

2022

FZDW TECHNICAL REPORTS [NO. 1]



FORSCHUNGSZENTRUM
DEMOGRAFISCHER
WANDEL (FZDW)

GESUNDHEITSVERHALTEN UND UNFALLGESCHEHEN IM SCHULALTER (GUS)
GENERIERTE VARIABLEN IM GUS-DATENSATZ
ROBERT LIPP



Autor:

Robert Lipp 

Zitationsempfehlung:

Lipp, R. (2022). Generierte Variablen im GUS-Datensatz. *FZDW Technical Reports, 1*. Frankfurt am Main, Forschungszentrum Demografischer Wandel (FZDW), Frankfurt University of Applied Sciences

Die Durchführung der GUS-Studie war eine große Gemeinschaftsleistung. *Der Beitrag der folgenden Personen wird dankend anerkannt:*

Andreas Klocke
Sven Stadtmüller
Andrea Giersiefen
Christina Wagner
Sarah Maier
Jannik Track
Christina Lenz-Bokhari
Ilona Kraus
Christine Leyhe
Mario Englert
Carina Schwaderer
Jessica Konschu

Kontakt:

Frankfurt University of Applied Sciences
Forschungszentrum Demografischer Wandel (FZDW)
Nibelungenplatz 1
60318 Frankfurt am Main
E-Mail: info@fzdw.de
Web: www.fzdw.de



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Inhaltsverzeichnis

1. Über diesen Bericht 4

2. Generierte Variablen 4

1. Über diesen Bericht

Im Rahmen der Datenedition wurden 136 neue Variablen generiert. Dabei handelt es sich um Skalen und Indizes, Harmonisierungen sowie auf andere Weise berechnete oder erstellte Hilfs- und Servicevariablen. Im vorliegenden Dokument sind diese Variablen und ihre Entstehungslogiken detailliert beschrieben und werden diejenigen Variablen gelistet, die zur Erstellung verwendet wurden. Ggf. sind auch Syntaxen und Anmerkungen zur Nutzung dokumentiert. Die Reihenfolge entspricht der Abfolge der Variablen im Datensatz. Die Syntaxen wurden in Stata (Version 14) erstellt.

2. Generierte Variablen

Variablenname: ID_t_g1	Ursprungsvariable(n): ID_t_short
Beschreibung: Numerische Entsprechung des selbstgenerierten ID-Codes der Schülerinnen und Schüler. Ursprünglich bestand dieser aus dem ersten Buchstaben des eigenen Vornamens, dem ersten Buchstaben des Vornamens der Mutter, dem ersten Buchstaben des Vornamens des Vaters und dem Geburtsmonat. Dieser Code wurde mittels eines Algorithmus nochmals umgeschlüsselt.	
Anmerkungen: Die Originalvariable ist aus Datenschutzgründen nicht im Datensatz enthalten.	
Syntax: -	

Variablenname: ID_t_g2	Ursprungsvariable(n): ID_t_short
Beschreibung: Numerische Entsprechung des selbstgenerierten ID-Codes der Schülerinnen und Schüler nach Record Linkage. Ursprünglich bestand dieser aus dem ersten Buchstaben des eigenen Vornamens, dem ersten Buchstaben des Vornamens der Mutter, dem ersten Buchstaben des Vornamens des Vaters und dem Geburtsmonat. Die Codes wurden dann mithilfe eines Bigram-Algorithmus mit den Codes der anderen Wellen verglichen. Auf diese Weise konnten ca. 5% der Fälle nachträglich noch anderen Fällen zugeordnet werden. Der Code wurde mittels eines Algorithmus numerisch umgeschlüsselt.	
Anmerkungen: Für Panelanalysen empfehlen wir diesen Code zu verwenden. Weitere Informationen zu den Record-Linkage Prozeduren finden Sie unter: https://doi.org/10.13094/SMIF-2021-00006 Die Originalvariable ist aus Datenschutzgründen nicht im Datensatz enthalten.	
Syntax: -	

Variablenname: s_east_west_g1	Ursprungsvariable(n): s_fedstate
Beschreibung: Dummyvariable mit den Ausprägungen „1“ für Westdeutschland und „2“ für Ostdeutschland.	
Anmerkungen: -	
Syntax: gen s_east_west_g1 = s_fedstate recode s_east_west_g1 (1/8=1) (9/14=2)	

Variablenname: i_sex_g1	Ursprungsvariable(n): s_interviewer
Beschreibung: Geschlecht der Interviewerin bzw. des Interviewers. Bei Interview im Tandem wurde für beide Interviewerinnen bzw. Interviewer das Geschlecht angegeben. Die Angaben wurden abgeleitet aus den Namen der Interviewerinnen und Interviewer.	
Anmerkungen: Die Ursprungsvariable ist aus Datenschutzgründen nicht im Datensatz enthalten.	
Syntax: -	

Variablenname: h_incl_g1	Ursprungsvariable(n): h_incl, h_incl_1, h_incl_2, h_incl_3, h_incl_4, h_incl_5, h_incl_6
Beschreibung: Wurde im Schulleiterfragebogen keine Angabe zur Gesamtzahl der Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf gemacht, wohl aber die Anzahl der Kinder und Jugendlichen nach bestimmten Förderschwerpunkten notiert, so wurden diese Angaben addiert und als Gesamtzahl imputiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: <pre>gen h_incl_g1 = h_incl replace h_inclusion_g1 = h_incl_1 + h_incl_2 + h_incl_3 + h_incl_4 + h_incl_5 + h_incl_6 if h_incl_g1 == .e & (!missing(h_incl_1) !missing(h_incl_2) !missing(h_incl_3) !missing(h_incl_4) !missing(h_incl_5) !missing(h_incl_6)) replace h_incl_g1 = .h if h_incl == .h</pre>	

