



Kinder- und Jugendgesundheitssurvey 2003-2006

Public Use File

KiGGS03_06

Dokumentation des Datensatzes

Robert Koch-Institut
FG 21 - Epidemiologisches Datenzentrum
Postfach 650261
13302 Berlin
Tel.: 030-18754-3320
E-Mail: EDZ@rki.de

Inhaltsübersicht

1.	Der Kinder- und Jugendgesundheitssurvey (KiGGS).....	4
2.	Module der KiGGS-Studie	4
3.	Untersuchungsteile/Datenquellen	4
4.	Datenprüfung und Korrekturen	5
5.	Aufbau des Datensatzes.....	5
5.1.	Nomenklatur	5
5.2.	Fehlende Werte	6
5.3.	Umgang mit Filterführung	7
5.4.	Umgang mit Fragen, bei denen Mehrfachantworten möglich sind	7
5.5.	Technische Anmerkungen zur SAS-Datei	8
5.6.	Zentrale Variablen	8
5.7.	Anmerkungen zu (neuen) Variablen der Fragebögen.....	9
5.8.	Anmerkungen zu Messbogenvariablen	12
5.9.	Der Sozialschicht-Index nach Winkler (2003/4).....	12
5.10.	Impfvariablen	18
5.11.	Neuvariablen zur Arzneimittelanwendung	18
5.12.	Variablen des Ernährungsfragebogens	19
5.13.	Die Laborvariablen.....	21
6.	Auswertung von KiGGS-Daten	23
6.1.	Altersspezifische Erhebungsinstrumente	23
6.2.	Stichprobengewichtung.....	23
6.3.	Ost-West-Vergleiche.....	26
6.4.	Varianzschätzung, Berechnung von p-Werten und Konfidenzintervallen.....	26

1. Der Kinder- und Jugendgesundheitssurvey (KiGGS)

Der Kinder- und Jugendgesundheitssurvey ist eine bevölkerungsrepräsentative Erhebung, deren Zielpopulation die in den Melderegistern der Bundesrepublik mit Hauptwohnsitz registrierte deutsche und ausländische Wohnbevölkerung der 0- bis 17-Jährigen ist. Die Baseline-Erhebung wurde von Mai 2003 bis Mai 2006 (3 Untersuchungsjahre) durchgeführt. Die Stichprobenziehung erfolgte über ein komplexes zweistufiges Stichprobenverfahren mit ungleichen Auswahlwahrscheinlichkeiten. Auf der ersten Stufe wurden 150 Sample Points (Untersuchungsorte) gezogen, wobei Gemeinden im Osten Deutschlands überproportional vertreten waren, um die Power für Ost-West-Vergleiche zu erhöhen. Von diesem Oversampling Ost abgesehen, bilden die Sample Points nach Gemeindetypen (BIK-Klassifikation) und räumlicher Verteilung die Struktur der Bundesrepublik ab. Die Ziehung der Gemeinden erfolgte proportional zur Zahl der Kinder und Jugendlichen von 0–17 Jahren. Die Ziehung wurde vom Zentrum für Umfragen, Methoden und Analysen (GESIS-ZUMA) in Mannheim durchgeführt. Auf der zweiten Stufe wurde innerhalb der Sample Points eine feste Zahl von Kindern und Jugendlichen pro Jahrgang gezogen. Insgesamt konnte bei einer Response von 66,6% eine Gesamtteilnehmerzahl von 17.641 (8.985 Jungen, 8.656 Mädchen) erreicht werden. Mit Hilfe besonderer Maßnahmen wie z.B. Oversampling ist es gelungen, ausländische Kinder und Jugendliche in einem Anteil an der Studie zu beteiligen, der ihrem Anteil an der Wohnbevölkerung entspricht. Der Routenplan für die Durchführung der Untersuchung in den einzelnen Gemeinden folgte einem Konzept, das zufällige regionale aber auch saisonale Einflüsse weitgehend ausschließen sollte. Die KiGGS-Basiserhebung ist repräsentativ für die Wohnbevölkerung der 0- bis 17-Jährigen in Deutschland für die Jahre 2003-2006, der Alters- und Geschlechtsaufbau sowie die Gewichtung der Regionen Ost/West/Berlin entspricht dem Stand vom 31.12.2004 (siehe dazu Publikationen in den Ordnern LIT2/LIT3).

Die englische Bezeichnung der Studie lautet „The German Interview and Examination Survey for Children and Adolescents (KIGGS)“.

Der Kinder- und Jugendgesundheitssurvey ist eine Querschnittserhebung. Bei der Interpretation der Ergebnisse ist daher zu beachten, dass aus den KiGGS-Daten Zusammenhänge zwischen Merkmalen abgeleitet werden können, aber i.A. keine kausalen Schlüsse gezogen werden können, da Querschnittsdaten keine Aussage über die zeitliche Abfolge zwischen potenzieller Ursache und potenzieller Wirkung erlauben.

2. Module der KiGGS-Studie

Die KiGGS-Studie setzt sich zusammen aus der Basiserhebung (Kernsurvey) und fünf weiteren Modulen:

- Kinder-Umwelt-Survey (KUS)
- Modul Psychische Gesundheit (Bella-Studie)
- Motorik-Modul (MoMo)
- Landesmodul Schleswig-Holstein
- Ernährungsmodul (EsKiMo, 2006)

Diesen Modulen liegen jeweils eine Teil- bzw. eine Aufstockungsstichprobe (im Landesmodul Schleswig-Holstein) zugrunde. Weitere Informationen dazu sind auf der KiGGS-Webseite (www.KiGGS.de unter: Die Studie – Organisation – Module) zu finden (siehe auch Publikationen im Ordner LIT2).

