen sehen dann folgendermaßen aus:

pt01

pt02

Grundlage für die Bestimmung der Reihenfolge und damit Ausgangspunkt für die fortlaufende Nummerierung innerhalb der Unterkategorien ist die Reihenfolge der Variablen im Datensatz der ALLBUS Kumulation 1980-2014. Variablen, die bereits in einer früheren Welle des ALLBUS gelaufen sind und nach 2016 zum ersten Mal repliziert werden oder mit der Erhebung 2016 zum ersten Mal erhoben werden, werden nach der Zuweisung zu den inhaltlichen Kategorien fortlaufend durchnummeriert. Werden neue Schwerpunktmodule entwickelt, werden neue Kategorien und Unterkategorien bestimmt.

5. Für Archiv- und Paradata sowie wenige inhaltliche Variablen werden sprechende Variablennamen verwendet.

Abweichend von den vierstelligen Stammmens werden für bestimmte, häufig wiederkehrende Variablen und Inhalte sprechende Variablennamen verwendet. Zum einen haben sich im Datenarchiv der GESIS für bestimmte technische Variablen (Version, Studiennummer, DOI) Standardnamen etabliert, die auch für den ALLBUS übernommen werden. Darüber hinaus werden Gewichtungs-, Split-, und methodische Variablen mit sprechenden Variablennamen benannt. Dies erleichtert die Identifikation der Variablen im Datensatz und der Inhalt der Variablen ist für die Nutzer auf diese Weise unmittelbar ersichtlich. Tabelle 3 gibt eine Übersicht über die verwendeten technischen Variablen mit sprechendem Variablennamen.

Tabelle 3: Sprechende Variablenamen für Archivvariablen und Paradata

Technische Variablen	
Studiennummer	za_nr
Digital Object Identifier	doi
Versionskennung	version
Erhebungsjahr	year
Identifikationsnummer des Studienteilnehmers	respid
Erhebungsmodus	capipapi
Stichprobenverfahren	samptype
Fragebogensplit 1980	splt80
Fragebogensplit 1982	splt82
Fragebogensplit 1988	splt88
Fragebogensplit 1990	splt90
Fragebogensplit 1994	splt94
Fragebogensplit 2000	splt00
Fragebogensplit 2006	splt06
Fragebogensplit 2010	splt10
Erster Fragebogensplit 2012	splt12_1
Zweiter Fragebogensplit 2012	splt12_2
Dritter Fragebogensplit 2012	splt12_3
Erster Fragebogensplit 2014	splt14_1
Zweiter Fragebogensplit 2014	splt14_2
Dritter Fragebogensplit 2014	splt14_3
Erster Fragebogensplit 2016	splt16_1
Zweiter Fragebogensplit 2016	splt16_2
Dritter Fragebogensplit 2016	splt16_3
Vierter Fragebogensplit 2016	splt16_4
Fünfter Fragebogensplit 2016	splt16_5
Sechster Fragebogensplit 2016	splt16_6
Siebter Fragebogensplit 2016	splt16_7
Achter Fragebogensplit 2016	splt16_8
Transformationsgewicht Person	wghtptr
Personenbezogenes Ost-West-Gewicht	wghtpew
Ost-West Transformationsgewicht Person	wghtptew

Technische Variablen

Transformationsgewicht Haushalt	wghtht
Haushaltsbezogenes Ost-West-Gewicht	wghthew
Ost-West Transformationsgewicht Haushalt	wghthtew

Einige inhaltliche Variablen werden aus Gründen der Eingängigkeit ebenfalls mit sprechenden Namen benannt. Die sprechenden Namen lassen sich besonders leicht merken und erleichtern damit die Arbeit mit den Daten. Zu diesen Variablen gehören insbesondere demographische Merkmale wie Alter, Geschlecht, Schulabschluss, der Erwerbsstatus und die daraus abgeleiteten Klassifikationen (ISCO, ISEI, SIOPS etc.). Auch „prominente“ Variablen wie der Postmaterialismus-Index werden der Einfachheit halber mit sprechenden Namen versehen. Eine Übersicht findet sich in Tabelle 4. Wie die Buchstabenkennungen orientieren sich die sprechenden Variablennamen an den englischen Begriffen, damit die Daten auch für internationale Nutzer einfach verwendbar sind.