Variablenname: h_incl_1_h_g1 - h_incl_8_h_g1	Ursprungsvariable(n): h_incl_1, h_incl_2, h_incl_3, h_incl_4_s, h_incl_5_s, h_incl_6_s, h_incl_w5_1, h_incl_w5_2, h_incl_w5_3, h_incl_w5_4, h_incl_w5_5, h_incl_w5_6, h_incl_w5_7, h_incl_w5_8
Beschreibung: Die Erhebung der Förderschwerpunkte erfuhr ab Welle 5 eine Überarbeitung. So wurde fortan nicht mehr die Anzahl der im jeweiligen Schwerpunkt geförderten Schülerinnen und Schüler erhoben, sondern lediglich das Vorhandensein des Förderschwerpunkts. Zudem wurden die Förderschwerpunkte „Sehen“, „Hören“, „Sprache“ und „Lernen“ nun kategorisiert mit abgefragt. Für das im Datensatz vorhandene Variablenset wurden zunächst die offenen Angaben zu anderen Schwerpunkten in den Wellen 3 und 4 in diese Kategorien nachcodiert. Anschließend wurden die Antworten in das Kategorienschema der Wellen 5 und 6 überführt und mit den dort gemachten Angaben zusammengeführt (= Harmonisierung).	
Anmerkungen: Aus Datenschutzgründen können die Ursprungsvariablen (inkl. der offenen Angaben zu anderen Förderschwerpunkten) nicht zur Verfügung gestellt werden.	
Syntax:	

Variablenname: t_duration_g1	Ursprungsvariable(n): t_date, t_endDate
Beschreibung: Berechnung der Befragungsdauer (in Minuten) aus den Angaben zum Startzeitpunkt (Betätigung der „Start“-Schaltfläche auf dem Tablet) und zum Endzeitpunkt der Befragung (Betätigung der Schaltfläche „Ende“).	
Anmerkungen: -	
Syntax: gen duration = t_endDate-t_date format duration %tc gen double day = dofC(duration) gen double hour = hhC(duration) gen double minute = mmC(duration) gen double second = ssC(duration) gen t_duration_g1 = (day*24*60)+(hour*60)+minute+(second/60) drop duration day hour minute second	

Variablenname: t_month_g1	Ursprungsvariable(n): t_date, t_endDate (nur zur Prüfung)
Beschreibung: Extraktion des Befragungsmonats aus den Angaben zum Startzeitpunkt (Betätigung der „Start“-Schaltfläche auf dem Tablet) der Befragung.	
Anmerkungen: -	
Syntax: gen t_month_g1 = dofC(t_date) gen endmonth = dofC(t_endDate) // Benötigt um zu prüfen, ob der Anfang und Ende im gleichen Monat lagen format t_month_g1 endmonth %27.0g replace t_month_g1 = month(t_month_g1) replace endmonth = month(endmonth) replace t_month_g1 = .g if t_month_g1!=endmonth	

Variablenname: t_classsize_g1	Ursprungsvariable(n): t_classsize
Beschreibung: Am häufigsten genannter Wert (= Modus) bei der Abfrage der Anzahl der Schülerinnen und Schüler in einer Klasse. Waren mehrere Modi vorhanden, wurde der höhere Modus verwendet.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Originalvariable.	
Syntax: bysort ID_c wave: egen t_classsize_g1 = mode(t_classsize), maxmode replace t_classsize_g1 = .b if t_classsize==.	

Variablenname: t_as003_g1	Ursprungsvariable(n): t_as003, t_as003_s
Beschreibung: Es wurde geprüft, ob die offenen Angaben bei Auswahl der Residualkategorie („Anderer Ort“) sich doch einer der geschlossenen Antwortmöglichkeiten zuordnen lassen. War dies der Fall, wurde die Auswahl entsprechend recodiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax: -	

Variablennamen: t_as006_1_h - t_as006_10_s_h	Ursprungsvariable(n): t_as006_w1_1 - t_as006_w1_9_s, t_as006_w2_1 - t_as006_w2_10_s
Beschreibung: In Welle 2 wurden bei der Mehrfachauswahl zwei neue Antwortmöglichkeiten hinzugenommen. Das harmonisierte Variablenset führt die Angaben dieser zwei Versionen der Frage zusammen.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax (exemplarisch): clonevar t_as006_1_h = t_as006_w2_1 replace t_as006_1_h = t_as006_w1_1 if wave==1	

Variablennamen: t_as006_1_h_g1 - t_as006_10_h_g1	Ursprungsvariable(n): t_as006_1_h - t_as006_10_h, t_as006_10_s_h
Beschreibung: Es wurde geprüft, ob die offenen Angaben bei Auswahl der Residualkategorie („Andere Schäden“) sich doch einer der geschlossenen Antwortmöglichkeiten zuordnen lassen. War dies der Fall, wurde die entsprechende Variable recodiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: -	

Variablenname: t_as010_g1	Ursprungsvariable(n): t_as010, t_as010_s
Beschreibung: Es wurde geprüft, ob die offenen Angaben bei Auswahl der Residualkategorie („Sonstiges“) sich doch einer der geschlossenen Antwortmöglichkeiten zuordnen lassen. War dies der Fall, wurde die Auswahl entsprechend recodiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax: -	

Variablenname: t_h003_g1	Ursprungsvariable(n): t_h003, t_h004_g3
<p>Beschreibung: Zunächst wurde aus den Angaben zur Anzahl der Schulstunden Sportunterricht und den verarbeiteten Angaben zur Dauer einer Schulstunde (siehe t_h004_g3) die Gesamtdauer des Sportunterrichts in einer Woche in Stunden berechnet. Wurden zur Anzahl der Schulstunden Sportunterricht keine Angaben gemacht, so wurde der am häufigsten genannte Wert (= Modus) in der Jahrgangsstufe der Schule imputiert. Waren mehrere Modi vorhanden, wurde der höhere Modus verwendet.</p>	
<p>Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.</p>	
<p>Syntax: gen t_h003_g1 = (t_h003*t_h004_g3)/60 bysort ID_s wave: egen sportmode = mode(t_h003_g1), maxmode replace t_h003_g1 = sportmode if t_h003==.e</p>	

Variablenname: t_sport_g1	Ursprungsvariable(n): t_h001, t_h002, t_h003_g1
<p>Beschreibung: Summe der Stunden, die in Freizeit, Verein und Unterricht mit Sport verbracht werden.</p>	
<p>Anmerkungen: -</p>	
<p>Syntax: gen t_sport_g1 = t_h001+t_h002+t_h003_g1 replace t_sport_g1 = .g if t_sport_g1==.</p>	