3. Untersuchungsteile/Datenquellen

KiGGS gliedert sich in einzelne Untersuchungsteile, die mit verschiedenen Erhebungsinstrumenten erhoben bzw. dokumentiert wurden. In der KiGGS-Datei sind die

Variablen aus den einzelnen Untersuchungsteilen zusammengefasst. Die Datei enthält alle Probanden, unabhängig davon, ob alle relevanten Untersuchungsteile vollständig vorliegen oder nicht.

Im Einzelnen liegen folgende Untersuchungsteile vor:

- Anthropometrische Messung und medizinische Untersuchung: Die Ergebnisse sind im Messbogen (MB) dokumentiert.
- Computergestütztes ärztliches Interview (CAPI) inkl. Medikamentenanamnese der letzten 7 Tage
- Eltern- und Kinderfragebögen inkl. Food Frequency Questionnaire (FFQ)
 - Die Eltern (in einigen Fällen auch andere Bezugspersonen) haben den Elternfragebogen (EFB) ausgefüllt, dessen Fragen mit der Altersgruppe variiert.
 - Kinder und Jugendliche von 11-17 Jahren haben einen eigenen Kinderfragebogen (KFB) ausgefüllt.
 - Ausländer mit nur geringen Deutschkenntnissen hatten die Möglichkeit, eine verkürzte Version des Elternfragebogens in ihrer Muttersprache auszufüllen (Ausländerfragebogen AFB 0-17 Jahre).
 - Für Ausländer mit geringer Teilnahmebereitschaft gab es darüber hinaus eine nochmals gekürzte Version sowohl des Eltern- als auch des Kinderfragebogens (Ausländerkurzfragebogen, sowohl für EFB als auch für KFB).

Die Angaben aus den Ausländer- bzw. Ausländerkurzfragebögen sind im Datensatz enthalten.

- Laboruntersuchungen (Urin- und Blutparameter)

4. Datenprüfung und Korrekturen

Die KiGGS-Daten wurden einer umfangreichen Konsistenzprüfung unterzogen. Die Datenprüfung umfasste formale ebenso wie inhaltliche Aspekte, wobei bei inhaltlichen Fragen ggf. der oder die für den jeweiligen Fragebogenteil Verantwortliche an der Entscheidung, wie mit Implausibilitäten umzugehen ist, beteiligt wurde. Zur Gewährleistung eines einheitlichen Vorgehens bei Datenänderungen wurden allgemeine Regeln zur Datenprüfung und -korrektur abgesprochen. Zudem wurden alle Editieranweisungen in einer Datenbank dokumentiert. Datenänderungen geschahen über eine Reihe von Korrektursyntaxen, so dass alle Änderungen nachvollziehbar und dokumentiert sind.

Das Ergebnis der Datenqualitätssicherung ist ein geprüfter, aber nicht vollständig widerspruchsfreier Datensatz. Vollständige Widerspruchsfreiheit ließe sich nur durch ein übermäßiges Glätten des Datensatzes mit der Gefahr der Verfälschung erreichen. Prinzipiell wurden keine Abgleiche zwischen verschiedenen Fragebogenteilen durchgeführt, etwa zwischen Eltern- und Kinderfragebogen, da die unterschiedlichen Sichtweisen von Kindern und Eltern selbst einen Aspekt der in KiGGS untersuchten Phänomene darstellen. Bei stetigen Variablen ist von Nutzern der KiGGS-Daten zu beachten, dass nicht alle Extremwerte („Ausreißer“) aus den Daten eliminiert wurden. Aufgabe der Datenqualitätssicherung ist es, Werte zu eliminieren, die mit hoher Wahrscheinlichkeit falsch sind. Die Frage, ob ein einzelner extremer, aber plausibler Wert etwa eine Regressionsanalyse über Gebühr beeinflusst, ist bei der Auswertung zu überprüfen.

5. Aufbau des Datensatzes

5.1. Nomenklatur

Variablenamen sollten möglichst so bezeichnet werden, dass sie ungekürzt in verschiedenen Auswertungsprogrammen (incl. deren Altversionen) verwendet werden können und in den gelabelten Erhebungsinstrumenten schnell auffindbar sind. Diese Vorgaben wurden in

Anbetracht der Menge und Verschiedenheit der im KiGGS eingesetzten Fragebögen so gut wie möglich umgesetzt.

Der Variablenname beginnt je nach Befragungs-Instrument mit einem Kennbuchstaben¹. Dann folgen eine 3-stellige Nummerierung, die sich nach der laufenden Fragenummer des Eltern-Fragebogens (EFB) für die 3- bis 6-Jährigen richtet, sowie eventuell weitere alphanumerische Zeichen. Die Nummerierung setzt sich bei den Items/Variablen fort, die nicht bei diesem als Standard dienenden Fragebogen vorkommen. Gibt es die gleiche Frage sowohl im EFB als auch im KFB (Kinder-Fragebogen), ist zwar der Anfangsbuchstabe verschieden, jedoch die Ziffernfolge identisch (z.B. E019 und K019). Weiteres zu den Fragebögen nun in Kurzform:

Kennungen:

E	Variable des Eltern- bzw. Ausländerfragebogens
K	Variable des Kinderfragebogens
C	Variable des CAPI (Teil Krankheiten)
I	Variable des CAPI (Teil Impfungen)
FQ	Variable des Ernährungsfragebogens (Eltern und Kinder)

Zusatzzeichen:

a, b	Unterfragen
M	Frage an die Mutter; beim CAPI: Monatsangabe
V	Frage an den Vater
J	Beim CAPI: Jahresangabe
t(x)	Freitext (lfd. Nr.)
c(x)	Freitext in Codeform (lfd. Nr.)
z(x)	Numerische Angabe wie Anzahl, Menge usw. (lfd. Nr.)
m(x)	Teil eines Mehrfachantwortenblocks (lfd. Nr.)