Tabelle 4: Sprechende Variablennamen für inhaltliche Variablen

Inhaltliche Variablen	
Geschlecht	sex
Geburtsmonat	mborn
Geburtsjahr	yborn
Alter	age
Deutsche Staatsangehörigkeit	german
Allgemeinbildender Schulabschluss	educ
Dauer der Schulausbildung in Jahren	educy
International Standard Classification of Education 1997	isced97
International Standard Classification of Education 2011	isced11
Erwerbsstatus	work
Internationale Standardklassifikation der Berufe 1988	isco88
Internationale Standardklassifikation der Berufe 2008	isco08
International Socio-Economic Index of Occupational Status 1988	isei88
International Socio-Economic Index of Occupational Status 2008	isei08
Standard International Occupational Prestige Scale 1988	siops88
Standard International Occupational Prestige Scale 2008	siops08
Treimanprestige I68	tps68
Magnitudeprestige I68	mps68
Magnitudeprestige I88	mps88
Goldthorpeklassen I68	gld68
Familienstand	mstat

Inhaltliche Variablen

Einkommen	inc
Haushaltseinkommen	hhinc
Verwandtschaftsgrad	kin
Inglehart-Index	ingle
Erhebungsgebiet alte oder neue Bundesländer	eastwest
Bundesland, in dem Befragter wohnt	land
BIK-Regionen	bik
Politische Gemeindegrößenklasse	gkpol
Boustedt-Gemeindetyp 7 Kategorien	bstdt07
Boustedt-Gemeindetyp 10 Kategorien	bstdt10

6. *Präfixe zum Stammmamen oder sprechenden Variablennamen zeigen den Personenbezug der Variable an.*

Eine Reihe von Merkmalen werden fast identisch für verschiedene Personen im Umfeld der Befragten erhoben (gegenwärtige und ehemalige Ehe- oder Lebenspartner, Mutter, Vater, Haushaltsmitglieder, etc.). Diese Variablen erhalten für alle Bezugspersonen den gleichen Stammmamen (zum Beispiel age für das Lebensalter). Eine ein- bis dreistellige Buchstabenkennung, die dem Stammmamen vorangestellt wird, bezeichnet die Person, für die das Merkmal erhoben wird. Dreistellige Präfixe werden insbesondere für die Haushaltsliste und die Kinder außer Haus benötigt. Tabelle 5 zeigt, welche Präfixe für den ALLBUS benötigt werden. Wird kein Präfix verwendet, beschreibt die Variable die Merkmale des Befragten selbst. Die Auswahl lehnt sich an die im General Social Survey verwendeten Suffixe zu den Variablennamen an.

Tabelle 5: Übersicht über Präfixe zum Variablennamen

Personenbezug	Präfix	Englischer Begriff
Gegenwärtiger Ehepartner	sc	spouse current
Früherer Ehepartner	sf	spouse former
Partner (nicht verheiratet)	p	partner
Xte Person im Haushalt ^a	hxx	household member x
Xtes Kind außer Haus ^b	kxx	kid away from home x
Mutter	m	mother
Vater	f	father
Großvater mütterlicherseits	gfm	grandfather maternal
Großvater väterlicherseits	gfp	grandfather paternal
Großmutter mütterlicherseits	gmm	grandmother maternal
Großmutter väterlicherseits	gmp	grandmother paternal

Kontakt A (Netzwerk)	ca	contact A
Kontakt B (Netzwerk)	cb	contact B
Kontakt C (Netzwerk)	cc	contact C
Kontakt D (Netzwerk)	cd	contact D
Kontakt E (Netzwerk)	ce	contact E

Bemerkung: ^a Es werden Daten für bis zu 8 Haushaltspersonen erhoben. ^b Es werden Daten für bis zu 10 Kindern außer Haus erhoben.

Tabelle 6 zeigt am Beispiel der Altersvariable (age) wie die Verwendung von Präfixen in der Praxis funktioniert.