Variablenname: t_h004_g1	Ursprungsvariable(n): t_h004
<p>Beschreibung: Korrigierte Angabe zur Dauer einer Schulstunde. Es wurde dabei für jede Schule die Angabe verwendet, die von der Mehrzahl der Schülerinnen und Schüler gemacht wurde. Diese wurde dann auf alle anderen Schülerinnen und Schüler übertragen.</p>	
<p>Anmerkungen: -</p>	
<p>Syntax: egen hour45 = sum(t_h004==1), by(ID_s wave) egen hour60 = sum(t_h004==2), by(ID_s wave) egen hourother = sum(t_h004==3), by(ID_s wave) gen t_h004_g1 = t_h004 bysort wave: replace t_h004_g1 = 1 if hour45>hour60 & hour45>hourother & t_h004!=.b bysort wave: replace t_h004_g1 = 2 if hour60>hour45 & hour60>hourother & t_h004!=.b bysort wave: replace t_h004_g1 = 3 if hourother>hour45 & hourother>hour60 & t_h004!=.b</p>	

Variablenname: t_h004_g2	Ursprungsvariable(n): t_h004_g1, t_h004_1
Beschreibung: Korrigierte, numerische Angabe zur Dauer einer Schulstunde. Sofern die Mehrzahl der Schülerinnen und Schüler eine andere Minutenzahl als 45 oder 60 ausgewählt hat, wurde die offene Angabe verwendet. Zudem wurden die Kategorien (1, 2 und 3) in Minuten (45, 60 usw.) umgewandelt.	
Anmerkungen: -	
Syntax: egen other_mode = mode(t_h004_1), by(ID_s wave) maxmode gen t_h004_g2 = . replace t_h004_g2 = other_mode if t_h004_g1==3 replace t_h004_g2 = 45 if t_h004_g1==1 replace t_h004_g2 = 60 if t_h004_g1==2 replace t_h004_g2 = .b if t_h004_g1==.b replace t_h004_g2 = .e if t_h004_g1==.e	

Variablenname: t_h004_g3	Ursprungsvariable(n): t_h004_g2
Beschreibung: Korrigierte, numerische Angabe zur Dauer einer Schulstunde inkl. Imputation. Die Dauer einer Schulstunde wurde nur in den Wellen 2 und 3 erfragt. In dieser Variable sind die Angaben auf die jeweils anderen Wellen mit fehlenden Werten übertragen worden. Dabei wurde, wenn möglich, in Welle 1 auf die Angaben in der Welle 2 und in den Wellen 4, 5 und 6 auf die Angaben in der Welle 3 zurückgegriffen. Es wurde jeweils der am häufigsten genannte Wert der Schule verwendet.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax: egen modew2 = mode(t_h004_g2) if wave==2, by(ID_s) maxmode bysort ID_s(modew2): replace modew2 = modew2[1] egen modew3 = mode(t_h004_g2) if wave==3, by(ID_s) maxmode bysort ID_s(modew3): replace modew3 = modew3[1] gen t_h004_g3 = . replace t_h004_g3 = modew2 if wave<=2 replace t_h004_g3 = modew3 if wave<=2 & missing(modew2) & !missing(modew3) replace t_h004_g3 = modew3 if wave>=3 replace t_h004_g3 = modew2 if wave>=3 & missing(modew3) & !missing(modew2)	

Variablenname: t_h011_g1	Ursprungsvariable(n): t_h011
Beschreibung: Stundenangabe der Uhrzeit, wann die Schülerinnen und Schüler in einer normalen Woche aufstehen. In den Wellen 1 und 2 wurden die Schülerinnen und Schüler gebeten ihre Aufstehzeit als drei- bzw. vierstellige Zahl anzugeben (z. B. 7:30 Uhr = 730). In den darauffolgenden Wellen konnte ein spezieller Variablentyp für Zeitangaben verwendet werden. Aufstehzeiten vor 4:30 Uhr und nach 10:00 Uhr wurden als unplausibel gewertet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_h011_g2	Ursprungsvariable(n): t_h011
Beschreibung: Minutenangabe der Uhrzeit, wann die Schülerinnen und Schüler in einer normalen Woche aufstehen. In den Wellen 1 und 2 wurden die Schülerinnen und Schüler gebeten ihre Aufstehzeit als drei- bzw. vierstellige Zahl anzugeben (z. B. 7:30 Uhr = 730). In den darauffolgenden Wellen konnte im Tabletfragebogen ein spezieller Variablentyp für Zeitangaben verwendet werden. Aufstehzeiten vor 4:30 Uhr und nach 10:00 Uhr wurden als unplausibel gewertet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_h012_g1	Ursprungsvariable(n): t_h012
Beschreibung: Stundenangabe der Uhrzeit, wann die Schülerinnen und Schüler in einer normalen Woche ins Bett gehen. In den Wellen 1 und 2 wurden die Schülerinnen und Schüler gebeten ihre Ins-Bett-Geh-Zeit als drei- bzw. vierstellige Zahl anzugeben (z. B. 21:30 Uhr = 2130). In den darauffolgenden Wellen konnte im Tabletfragebogen ein spezieller Variablentyp für Zeitangaben verwendet werden. Ins-Bett-Geh-Zeiten vor 17:00 Uhr und nach 3:00 Uhr wurden als unplausibel gewertet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_h012_g2	Ursprungsvariable(n): t_h012
Beschreibung: Minutenangabe der Uhrzeit, wann die Schülerinnen und Schüler in einer normalen Woche ins Bett gehen. In den Wellen 1 und 2 wurden die Schülerinnen und Schüler gebeten ihre Ins-Bett-Geh-Zeit als drei- bzw. vierstellige Zahl anzugeben (z. B. 21:30 Uhr = 2130). In den darauffolgenden Wellen konnte im Tabletfragebogen ein spezieller Variablentyp für Zeitangaben verwendet werden. Ins-Bett-Geh-Zeiten vor 17:00 Uhr und nach 3:00 Uhr wurden als unplausibel gewertet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_sleep_g1	Ursprungsvariable(n): t_h011_g1, t_h011_g2, t_h012_g1, t_h012_g2
Beschreibung: Dauer zwischen der Uhrzeit, zu der die Schülerinnen und Schüler ins Bett gehen und der Uhrzeit, zu der sie wieder aufstehen in Stunden.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax: gen t_sleep_g1 = t_h011_g1+(t_h011_g2/60) + (24-(t_h012_g1+(t_h012_g2/60))) replace t_sleep_g1 = t_h011_g1+(t_h011_g2/60) - (t_h012_g1+(t_h012_g2/60)) /// if t_h012_g1+(t_h012_g2/60)<12 replace t_sleep_g1 = .c if t_sleep_g1<0 (t_sleep_g1>24 & t_sleep_g1<.) replace t_sleep_g1 = .g if t_sleep_g1==.	