Allein schon wegen der strengen Vorgabe, für die Variablennamen maximal 8 Zeichen zu verwenden, kann es hier Ausnahmen von den Regeln geben; dies ist auch der Fall bei neugebildeten Variablen wie z.B. Aggregat- und Indexvariablen. Die Namensgebungen für die Laborparameter bzw. für die verschiedenen Messwerte beruhen zumeist auf tradierten Kurzbezeichnungen früherer Erhebungen.

Auch bei den Wertelabeln wurde eine Vereinheitlichung angestrebt. So gilt für alle Ja/Nein-Variablen „1 = Ja“ und „2 = Nein“. Bei ordinalskalierten Variablen sind die Wertelabel auf- bzw. absteigend vergeben worden.

5.2. Fehlende Werte

Zur Markierung spezieller fehlender Werte werden User-definierte Missings verwendet. In SPSS sind dabei –7, –8 und –9 als Missing erklärt. In SAS werden .N, .M und .W verwendet. Im Einzelnen haben die speziellen Missings folgende Bedeutungen:

5.2.1. Spezielles Missing –7 = .N = Nicht zutreffend für das Alter

Nicht alle Fragen bzw. Untersuchungsteile wurden allen Altersgruppen gestellt. Die sich daraus ergebenden Missings in den jeweils nicht zu dieser Variable befragten bzw. untersuchten Altersgruppen sind durch –7 bzw. .N gekennzeichnet.

5.2.2. Spezielles Missing –8 = .M = Missing gesetzt

Extremwerte, die als unplausibel gelten müssen, wurden bei der Datenprüfung auf –8 bzw. .M gesetzt.

¹ Nur beim FQ sind es zwei Kennbuchstaben.

5.2.3. Spezielles Missing –9 = .W = Weiß nicht

Bei einigen stetigen Variablen (z.B. Geburtsgewicht) gab es im Fragebogen die Möglichkeit, die Frage mit „Weiß nicht“ zu beantworten. Diese Antworten sind als –9 bzw. .W im Datensatz enthalten.

Bei kategoriellen Variablen mit der Antwortmöglichkeit „Weiß nicht“ wurde dieses spezielle Missing nicht verwendet, sondern „Weiß nicht“ als gültige Antwortmöglichkeit kodiert.

Alle anderen fehlenden Werte sind als normales System-Missing kodiert. Das gilt z.B. auch für fehlende Werte, die sich durch Filterfragen ergeben (siehe Absatz 5.3 „Umgang mit Filterführung“).

Beim Arbeiten mit dem Datensatz ist zu beachten, dass die Abfrage nach System-Missing (SYSMIS in SPSS) bzw. nach „.“ (in SAS) die User-definierten Missings nicht mit erfasst.

In SPSS werden beide Arten von fehlenden Werten gemeinsam mit der Funktion MISSING(Variable) angesprochen. In SPSS ist zudem zu beachten, dass bei der Bildung neuer Variablen aus Variablen mit User-definierten Missings mit dem COMPUTE-Befehl die Werte –7, –8 und –9 als System-Missings übernommen werden. Verwendet man dagegen den Befehl RECODE mit dem Unterbefehl COPY (etwa in der Form ELSE=COPY), werden die Werte –7, –8 und –9 zu normalen Variablenwerten umkodiert. Will man dieses verhindern, müssen die User-Missings mit dem Befehl MISSING VALUES Variable (-7,-8,-9) zu User-definierten Werten erklärt werden. Dabei ist zu beachten, dass für jede Variable nur der letzte MISSING VALUES-Befehl gültig ist. Mit RECODE Variable (MISSING=SYSMIS) können alle User-Missings unproblematisch in normale System-Missings umgewandelt werden.

In SAS müssen Abfragen der Form `variable = .` ersetzt werden durch `variable <= .Z`, da die Sortierreihenfolge für numerische Variablen lautet: `._` (Punkt Unterstrich, der kleinste Wert),

`.` (Punkt, das normale Missing), `.A` bis `.Z` (Punkt A bis Punkt Z), negative Zahlen, Null, positive Zahlen. Entsprechend müssen Abfragen der Form `variable ne .` ersetzt werden durch `variable > .Z`.

Die bei einigen Laborvariablen vorkommenden Wertelabel „99999,0“ bzw. „-99999,0“ sind nicht als fehlende Werte zu interpretieren, sondern markieren Messwerte jenseits der jeweiligen technischen Messgrenzen.

5.3. Umgang mit Filterführung

Falls die Eingangsfrage mit "Nein" beantwortet ist, stehen die Unterfragen auf Missing.

Bsp. - Für Nicht-Raucher steht die Zahl der gerauchten Zigaretten auf Missing (nicht auf Null).
 - Wenn die Frage E026 („Ist Ihr Kind dauerhaft schwerhörig?“) mit „Nein“ beantwortet ist, steht die Unterfrage „Wenn ja, hat Ihr Kind ein Hörgerät?“ auf Missing (nicht auf „Nein“).
 - Wenn keine der beiden Fragen CA01 (Heuschnupfen) und CA02 (Neurodermitis etc.) mit „Ja“ beantwortet ist, stehen die Unterfragen Allergie (C17 bis C19a) auf Missing.

Falls die Unterfragen positiv beantwortet waren, aber die Eingangsfrage verneint war, wurde die Eingangsfrage auf "Ja" korrigiert.

5.4. Umgang mit Fragen, bei denen Mehrfachantworten möglich sind

Bei einigen Fragen ist es erlaubt, mehr als eine Antwortkategorie anzukreuzen (z.B. E018M1-E018M6, Probleme bei der Geburt). In diesem Fall wurden alle nicht angekreuzten Antwortmöglichkeiten durch eine sog. technische Null gekennzeichnet (d.h. eine Null ohne Label). Falls die Frage gar nicht beantwortet wurde, stehen alle Antwortmöglichkeiten auf Missing.