Tabelle 6: Beispiel für die Verwendung von Präfixen

Personenbezug	Stammmname	Präfix	Variablenname
Alter des Befragten	age		age
Alter des gegenwärtigen Ehepartners	age	sc	scage
Alter des früheren Ehepartners	age	sf	sfage
Alter des Partners	age	p	page
Alter Haushaltsperson 2	age	hh2	hh2age
Alter Haushaltsperson 3	age	hh3	hh3age
Alter des 1. Kindes außer Haus	age	kh1	kh1age
Alter des 2. Kindes außer Haus	age	kh2	kh2age
Mutter	age	m	mage
Vater	age	f	fage
Kontakt A	age	ca	caage
Kontakt B	age	cb	cbage
Kontakt C	age	cc	ccage
Kontakt D	age	cd	cdage
Kontakt E	age	ce	ceage

Die Präfixe werden insbesondere für die Variablen Geburtsjahr, Geburtsmonat, Alter, Geschlecht, Staatsbürgerschaft sowie die Bildungs- und Berufsvariablen verwendet. Die Verwendung von Präfixen hat gegenüber Suffixen (wie im General Social Survey) den Vorteil, dass die Datensätze einfach nach der jeweiligen Bezugsperson geordnet werden können. Das ist bei der Verwendung von Suffixen komplizierter. Darüber hinaus sind die Variablen zu den Bezugspersonen des Befragten im Datensatz gruppiert. Bei der Suche im Datensatz sind diese Blöcke über die Präfixe erkennbar.

7. *Suffixe zum Stammmamen machen Varianten gleicher bzw. ähnlicher Variablen kenntlich.*

In bestimmten Fällen werden auch Suffixe zu den Stammmamen und den sprechenden Variablenamen verwendet werden. Das betrifft in erster Linie Ableitungen und Kategorisierungen von Variablen. In den ALLBUS Datensätzen werden zum Beispiel viele metrische Variablen auch als kategorisierte Variablen angeboten. Diesen Variablen wird einheitlich das Suffix *c* beigefügt. Dies wird am Beispiel der Altersvariablen veranschaulicht (siehe Tabelle 7).

Tabelle 7: Beispiel für die Verwendung des Suffixes *c* für die Benennung von Kategorisierungen

Variablenlabel	Stammmame	Präfix	Suffix	Variablenname
Alter des Befragten	age	-	-	age
Alter des Befragten kategorisiert	age	-	c	agec
Alter des gegenwärtigen Ehepartners	age	sc	-	scage
Alter des gegenwärtigen Ehepartners kategorisiert	age	sc	c	scagec

Systematisch aufeinander bezogene Fragen werden ebenfalls mit dem gleichen Stammmamen versehen und über Suffixe *a* und *b* differenziert. Dies betrifft etwa verschiedene Varianten einer Frage, die in Fragebogensplits erhoben werden. Ein Beispiel dafür sind die heterodoxen (Glaubens)-Praktiken, die im Religionsschwerpunkt wiederholt gefragt wurden. Eine Variable beschreibt die Bekanntheit der Praktiken, die zweite die Beurteilung der Praktiken. Hier kann die Benennung der Variablen wie in Tabelle 8 vorgenommen werden.

Tabelle 8: Verwendung des Suffixes zur Markierung inhaltlich verbundener Variablen am Beispiel heterodoxer Praktiken

Variablenlabel	Name mit Suffix
ERFAHRUNG: NEW AGE, WASSERMANN AERA	rh01a
HALTE VON: NEW AGE, WASSERMANN AERA	rh01b
ERFAHRUNG: ZEN-MEDITATION, WEISHEITEN	rh02a
HALTE VON: ZEN-MEDITATION, WEISHEITEN	rh02b
ERFAHRUNG: ANTHROPOSOPHIE, THEOSOPHIE	rh03a
HALTE VON: ANTHROPOSOPHIE, THEOSOPHIE	rh03b
ERFAHRUNG: REINKARNATION	rh04a
HALTE VON: REINKARNATION	rh04b

3 Umsetzung des neuen Konzeptes

Das neue Konzept wird zum ersten Mal für den Querschnitt 2016 umgesetzt. Im Anschluss daran werden die festen Variablennamen auch für die Kumulation 1980 bis 2016 angewendet sowie für alle folgenden Datensätze. Eine Aktualisierung der alten Datensätze mit den neuen Variablennamen ist nicht vorgesehen.