Variablenamen: t_h020_1_g1, t_h020_2_g1, t_h020_3_g1, t_h020_4_g1, t_h020_5_g1,	Ursprungsvariable(n): t_h020_1, t_h020_2, t_h020_3, t_h020_4, t_h020_5
Beschreibung: Stundenangabe der Uhrzeit des Beginns des Sportunterrichts. Es konnte der Beginn von bis zu fünf Einheiten Sportunterricht dokumentiert werden.	
Anmerkungen: -	
Syntax: foreach var of varlist t_h020_1 t_h020_2 t_h020_3 t_h020_4 t_h020_5 { gen double `var'_g1 = hh(`var') replace `var'_g1 = `var' if missing(`var'_g1) & missing(`var') replace `var'_g1 = .b if `var'_g1==0 & `var'_g2==0 }	

Variablenamen: t_h020_1_g2, t_h020_2_g2, t_h020_3_g2, t_h020_4_g2, t_h020_5_g2,	Ursprungsvariable(n): t_h020_1, t_h020_2, t_h020_3, t_h020_4, t_h020_5
Beschreibung: Minutenangabe der Uhrzeit des Beginns des Sportunterrichts. Es konnte der Beginn von bis zu fünf Einheiten Sportunterricht dokumentiert werden.	
Anmerkungen: -	
Syntax: foreach var of varlist t_h020_1 t_h020_2 t_h020_3 t_h020_4 t_h020_5 { gen double `var'_g2 = mm(`var') replace `var'_g2 = `var' if missing(`var'_g2) & missing(`var') replace `var'_g2 = .b if (`var'_g1==0 `var'_g1==.b) & `var'_g2==0 }	

Variablenamen: t_h020_1_g3, t_h020_2_g3, t_h020_3_g3, t_h020_4_g3, t_h020_5_g3,	Ursprungsvariable(n): t_h020_1, t_h020_2, t_h020_3, t_h020_4, t_h020_5
Beschreibung: Kategorisierung des Beginns des Sportunterrichts in die Zeiträume „vor 8 Uhr“, „8 bis vor 10 Uhr“, „10 bis vor 12 Uhr“, „12 bis vor 14 Uhr“ und „ab 14 Uhr“.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: foreach var of varlist t_h020_1 t_h020_2 t_h020_3 t_h020_4 t_h020_5 { recode `var'_g1 (min/7 = 1) (8/9 = 2) (10/11 = 3) (12/13 = 4) (14/max = 5), gen(`var'_g3) replace `var'_g3 = .b if `var'_g1==.b & `var'_g2==.b }	

Variablenamen: t_h032_1_g1 - t_h032_8_g1	Ursprungsvariable(n): t_h032_1 - t_h032_8, t_h032_8_s
Beschreibung: Es wurde geprüft, ob die offenen Angaben bei Auswahl der Residualkategorie („Sonstiges, und zwar“) sich doch einer der geschlossenen Antwortmöglichkeiten zuordnen lassen. War dies der Fall, wurde die entsprechende Variable recodiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: -	

Variablenamen: t_s001_1_g1, t_s002_1_g1, t_s003_1_g1, t_s004_1_g1, t_s005_1_g1	Ursprungsvariable(n): t_s001_1, t_s002_1, t_s003_1, t_s004_1, t_s005_1
Beschreibung: Stundenangabe der Uhrzeit, wann die Schülerinnen und Schüler am jeweiligen Wochentag (Montag bis Freitag) in der Schule ankommen.	
Anmerkungen: Auf einigen Tablets waren die Zeitangaben nicht im 24h-Format möglich, sondern es musste gesondert „am“ oder „pm“ für vormittags und nachmittags ausgewählt werden. Da dies bei vielen Schülerinnen und Schülern zu Verwirrung führte, wurden unplausible Werte am Nachmittag in ihr jeweiliges Pendant am Vormittag umkodiert. Ob die Angabe von einem Tablet mit entsprechender Formatvorgabe stammt ist der Variable t_ampm_g1 zu entnehmen.	
Syntax:	

Variablenamen: t_s001_1_g2, t_s002_1_g2, t_s003_1_g2, t_s004_1_g2, t_s005_1_g2	Ursprungsvariable(n): t_s001_1, t_s002_1, t_s003_1, t_s004_1, t_s005_1
Beschreibung: Minutenangabe der Uhrzeit, wann die Schülerinnen und Schüler am jeweiligen Wochentag (Montag bis Freitag) in der Schule ankommen.	
Anmerkungen: Auf einigen Tablets waren die Zeitangaben nicht im 24h-Format möglich, sondern es musste gesondert „am“ oder „pm“ für vormittags und nachmittags ausgewählt werden. Da dies bei vielen Schülerinnen und Schülern zu Verwirrung führte, wurden unplausible Werte am Nachmittag in ihr jeweiliges Pendant am Vormittag umkodiert. Ob die Angabe von einem Tablet mit entsprechender Formatvorgabe stammt ist der Variable t_ampm_g1 zu entnehmen.	
Syntax:	

Variablenamen: t_s001_2_g1, t_s002_2_g1, t_s003_2_g1, t_s004_2_g1, t_s005_2_g1	Ursprungsvariable(n): t_s001_2, t_s002_2, t_s003_2, t_s004_2, t_s005_2
Beschreibung: Stundenangabe der Uhrzeit, wann die Schülerinnen und Schüler am jeweiligen Wochentag (Montag bis Freitag) die Schule verlassen.	
Anmerkungen: Auf einigen Tablets waren die Zeitangaben nicht im 24h-Format möglich, sondern es musste gesondert „am“ oder „pm“ für vormittags und nachmittags ausgewählt werden. Da dies bei vielen Schülerinnen und Schülern zu Verwirrung führte, wurden unplausible Werte am Nachmittag in ihr jeweiliges Pendant am Vormittag umkodiert. Ob die Angabe von einem Tablet mit entsprechender Formatvorgabe stammt ist der Variable t_ampm_g1 zu entnehmen.	
Syntax:	