Es gibt Fragebatterien, bei denen jede einzelne Frage mit „Ja“ oder „Nein“ zu beantworten ist (z.B. E03301-E03311, Schmerzen). In diesem Fall wurden keine Korrekturen durchgeführt, wenn einzelne Fragen nicht beantwortet wurden.

5.5. Technische Anmerkungen zur SAS-Datei

Beim Arbeiten mit der SAS-Datei ist unter anderem der Umgang mit den benutzerdefinierten Missings zu beachten, wie bereits oben beschrieben.

Die SAS-Formate sind im Formatkatalog *kiggs03_06.sas7bcat* abgelegt. Damit SAS auf die Formate zugreifen kann, ist folgende Syntaxzeile erforderlich:

```
OPTIONS FMTSEARCH=(WORK libname.kiggs03_06);
```

Der Libname muss dabei auf das Verzeichnis verweisen, in dem der Formatkatalog abgelegt ist. Alternativ kann der Formatkatalog umbenannt werden in *formats.sas7bcat*. Wenn dann dem Verzeichnis, in dem der Formatkatalog abgelegt ist, der Libname *library* zugewiesen wird, findet SAS die Formate automatisch.

Neben der SAS-Datei (*kiggs03_06.sas7bdat*) und dem Formatkatalog (*kiggs03_06.sas7bcat*) wird auch eine Indexdatei (*kiggs03_06.sas7bndx*) mit ausgeliefert. Die Indexdatei muss im selben Verzeichnis abgelegt werden wie die SAS-Datei. Es sind verschiedene Indizes definiert, die über PROC CONTENTS angesehen werden können. Die Indizes erlauben das Arbeiten mit der SAS-Datei, so als ob sie gleichzeitig nach unterschiedlichen Kriterien sortiert wäre. So kann man etwa „BY alt sexa“ auf die Datei zugreifen, aber auch „BY sexa alt“, ohne dass eine vorherige Sortierung erforderlich wäre. Dies geht deshalb, weil sowohl ein Index für die Sortierreihenfolge (ALT SEXA) als auch ein Index für die Sortierreihenfolge (SEXA ALT) definiert ist. Der Zugriff auf die Indizes geschieht automatisch; der Indexname muss nicht angegeben werden.

5.6. Zentrale Variablen

An dieser Stelle sollen – beginnend mit den zentralen Variablen - insbesondere die Variablen, die nicht in den mit Labeln versehenen Fragebögen vorkommen (siehe Ordner FB), kurz erläutert werden. In einigen Fällen wird auf die jeweils relevante Originalpublikation (siehe Ordner LIT2) mit ihren Sekundärzitationen hingewiesen. Viele der Neuvariablen sind in erster Linie als Angebot für die Auswertenden gedacht, zumal die „Originalvariablen“ im Datensatz belassen wurden.

5.6.1. Erhebungsinstrumente

Die Steuervariablen **MBTYP**, **EFBTYP** und **KFBTYP** geben an, welcher Messbogen (d.h. für welche Altersgruppe) und welcher Fragebogen zur Anwendung kamen (und ob überhaupt ein Fragebogen ausgefüllt wurde). Die Steuervariablen **FFQ** (Ernährungsfragebogen), **CAPI** (Arztinterview) und **LAB** (Blut- und/oder Urin-Werte) zeigen an, ob die jeweiligen Untersuchungsteile vorhanden sind.

5.6.2. Alter

Für das Alter zum Zeitpunkt der Erhebung stehen folgende Variablen zur Verfügung:

ALT	Alter in vollendeten Lebensjahren (0, 1, 2, ..., 17)
ALTK	exaktes Alter, taggenau berechnet, mit 3 Nachkommastellen (für Analysen mit Alter als stetiger bzw. metrischer Variable)
AGEGRP	Alter in den KiGGS-Altersgruppen 0-2, 3-6, 7-10, 11-13 und 14-17 Jahre (Datenerhebung und Untersuchungsprogramm variieren z.T. zwischen diesen Altersgruppen)

5.6.3. Geschlecht

In der Variable **SEX** ist das Geschlecht als 1 = männlich, 2 = weiblich gelabelt. In der Variablen **SEXA** lautet die Labelung: 1 = Jungen, 2 = Mädchen.

5.6.4. Migrantenvariablen

Die Angabe zur Staatsangehörigkeit des Kindes (laut Einwohnermeldeamt) wurde in die Datei übernommen (Variable **AUSL**). Für migrantenspezifische Auswertungen sollte jedoch die Variable **MIGRANT** verwendet werden.

Zur Konstruktion der Variablen **MIGRANT** wurde herangezogen a) das Geburtsland des Kindes, b) das Geburtsland von Vater und Mutter sowie c) die Staatsangehörigkeit von Vater und Mutter. Als Migranten werden Kinder und Jugendliche definiert, die selbst aus einem anderen Land zugewandert sind und von denen mindestens ein Elternteil nicht in Deutschland geboren ist, oder von denen beide Eltern zugewandert oder nichtdeutscher Staatsangehörigkeit sind. Kinder und Jugendliche mit einseitigem Migrationshintergrund, also Kinder, die in Deutschland geboren und von denen nur ein Elternteil aus einem anderen Land zugewandert und/oder nichtdeutscher Staatsangehörigkeit ist, werden den Nicht-Migranten zugerechnet. Bei alleinerziehenden Eltern gilt der Status des erziehenden Elternteils für das Kind. Sofern keine diesbezgl. Informationen im Elternfragebogen vorlagen, wurde auf die Angaben des Kindes zum Geburtsland von Vater und Mutter zurückgegriffen. Waren keinerlei Informationen zu sämtlichen genannten Variablen vorhanden, wurde die zu Hause gesprochene Sprache als Indikator für einen Migrationshintergrund berücksichtigt (LIT2: Schenk_Migration.pdf).