Variablenamen: t_s001_2_g2, t_s002_2_g2, t_s003_2_g2, t_s004_2_g2, t_s005_2_g2	Ursprungsvariable(n): t_s001_2, t_s002_2, t_s003_2, t_s004_2, t_s005_2
Beschreibung: Minutenangabe der Uhrzeit, wann die Schülerinnen und Schüler am jeweiligen Wochentag (Montag bis Freitag) die Schule verlassen.	
Anmerkungen: Auf einigen Tablets waren die Zeitangaben nicht im 24h-Format möglich, sondern es musste gesondert „am“ oder „pm“ für vormittags und nachmittags ausgewählt werden. Da dies bei vielen Schülerinnen und Schülern zu Verwirrung führte, wurden unplausible Werte am Nachmittag in ihr jeweiliges Pendant am Vormittag umkodiert. Ob die Angabe von einem Tablet mit entsprechender Formatvorgabe stammt ist der Variable t_ampm_g1 zu entnehmen.	
Syntax:	

Variablenamen: t_s001_g1, t_s002_g1, t_s003_g1, t_s004_g1, t_s005_g1	Ursprungsvariable(n): t_s001_1_g1, t_s002_1_g1, t_s003_1_g1, t_s004_1_g1, t_s005_1_g1 t_s001_1_g2, t_s002_1_g2, t_s003_1_g2, t_s004_1_g2, t_s005_1_g2 t_s001_2_g1, t_s002_2_g1, t_s003_2_g1, t_s004_2_g1, t_s005_2_g1 t_s001_2_g2, t_s002_2_g2, t_s003_2_g2, t_s004_2_g2, t_s005_2_g2
Beschreibung: Aufenthaltsdauer der Schülerinnen und Schüler in der Schule (= Expositionszeit). Hierfür wurde die Zeit der Ankunft in der Schule von der Zeit des Verlassens der Schule abgezogen. In Welle 1 wurde die Uhrzeit der Ankunft in der Schule noch nicht erfragt. Daher wurden für diese Welle die Angaben imputiert. Wo möglich, wurden dafür die individuellen Angaben aus Welle 2 verwendet. Wo keine individuellen Werte vorlagen, wurde der am häufigsten genannte Wert in der jeweiligen Schule verwendet. Lagen dazu auch keine Angaben vor, wurde der insgesamt am häufigsten genannte Wert insgesamt imputiert.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax:	

Variablenname: t_s007_h	Ursprungsvariable(n): t_s007_to, t_s007_from
Beschreibung: Harmonisierte Angaben zur Gefährlichkeit des Schulwegs. In den Wellen 1 und 2 wurde die Gefährlichkeit sowohl auf dem Hinweg zur Schule als auch auf dem Rückweg erfragt. In der harmonisierten Variable wurden fehlende Angaben zum Hinweg, mit gültigen Angaben zum Rückweg ersetzt. Ab Welle 3 wurde die Gefährlichkeit des Schulwegs nur noch auf dem Hinweg zur Schule erfragt.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: gen t_s007_h = t_s007_to replace t_s007_h = t_s007_from if missing(t_s007_h) & !missing(t_s007_from) replace t_s007_h = .e if t_s007_from==.e replace t_s007_h = .a if t_s007_from==.a	

Variablenname: t_s008_h	Ursprungsvariable(n): t_s008_to, t_s008_from
Beschreibung: Harmonisierte Angaben zur Dauer des Schulwegs. In den Wellen 1 und 2 wurde die Schulwegdauer sowohl auf dem Hinweg zur Schule als auch auf dem Rückweg erfragt. In der harmonisierten Variable wurden fehlende Angaben zum Hinweg, mit gültigen Angaben zum Rückweg ergänzt. In den Wellen 4 bis 6 wurde die Dauer des Schulwegs nur noch auf dem Hinweg zur Schule erfragt.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: gen t_s007_h = t_s007_to replace t_s007_h = t_s007_from if missing(t_s007_h) & !missing(t_s007_from) replace t_s007_h = .e if t_s007_from==.e replace t_s007_h = .a if t_s007_from==.a	

Variablenname: t_sex_g1	Ursprungsvariable(n): t_sex
Beschreibung: Fehlende Angaben zum Geschlecht in einer Welle, wurden mit gültigen Angaben aus den anderen Wellen ergänzt. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_sex_g1_h	Ursprungsvariable(n): t_sex_g1
Beschreibung: Die Variable harmonisiert die Geschlechtsangaben über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax:	

Variablenname: t_birth_g1	Ursprungsvariable(n): t_birth
Beschreibung: Fehlende Angaben zum Geburtsjahr in einer Welle, wurden mit gültigen Angaben aus den anderen Wellen ergänzt. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_birth_g1_h	Ursprungsvariable(n): t_birth_g1
Beschreibung: Die Variable harmonisiert die Angaben zum Geburtsjahr über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: -	
Syntax:	

Variablenname: t_age_g1	Ursprungsvariable(n): t_birth_g1_h, ID_t_4 (ursprüngliche Angabe des Geburtsmonats), t_date
Beschreibung:	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariable.	
Syntax: <pre>gen byear = 2000 + t_birth_g1_h gen bdate = byear+(ID_t_4/12) replace bdate = .e if byear==. gen date = dofc(t_date) format date %td gen double syear = year(date) gen double smonth = month(date) gen sdate = syear+(smonth/12) gen t_age_g1 = sdate-bdate replace t_age_g1 = .g if missing(t_birth_g1_h) missing(t_date)</pre>	

Variablenname: t_gert_g1	Ursprungsvariable(n): t_gert
Beschreibung: Variable zum Migrationsstatus der Schülerinnen und Schüler (in Deutschland geboren ja/nein), bei der fehlende Angaben in einer Welle, ersetzt wurden mit gültigen Angaben aus den anderen Wellen. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_gert_g1_h	Ursprungsvariable(n): t_gert_g1
Beschreibung: Die Variable harmonisiert die Angaben zum Migrationsstatus der Schülerinnen und Schüler über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: -	
Syntax:	

Variablenname: t_germ_g1	Ursprungsvariable(n): t_germ
Beschreibung: Variable zum Migrationsstatus der Mütter der Schülerinnen und Schüler (in Deutschland geboren ja/nein), bei der fehlende Angaben in einer Welle, ersetzt wurden mit gültigen Angaben aus den anderen Wellen. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_germ_g1_h	Ursprungsvariable(n): t_germ_g1
Beschreibung: Die Variable harmonisiert die Angaben zum Migrationsstatus der Mütter der Schülerinnen und Schüler über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: -	
Syntax:	

Variablenname: t_germ_s_g1	Ursprungsvariable(n): t_germ_s
Beschreibung: Codierte Angaben zum Geburtsland der Mutter. Die offenen Angaben wurden in Zahlenwerte nach ISO 3166 numerisch überführt.	
Anmerkungen: -	
Syntax: -	

Variablenname: t_gerf_g1	Ursprungsvariable(n): t_gerf
Beschreibung: Variable zum Migrationsstatus der Väter der Schülerinnen und Schüler (in Deutschland geboren ja/nein), bei der fehlende Angaben in einer Welle, ersetzt wurden mit gültigen Angaben aus den anderen Wellen. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_gerf_g1_h	Ursprungsvariable(n): t_gerf_g1
Beschreibung: Die Variable harmonisiert die Angaben zum Migrationsstatus der Väter der Schülerinnen und Schüler über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: -	
Syntax:	

Variablenname: t_gerf_s_g1	Ursprungsvariable(n): t_gerf_s
Beschreibung: Codierte Angaben zum Geburtsland des Vaters. Die offenen Angaben wurden in Zahlenwerte nach ISO 3166 numerisch überführt.	
Anmerkungen: -	
Syntax: -	

Variablenname: t_mig_g1	Ursprungsvariable(n): t_gert_g1_h, t_germ_g1_h, t_gerf_g1_h
Beschreibung: Variable zum Migrationshintergrund der Schülerinnen und Schüler. Wenn angegeben wurde, dass beide Elternteile in Deutschland geboren sind, wird ein deutscher Migrationshintergrund vergeben. Wenn angegeben wurde, dass ein Elternteil aus einem anderen Land als Deutschland kommt, wurde ein einseitiger Migrationshintergrund vergeben. Wenn für beide Elternteile angegeben wurde, dass sie aus einem anderen Land kommen, wurde ein zweiseitiger Migrationshintergrund vergeben.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax: <pre>gen t_mig_g1 = 1 if t_gert_g1_h==1 & t_germ_g1_h==1 & t_gerf_g1_h==1 replace t_mig_g1 = 1 if t_gert_g1_h==2 & t_germ_g1_h==1 & t_gerf_g1_h==1 replace t_mig_g1 = 2 if t_gert_g1_h==2 t_germ_g1_h==2 t_gerf_g1_h==2 replace t_mig_g1 = 3 if t_gert_g1_h==2 & t_germ_g1_h==2 & t_gerf_g1_h==2 replace t_mig_g1 = 3 if t_gert_g1_h==1 & t_germ_g1_h==2 & t_gerf_g1_h==2 replace t_mig_g1 = .g if missing(t_gert_g1_h) missing(t_germ_g1_h) missing(t_gerf_g1_h)</pre>	

Variablenname: t_mig_g2	Ursprungsvariable(n): t_mig_g1
Beschreibung: Größere Kategorisierung der Variable zum Migrationshintergrund. Hierbei wird nun noch unterschieden zwischen einem vorhandenen („1“) und einem nicht-vorhandenen („0“) Migrationshintergrund.	
Anmerkungen: -	
Syntax: <pre>gen t_mig_g2 = t_mig_g1 recode t_mig_g2 (1=0) (2/3=1)</pre>	

Variablennamen: t_f001_1_h - t_f001_9_h	Ursprungsvariable(n): t_f001_w1_1 - t_f001_w1_7, t_f001_w2_1 - t_f009_w9_1
Beschreibung: Harmonisierte Angaben zu den Personen, mit denen die Schülerinnen und Schüler zusammenleben. In Welle 2 wurden zwei Kategorien neu hinzugefügt. Die hier generierten Variablen integrieren beide Versionen der Frage. Die Kategorien sind in Welle 1 „designbedingt fehlend“.	
Anmerkungen: Für Analysen empfehlen wir die Verwendung dieser statt der Ursprungsvariablen.	
Syntax (exemplarisch): <pre>gen t_f001_1_h = . replace t_f001_1_h = t_f001_w1_1 if wave==1 replace t_f001_1_h = t_f001_w2_1 if wave!=1</pre>	

Variablenname: t_f002_2_g1	Ursprungsvariable(n): t_f002_2
Beschreibung: Variable zur Anzahl älterer Geschwister der Schülerinnen und Schüler, bei der fehlende Angaben in einer Welle, ersetzt wurden mit gültigen Angaben aus den anderen Wellen. Dabei wurde jeweils die als letztes gemachte Angabe verwendet.	
Anmerkungen: -	
Syntax:	

Variablenname: t_f002_2_g1_h	Ursprungsvariable(n): t_f002_2_g1
Beschreibung: Die Variable harmonisiert die Angaben zur Anzahl älterer Geschwister der Schülerinnen und Schüler über die Wellen hinweg. Dabei wird für jede befragte Person jeweils die als letztes gemachte Angabe verwendet und die vorherigen Angaben überschrieben.	
Anmerkungen: -	
Syntax:	

Variablenname: t_FAS_tot_g1	Ursprungsvariable(n): t_f004, t_f005, t_f007, t_f008
Beschreibung: Family Affluence Scale (FAS) zur Messung des sozioökonomischen Status bei Kindern und Jugendlichen (siehe z. B. Hobza et al. 2017, https://www.mdpi.com/1660-4601/14/12/1540). Der Index summiert die wellenspezifisch z-standardisierten Werte der vier Items auf. Weist eines der Items einen fehlenden Wert auf, wird auch der FAS ein fehlender Wert zugewiesen.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): egen fasmiss = rowmiss(t_f004 t_f005 t_f007 t_f008) if wave==1 sum t_f004 if wave==1 & fasmiss==0 gen zt_f004 = (t_f004 - r(mean))/(r(sd)) if wave==1 & fasmiss==0 sum t_f005 if wave==1 & fasmiss==0 gen zt_f005 = (t_f005 - r(mean))/(r(sd)) if wave==1 & fasmiss==0 sum t_f007 if wave==1 & fasmiss==0 gen zt_f007 = (t_f007 - r(mean))/(r(sd)) if wave==1 & fasmiss==0 sum t_f008 if wave==1 & fasmiss==0 gen zt_f008 = (t_f008 - r(mean))/(r(sd)) if wave==1 & fasmiss==0 gen t_FAS_tot_g1 = zt_f004 + zt_f005 + zt_f007 + zt_f008	