5.6.5. Räumlich-geografische Klassifikation

OW Ost-West-Einteilung, wobei Berlin zum Osten gezählt wird
STALA Stadt-Land-Einteilung anhand der Größe des Wohnorts
 (administrative Gemeindegröße)

Für Stadt-Land-Vergleiche hat sich die Variable **STALA** bewährt, die eine aggregierte Form der 7-stufigen politischen Gemeindegrößenklasse darstellt.

Folgende Regionen wurden definiert:

- | | |
|--------------------|------------------------------|
| 1. Ländlich | < 5.000 Einwohner |
| 2. Kleinstädtisch | 5.000 - < 20.000 Einwohner |
| 3. Mittelstädtisch | 20.000 - < 100.000 Einwohner |
| 4. Großstädtisch | 100.000 und mehr Einwohner |

5.6.6. Variablen für Analyse und Varianzschätzung

Die beiden Variablen **wKiGGS** (Stichprobengewichte) und **PPOINT** (Sample Point) werden für die Analyse benötigt (Erläuterung siehe 6.). Für die Analyse von Laborwerten gibt es die Gewichtungsvariablen **wKGSLab** und **wKGSVitD**.

5.7. Anmerkungen zu (neuen) Variablen der Fragebögen

Anstelle der erhobenen Freitexte (Endung = **t**) werden nur kodierte Variablen (Endung = **c**) ausgeliefert, deren (freier) Codeplan in Form von Wertelabeln beigefügt ist. Es handelt sich hier um die folgenden Neuvariablen:

E0152c1-3, E034c, E03910c, E03911c, E041c, E080c, E0864c, Ther4_c,
K099c, K03312c, K034c, K03910c, K03911c, K080c, C406c1-4, Se_c.

Freitexte zu Krankheiten wurden gemäß ICD10 vercodet; somit entstanden neu:

E031icd1-3, C15t1icd – C15t3icd, C33t1icd – C33t4icd.

Für die Vercodung der Staatsangehörigkeiten/Staaten wurde der Staatsangehörigkeits- und Gebietsschlüssel des Stat. Bundesamtes (2006) angewendet. Da z. T. auch zwei Angaben hierzu vorlagen, wurden generell 6 Stellen abgespeichert (2mal ein 3stelliger Schlüssel). Die entsprechenden Ergebnisvariablen lauten:

E007Mc, E007Vc, E008Mc, E008Vc, E010m2c, K008c, K008Mc, K008Vc.

Die Geschwisterzahl (E005x, E006x) umfasst auch Halbgeschwister und Kinder von neuen Partnern der Eltern.

Die Body-Mass-Indexwerte **MBMI** bzw. **VBMI** basieren auf subjektiven Angaben zur Körpergröße und Körpergewicht der Mutter bzw. des Vaters. Wie auch bei E0012xz und E013xxz beziehen sich die Angaben auf die sozialen Eltern, die nicht unbedingt auch die leiblichen Eltern sind (vgl. dazu E001, K001).

Die Variablen **E025az1** (Jahre) und **E025az2** (Monate) ersetzen zusammen E025az. Dies war notwendig, da bei den 0- bis 2-Jährigen die Altersangabe nur in Monaten erhoben wurde, ansonsten jedoch in Jahren.

Als Aufmerksamkeitsdefizit-/Hyperaktivitätsstörungsfälle (**ADHSFall = 1**) wurden Kinder eingestuft, wenn ADHS bei ihnen explizit von einem Arzt oder Psychologen (E027 = 1, E027m1 = 1, E027m2 = 1) diagnostiziert worden war; falls E027 = 2 (Nein) gilt auch **ADHSFall = 2**. **ADHSFall** erhält den Status „missing“, falls E027mx leer ist oder E027 = "Weiß nicht" oder E027 = ja und es ist nicht angekreuzt, wer die Diagnose gestellt hat; weiterhin bei E027 = ja und e027m3 = ja und E027m1 und E027m2 nicht angekreuzt. Als ADHS-Verdachtsfälle (**ADHSVer = 1**) wurden Teilnehmer/innen betrachtet, die Werte von größer gleich 7 auf der Unaufmerksamkeits-/Hyperaktivitätsskala des SDQ (Elternurteil: **hypauf_e** = 2) aufwiesen. Wurde das Kind als normal oder grenzwertig eingestuft, ist **ADHSVer = 2** (Nein) (LIT2: Schlack_ADHS.pdf).

Im Rahmen des KiGGS wurde die deutsche Übersetzung des **CSHCN**- (Children with Special Health Care Needs) Screener als Teil des Elternfragebogens eingesetzt. Die Definition eines speziellen Versorgungsbedarfes (CSHCN-positives Screening) stützt sich auf die Bejahung von mindestens einer der 5 Hauptfragen (vgl. E030a1 – E030e3) einschl. aller zugehörigen Unterfragen (LIT2: Scheidt-Nave_spez.Versorgungsbedarf.pdf).

Da der Fragenblock zum letzten Unfall des Kindes (E04201 – E04216) im Elternfragebogen z.T. detaillierter erhoben wurde als im Ausländerfragebogen, sind die entsprechenden zusätzlichen Items pro Hauptfrage jeweils in einer separaten Variablen gespeichert worden (E04202a, E04203a, E04204a, E04208a, E04209a, E04211a, E04212a, E04214a).