Variablenname: t_FAS_5_g1	Ursprungsvariable(n): t_FAS_tot_g1
Beschreibung: Kategorisierung der FAS in wellenspezifische Quintile.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): xtile t_FAS_5_g1 = t_FAS_tot_g1 if wave==1, nq(5)	

Variablenname: t_FAS_3_g1	Ursprungsvariable(n): t_FAS_tot_g1
Beschreibung: Kategorisierung der FAS in wellenspezifische Terzile.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): xtile t_FAS_3_g1 = t_FAS_tot_g1 if wave==1, nq(3)	

Variablenname: t_month_allowance_g1	Ursprungsvariable(n): t_c005, t_c006, t_c007
Beschreibung: Aus den Angaben zur Häufigkeit und Höhe des ausgezahlten Taschengelds, wurde die Höhe des monatlichen Taschengelds berechnet. Dabei wurde von vier Wochen pro Monat ausgegangen.	
Anmerkungen: -	
Syntax: gen t_month_allowance_g1 = . replace t_month_allowance_g1 = t_c007 if !missing(t_c007) replace t_month_allowance_g1 = t_c006*4 if !missing(t_c006) replace t_month_allowance_g1 = 0 if t_c005==4 replace t_month_allowance_g1 = .a if t_c005==0 t_c005==3 replace t_month_allowance_g1 = .b if t_c005==.b replace t_month_allowance_g1 = .e if (t_c005==1 & t_c006==.e) (t_c005==2 & t_c007==.e) replace t_month_allowance_g1 = .e if t_c005==.e & t_c006==.e & t_c007==.e	

Variablenname: t_prob_g1	Ursprungsvariable(n): externe Quellen
Beschreibung: Wahrscheinlichkeit Schule, Teil des Samples zu sein. Diese ergibt sich aus dem Stichprobendesign einer disproportional geschichteten Zufallsstichprobe mit den Schichtungsmerkmalen Bundesland, Schulgröße, Kreis, Schulform und Urbanisierung (siehe Zins 2014 in den Dokumentationsmaterialien zur GUS Studie).	
Anmerkungen: -	
Syntax: -	

Variablenname: t_dweight_g1	Ursprungsvariable(n): t_prob_g1
Beschreibung: Designgewicht um die disproportionale Schichtung der Stichprobe auszugleichen.	
Anmerkungen: -	
Syntax: gen t_dweight_g1 = 1/t_prob_g1	

Variablenname: t_misscount_g1	Ursprungsvariable(n): Alle erhobenen Items im Schüler*innenfragebogen
Beschreibung: Anzahl der nicht-beantworteten Items im Fragebogen (Befragung der Schülerinnen und Schüler). Es werden alle Variablen mit Missing-Codes für „unplausibler Wert“, „weiß nicht“ und „keine Angabe“ gezählt.	
Anmerkungen: -	
Syntax: quietly: mvencode _all, mv(.c=99998\d=99997\,e=99996) egen t_misscount_g1 = anycount(_all), values(99998 99997 99996)	

Variablenname: t_missratio_g1	Ursprungsvariable(n): t_misscount_g1
Beschreibung: Anteil der nicht-beantworteten Items im Fragebogen an der Gesamtzahl der in der Welle erfragten Items (Befragung der Schülerinnen und Schüler). Überfilterte Fragen zählen nicht zur Gesamtzahl.	
Anmerkungen: Die dokumentierte Syntax wurde zu einem frühen Zeitpunkt der Datenaufbereitung ausgeführt. Bei einer Berechnung der Gesamtzahl der Variablen im Nachhinein müssen zunächst weitere Variablen (insb. die hier beschriebenen, generierten Variablen) von der Zählung ausgeschlossen werden.	
Syntax (exemplarisch für Welle 1): keep if wave==1 mvencode _all, mv(.a=99999) egen filter = anycount(_all), values(99999) quietly: describe t_* gen totalvar = r(k) - filter - 1 gen t_missratio_g1 = t_misscount_g1/totalvar	

Variablenname: t_classcount_g1	Ursprungsvariable(n): ID_c, wave
Beschreibung: Anzahl der befragten Schülerinnen und Schüler in der Klasse in einer Welle.	
Anmerkungen: -	
Syntax: egen t_classcount_g1 = count(1), by(ID_c wave) replace t_classcount_g1 = . if ID_c==.	

Variablenname: t_ampm_g1	Ursprungsvariable(n): t_device, externe Referenzliste
Beschreibung: Die Variable gibt an, ob die Befragung mit einem Tablet durchgeführt wurde, das bei Zeitangaben eine Unterscheidung in vormittags („am“) und nachmittags („pm“) verlangte.	
Anmerkungen: In den Wellen 1 und 2 gab es in der Befragungssoftware noch keine Möglichkeit Uhrzeiten strukturiert zu erfassen. Daher wurde die Variable in diesen Wellen auf „designbedingt fehlend“ gesetzt.	
Syntax: import excel using "\${reflists}/AM-PM.xlsx", firstrow clear tempfile ampm save `ampm' use "GUS_Data_D_1-0.dta", clear merge m:1 t_device using `ampm', keep(master match) nogen replace am_pm = .b if wave==1 wave==2 rename am_pm t_ampm_g1	

Variablenname: t_schoolinj_g1	Ursprungsvariable(n): t_as001, t_as003, t_as005_1
Beschreibung: Dummy-Variable, die angibt, ob in den letzten 12 Monaten (oder seit der letzten Befragung) eine Schulverletzung erlitten wurde. Dabei wurden zwei Kontrollvariablen berücksichtigt, die sicherstellen sollen dass a) die berichtete Verletzung auch wirklich im Schulkontext passiert ist und b) die Verletzung eine Behandlung durch eine Ärztin oder einen Arzt erforderte. Traf eine dieser Bedingungen nicht zu, so wurde die Schülerin bzw. der Schüler als nicht schulverletzt eingestuft. Wurden bei einer der drei Ursprungsvariablen keine Angaben gemacht, wurde die Variable auf „nicht ermittelbar“ gesetzt.	
Anmerkungen: -	
Syntax: generate t_schoolinj_g1 = . replace t_schoolinj_g1 = 1 if (t_as001 == 2 t_as001 == 3 t_as001 == 4) & /// (t_as003 == 1 t_as003 == 2 t_as003 == 3 t_as003 == 4 t_as003 == 5 t_as003 == 6) & /// t_as005_1 == 1 replace t_schoolinj_g1 = 0 if t_as001 == 1 replace t_schoolinj_g1 = 0 if (t_as001 == 2 t_as001 == 3 t_as001 == 4) & /// (t_as003 == 7 t_as005_1 == 0) replace t_schoolinj_g1 = .g if t_as001 == .e t_as003 == .e t_as005_1 == .e	