Zur Erfassung der gesundheitsbezogenen Lebensqualität von Kindern und Jugendlichen wurde bei KiGGS der KINDL-R-Fragebogen eingesetzt. Zum einen haben die Eltern für die 3- bis 17-Jährigen diesen Fragenkomplex beantwortet, zum anderen war der KINDL-R in den Kinderfragebögen integriert. Der KINDL-R ist ein Fragebogen mit 24 likert-skalierten Items, der folgende sechs Dimensionen der Lebensqualität im Rückblick auf die letzte Woche erfasst:

- „Körperliches Wohlbefinden“ (E0491 – E0494 bzw. K0491 – K0494) **kw100_e/k**,
- „Emotionales Wohlbefinden“ (E0501 – E0504 bzw. K0501 – K0504) **pw100e/k**,
- „Selbstwert“ (E0511 – E0514 bzw. K0511 – K0514) **sel100e/k**,
- „Wohlbefinden in der Familie“ (E0521 – E0524 bzw. K0521 – K0524) **fam100e/k**,
- „Wohlbefinden in Bezug auf Freunde/Gleichaltrige“ (E0531 – E0534 bzw. K0531 – K0534) **fre100e/k**,
- „Schulisches Wohlbefinden“ (E0541 – E0546 bzw. K0541 – K0546) **sch100e/k**.

Der KINDL-R Fragebogen liegt sowohl in einer Selbstbeurteilungsversion als auch in einer Fremdbeurteilungsversion (Elternurteil) vor. Die Befragten konnten die Items mittels fünfstufiger Antwort-kategorien (nie, selten, manchmal, oft, immer) beantworten. Außerdem wurde ein übergreifender Messwert der gesundheitsbezogenen Lebensqualität aus allen 24 Items berechnet werden (**tot100_e bzw. tot100_k**). Alle Messwerte werden auf Skalen von 0-100 Punkten angegeben, wobei höhere Werte eine bessere Lebensqualität bezeichnen (LIT2: Ravens-Sieberer_Lebensqualitaet.pdf).

Die Variablenbereiche E05501 – E05512 bzw. K05501 – K05512 eignen sich zur Erstellung von Skalenwerten der Dimension „Familiäre Schutzfaktoren“. Skalenwerte für „Personale Schutzfaktoren“ lassen sich bei den 11- bis 13-Jährigen auf Basis der Variablen K05513 – K05517 aufstellen; für die Jugendlichen stehen hierfür die Variablen K05516, K05517, K05524, K05525 und K05526 zur Verfügung. (LIT2: Erhart_psychischeEntw.pdf)

Im KiGGS wurden Hinweise auf psychische Auffälligkeiten und Stärken mit dem Strengths and Difficulties Questionnaire (SDQ) erfasst. Dieses Screening-Instrument kam sowohl im Elternfragebogen ab 3 Jahren (E056xx) als auch in den Fragebögen für die Kinder und Jugendlichen ab 11 Jahren (K056xx) zum Einsatz. Der SDQ erfragt psychische Schwächen und Stärken in den 5 Bereichen „Emotionale Probleme“ **emo_e/k**, „Hyperaktivitätsprobleme“ **hyp_e/k**, „Verhaltensprobleme“ **verh_e/k**, „Probleme mit Gleichaltrigen“ **peer_e/k** und „Prosoziales Verhalten“ **proso_e/k**, welches im Gegensatz zu den vier erstgenannten Bereichen einen Stärkenbereich betrifft. Für jeden der einzelnen Bereiche wurde aus den Itemantworten ein Testwert gebildet. Außerdem wurde aus der Summe der Problemskalen ein Gesamtproblemwert **totsdq_e/k** berechnet. Die Stärkenskala „Prosoziales Verhalten“ geht in den so genannten SDQ-Gesamtproblemwert nicht mit ein. Für den Gesamtproblemwert und die einzelnen Skalen wurden die Cut-off-Werte der englischen Normstichprobe zu Grunde gelegt. Die Kinder und Jugendlichen wurden danach als „unauffällig“, „grenzwertig“ oder „auffällig“ hinsichtlich des entsprechenden Problembereichs klassifiziert (**emoauf_e/k**, **hypauf_e/k**, **verhauf_e/k**, **peeauf_e/k**, **proauf_e/k**, **sdqauf_e/k**).

(LIT2: Hoelling_Verhaltensauffaelligkeiten.pdf).

In der neuen Variablen **STILLEN** wurde versucht alle Aspekte hierzu in 8 Kategorien abzubilden (Basis: E066, E067, E067z). Aktuell noch gestillte Kinder im 6. Lebensmonat oder älter werden in der Kategorie 6 mitgezählt. Zusätzlich wird noch **STILLV** („Kind jemals voll gestillt“) angegeben, wobei die „Weiß nicht“-Angaben als fehlende Werte eingestuft wurden.

Da das Rauchen bei den Ausländer-Kurzfragebögen nur als Ja/Nein-Variable erhoben wurde, in den anderen Fragebögen jedoch detaillierter, wurden die neuen Variablen **E070Ma**, **E070Va** und **K119a** gebildet. Die Informationen aus E120z1 und E120z2 stehen als **KZIG** (Anzahl Zigaretten/Tag) zur Verfügung. Das Alter bei Rauchbeginn und die Anzahl Zigaretten wurden nur bei Rauchern gefragt, die mindestens einmal pro Woche rauchen. Eine getrennte Ermittlung der Extraucher ist hier nicht möglich.

Die Neuvariable **E099n** fasst die Informationen aus E099 (noch nicht in Schule), E099a (nicht mehr zur Schule) und E099z (Schulklasse) zusammen. Für die 11- bis 13-Jährigen ist diese Variable entweder „missing“ oder auf Kategorie 2 (Klasse angegeben) gesetzt.

Die Frage zum allgemeinen Gesundheitszustand (K019) wurde im KFB11-13 erst ab November 2004 nachträglich eingefügt.

Falls beim Fragenkomplex „Inanspruchnahme“ eine Arzt- bzw. Therapeutengruppe (K085zxx, K086x bzw. analog bei E085zxx und E086x) nicht in Anspruch genommen wurde, ist die Anzahl der Besuche auf Null gesetzt worden. Die im EFB vorhandene, im KFB jedoch fehlende Abfrage zur Inanspruchnahme von Therapeuten wurde als **K086x** nachgebildet.

Beim Antwortenblock zum Zähneputzen (**K075mx**) wurden unplausible Mehrfachangaben auf die ungünstigste Antwort gesetzt. Als plausibel gelten alle Kombinationen der ersten 3 Items.

Bei der Abfrage zur körperlichen Aktivität im KFB (**K111**, **K111z**) wird der Schulsport nicht berücksichtigt. Da die Aktivität nur als ganze Stundenzahl eintragbar ist, kann eine Null auch Werte unterhalb 1 Stunde bedeuten.

Mit der Variablen **SCHULTYP** wird versucht die Angaben zum Schultyp (K106 bei Schülern) und zum Schulabschluss (K099c bei Schulabgängern) zusammenzufassen. Falls die Kinderangaben fehlen und die Eltern den Ausländer-Kurzfragebogen ausgefüllt haben, werden die Elternangaben berücksichtigt (E106). Die wenigen Angaben in beiden Ausländer-Kurzfragebögen zu „Berufsschule“ bzw. „Sonstiges“ wurden auf „missing“ gesetzt, zumal diese Kategorien sonst nirgends vorkamen.

Die neue Variable **SCOFF** charakterisiert eine mögliche Essstörung. Falls mindestens 2 Fragen zum Essverhalten (K1181 – K1185) mit „Ja“ beantwortet wurden, ist SCOFF = „Auffällig“ gesetzt worden, ansonsten auf „Nicht auffällig“. SCOFF ist „missing“, falls einzelne Fragen nicht beantwortet wurden. Aber: Fehlt nur eine Antwort und die vier übrigen Fragen sind mit „Nein“ beantwortet, dann steht der SCOFF auf „Nicht auffällig“ (LIT2: Hoelling_Esstörungen.pdf).

Bei einigen Krankheitsangaben im CAPI (CC01, CC02, CC03 und CC06) wurde zwar angegeben, dass hierfür in den letzten 12 Monaten Medikamente eingenommen wurden, jedoch wurde das Auftreten der Krankheit verneint. Eine Möglichkeit wäre, beide Informationen zusätzlich jeweils in einer Neuvariablen zusammenzufassen.

5.8. Anmerkungen zu Messbogenvariablen

Zu den anthropometrischen Messungen und den daraus abgeleiteten Größen sind erste Ergebnisauswertungen erfolgt (LIT2: Stolzenberg_Körpermaße.pdf, Kurth_Übergewicht.pdf). Ein Band „Referenzkurven“ mit geglätteten Perzentilkurven (LMS-Verfahren nach Cole) wird hierzu Anfang 2009 erscheinen. Im Folgenden werden die Definitionen der berechneten anthropometrischen Kennwerte kurz vorgestellt.

Der Body-Mass-Index (**BMI**), der quasi ein körperlängenkorrigiertes Relativgewicht darstellt, errechnet sich gemäß

$$\text{BMI [kg/m}^2\text{]} = \text{Körpergewicht [kg]} / \text{Körpergröße}^2 \text{ [m}^2\text{]}$$

Der Waist-to-Hip Ratio (**WHR**), der auch als Taille-Hüfte-Index bezeichnet wird, dient zur Einschätzung des Fettverteilungsmusters und ist definiert gemäß

$$\text{WHR} = \text{Taillenumfang [cm]} / \text{Hüftumfang [cm]}$$

Dem Datensatz beigelegt ist auch der Waist-to-Height Ratio (**WHtR**), der in den letzten Jahren immer mehr als effektiver Indikator für übergewichtsrelevante Gesundheitsrisiken erkannt wurde. Er ist einfach zu messen; der Grenzwert von 0.5 wurde als Cutpoint für beide Geschlechter, für Kinder ebenso wie für Erwachsene und für alle ethnischen Gruppen als gut geeignet vorgeschlagen (McCarthy HD, Ashwell M: A study of central fatness using waist-to-height ratios in UK children and adolescents over two decades supports the simple message - 'keep your waist circumference to less than half your height'. Int J Obes 2006; 30: 988-92).

Der Waist-to-Height Ratio wird berechnet als

$$\text{WHtR} = \text{Taillenumfang [cm]} / \text{Körpergröße [cm]}$$

WHR und **WHtR** sind auf 4 Nachkommastellen angegeben; in Anbetracht der erzielten Messgenauigkeiten von +/- 0,5 cm ist für die Angabe von Einzelwerten die Angabe von 2 Nachkommastellen gerechtfertigt.

Die Variable **BMKH** enthält die vorläufige Kategorisierung des **BMI** anhand von deutschen Referenzwerten (Kromeyer-Hauschild K, Wabitsch M et al., Perzentile für den Body-Mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. Monatsschr Kinderheilkd 2001, 149:807-818). Daraus abgeleitet ist **UeberKH** (Übergewicht: BMI >90. alters- und geschlechtsspezif. Perzentil) und **AdiposKH** (Adipositas: >97. alters- und geschlechts-spezif. Perzentil). Alternative Definitionen hierzu (**UeberIO** bzw. **AdiposIO**) beziehen sich auf die IOTF-Referenzpopulation aus verschiedenen Ländern (Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. British Medical Journal. 2000, 320:1240-3). Somit beginnt hier Übergewicht bei dem alters- und geschlechtsspezif. Perzentil, das bei 18 Jahren einem BMI von 25 kg/m² entspricht. Entsprechend gilt das alters- und geschlechtsspezifische Perzentil, das bei 18 Jahren einem BMI von 30 kg/m² entspricht, als Indikator für Adipositas. Beide Ansätze sind im KiGGS-Datensatz nur für Kinder ab 3 Jahren definiert.

Die neue Variable **MENAR**, die sich von PERI im Messbogen ableitet, kennzeichnet das Eintreten der Menarche bei Mädchen.