Variablenname: t_sc_tot_g1	Ursprungsvariable(n): t_f003_1, t_f003_2, t_f003_3, t_f003_4, t_s015_1, t_s015_2, t_s015_3, t_c001_1, t_c001_2, t_c001_3
Beschreibung: Sozialkapitalindex zur Messung der Qualität der sozialen Beziehungen der Schülerinnen und Schüler (siehe Klocke & Stadtmüller 2019 https://doi.org/10.1007/s12187-018-9583-y). Der Index summiert die wellenspezifisch z-standardisierten Werte der verwendeten Items auf. Zwei davon beziehen sich auf a) die Familie, drei auf b) das Schulklima und drei weitere auf c) die Nachbarschaft, in der die Schülerinnen und Schüler leben. Damit alle drei Bereiche mit gleicher Bedeutsamkeit in den Index eingehen, wurde der Bereich Familie höher gewichtet (Faktor 1,5).	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): Analog zu t_FAS_tot_g1 (ausgenommen Umpolung von Items und Gewichtung).	

Variablenname: t_sc_5_g1	Ursprungsvariable(n): t_sc_tot_g1
Beschreibung: Kategorisierung des Sozialkapitalindex in wellenspezifische Quintile.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): xtile t_s_5_g1 = t_s_tot_g1 if wave==1, nq(5)	

Variablenname: t_sc_3_g1	Ursprungsvariable(n): t_sc_tot_g1
Beschreibung: Kategorisierung des Sozialkapitalindex in wellenspezifische Terzile.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): xtile t_sc_3_g1 = t_sc_tot_g1 if wave==1, nq(3)	

Variablenname: t_linkage_g1	Ursprungsvariable(n): ID_t_g1, ID_t_g2
Beschreibung: Die Variable zeigt an, ob bei einer Schülerin bzw. einem Schüler im Verlauf der Studie mindestens ID-Match in ID_t_g2 durch das Record Linkage Verfahren zustande gekommen ist. Dabei wird jeweils die ID vor dem Record Linkage (ID_t_g1) mit der ID nach dem Record Linkage (ID_t_g2) verglichen. Gibt es dabei Unterschiede, so werden auch die anderen Fälle, die dem Schulkind zuzuordnen sind mit „1“ markiert.	
Anmerkungen: -	
Syntax: egen t_linkage_g1 = diff(ID_t_short ID_t_short_g*) replace t_linkage_g1 = 1 if t_linkage_g1[_n-1]==1 & ID_st_short_g2[_n-1]==ID_st_short_ replace t_linkage_g1 = 1 if t_linkage_g1[_n+1]==1 & ID_st_short_g2[_n+1]==ID_st_short_g2	

Variablenname: t_recourse_g1	Ursprungsvariable(n): ID_t_g1, wave
Beschreibung: Die Variable zeigt an, in welchen Wellen eine Schülerin bzw. ein Schüler an der Studie teilgenommen hat. Der Wert „111111“ steht dabei für eine Teilnahme in allen sechs Erhebungswellen. Eine Teilnahme nur in Welle 4 ist bspw. mit „...1..“ gekennzeichnet. Die Variable bezieht sich auf die IDs ohne Record Linkage.	
Anmerkungen: -	
Syntax: xtset ID_t_g1 wave xtpatternvar, gen(t_recourse_g1)	

Variablenname: t_recourse_g2	Ursprungsvariable(n): ID_t_g2, wave
Beschreibung: Die Variable zeigt an, in welchen Wellen eine Schülerin bzw. ein Schüler an der Studie teilgenommen hat. Der Wert „111111“ steht dabei für eine Teilnahme in allen sechs Erhebungswellen. Eine Teilnahme nur in Welle 4 ist bspw. mit „...1..“ gekennzeichnet. Die Variable bezieht sich auf die IDs nach Record Linkage.	
Anmerkungen: -	
Syntax: xtset ID_t_g2wave xtpatternvar, gen(t_recourse_g2)	

Variablenname: recourse_s_g1	Ursprungsvariable(n): ID_s, t_recourse_g2
Beschreibung: Die Variable zeigt an, in welchen Wellen die Schulen an der Studie teilgenommen haben. Der Wert „111111“ steht dabei für eine Teilnahme in allen sechs Erhebungswellen, während bspw. eine Teilnahme nur in Welle 4 mit „...1..“ gekennzeichnet ist. Als Basis gilt das Vorliegen mindestens eines Fragebogens für Schülerinnen und Schüler in der betreffenden Welle.	
Anmerkungen: -	
Syntax (exemplarisch für Welle 1): gen str recourse_s_g1 = "....." quietly: fre ID_s foreach school in `r(lab_valid)' { quietly: count if usubstr(t_recourse_g2,1,1)=="1" & ID_s=="school" quietly: replace recourse_s_g1 = "1" + usubstr(recourse_s_g1,2,.) if r(N)>0 & r(N)<. /// & ID_s=="school" }	

Variablenname: t_fromschool_g1	Ursprungsvariable(n): -
Beschreibung: In den Wellen 1 und 2 gab es unterschiedliche Fragebogenversionen in Bezug auf den Schulweg (Fragen t_s006, t_s007 und t_s008). Da sich Hin- und Rückweg potenziell unterscheiden können, wurde in diesen Wellen rund die Hälfte der Schülerinnen und Schüler nach dem Hinweg und die andere Hälfte nach dem Rückweg gefragt. Die Variable gibt an, zu welcher Gruppe die Teilnehmenden jeweils gehörten. Ab Welle 3 wurde nur noch nach dem Hinweg gefragt.	
Anmerkungen: -	
Syntax: -	