5.9. Der Sozialschicht-Index nach Winkler (2003/4)

Der Sozialschicht-Index nach Winkler wird durch die Variable **WINDEXZ** abgebildet, deren Skalenwerte metrisch interpretiert werden sollten. Daraus abgeleitet ist die Sozialschicht

(SCHICHTZ) mit ihren 3 Kategorien. Wer hier die (sprachliche) Einteilung in Ober-/Mittel- und Unterschicht vermeiden will, dem sei der Begriff „Sozialstatus“ mit den 3 Attributen „hoch“, „mittel“ und „niedrig“ empfohlen (LIT2: Lange_soziol.Merkmale.pdf).

Der soziale Status wird gemessen anhand von Angaben der Eltern zu ihrer Bildung inkl. Ausbildung, ihrer beruflichen Stellung sowie zum Nettoeinkommen aller Haushaltsmitglieder; die genauen Operationalisierungsschritte sind weiter unten beschrieben.

Da die Methodik des Winkler-Index von 1998 (basierend auf Daten des BGS98) nachweislich nicht mehr auf die KiGGS-Studie von 2003-2006 angewendet werden konnte (Inflation, Veränderungen bei der Bildung), musste eine Neuadjustierung des Index erfolgen. Als Basis dienten hier die beiden im RKI durchgeführten Telefon-Surveys GSTel03 und GSTel04, von denen repräsentative Daten für Deutschland 2003 bzw. 2004 vorlagen. Der gebildete neue Index, der den Teilscore zur beruflichen Stellung in Abhängigkeit der Beantwortung der Frage zum Hauptverdiener ermittelt, musste noch modifiziert werden, da bei KiGGS diese Frage fehlt dafür aber Angaben von beiden Eltern zu den Sozialvariablen vorlagen. Es war schließlich das Ziel, mit Hilfe des Sozialindex den Haushalt zu charakterisieren, in welchem das Kind hauptsächlich lebt.

Die Schicht-Variable ist nützlich für tabellarische Grundausswertungen oder bei statistischen Analysen (Fallzahlproblem usw.). Ist bekannt bzw. wurde ermittelt, dass nur einer der 3 Faktoren für den betrachteten Zusammenhang von Bedeutung ist, so sollte besser der jeweilige Teilscore (educz oder prof oder moneyz) zur Modellierung benutzt werden.

a) Operationalisierung von Bildung/Ausbildung für Mutter/Vater:

Variablen: E089M/V, E090M/V, (E091M/V = 6)

	Bildung		Ausbildung
educzm/v = 1	Noch kein Schulabschluss	oder	
	Haupt-/Volksschule Realschule/mittl. Reife POS/10.Klasse Fachhochschulreife/Fachoberschule Anderer Schulabschluss Schule beendet ohne Abschluss	und	Anderer Berufsabschluss Keinen Berufsabschluss (Auszubildender)
educzm/v = 2	Haupt-/Volksschule Anderer Schulabschluss Schule beendet ohne Abschluss	und	Lehre, berufl.-betriebl. Ausbildung BFS, Handelsschule, berufl.-schul. Ausbildung Fachschule Noch in berufl. Ausbildung
educzm/v = 3	Realschule/mittl. Reife	und	Lehre, berufl.-betriebl. Ausbildung BFS, Handelsschule, berufl.-schul. Ausbildung Fachschule Noch in berufl. Ausbildung
educzm/v = 4	POS/10.Klasse Fachhochschulreife/Fachoberschule	und	Lehre, berufl.-betriebl. Ausbildung BFS, Handelsschule, berufl.-schul. Ausbildung Fachschule Noch in berufl. Ausbildung
educzm/v = 5	Abitur, EOS, allgem. fachgeb. Hochschulreife	und	Lehre, berufl.-betriebl. Ausbildung BFS, Handelsschule, berufl.-schul. Ausbildung Fachschule Anderer Berufsabschluss Keinen Berufsabschluss Noch in berufl. Ausbildung
educzm/v = 6			Fachhochschule, Ingenieursschule
educzm/v = 7			Universität, Hochschule

b) Operationalisierung von Einkommen:

Variable: E093

moneyz = 1	< 1250 €
moneyz = 2	1250 - < 1750 €
moneyz = 3	1750 - < 2250 €
moneyz = 4	2250 - < 3000 €
moneyz = 5	3000 - < 4000 €
moneyz = 6	4000 - < 5000 €
moneyz = 7	>= 5000 €

c) Operationalisierung von berufl. Stellung von Mutter/Vater:

Variablen: E092M/V

profm/v = 1	Schüler, Student oder Auszubildender (zB. In Lehre) Ungelernter Arbeiter Hausfrau/mann
profm/v = 2	Angelernter oder gelernter Arbeiter oder Facharbeiter Selbst. Landwirt/Genossenschaftsbauer
profm/v = 3	Vorarbeiter, Meister, Polier usw. Angestellter mit einf. Tätigkeit Beamter einfacher Dienst Mithelfender Familienangehöriger
profm/v = 4	Angestellter Industrie-/Werkmeister oder Angestellter mit qualifizierter Tätigkeit Beamter mittlerer Dienst
profm/v = 5	Sonstiger Selbständiger mit bis zu 9 Mitarbeitern/Partnern
profm/v = 6	Angestellter mit hochqualifizierter Tätigkeit oder Leitungsfunktion Beamter gehobener Dienst Freiberuflich, selbständiger Akademiker
profm/v = 7	Angestellter mit umfassender Führungstätigkeit u. Entscheidungsbefugnissen Beamter höherer Dienst Sonstiger Selbständiger mit 10 und mehr Mitarbeitern/Partnern

d) Ermittlung des Indexscores und der Schichtkategorien (Winkler D2003/4):

Die 1-7 Punkte umfassenden Teilscores werden jeweils addiert gemäß

$$\text{windexzm} = \text{educzm} + \text{moneyz} + \text{profm}$$

bzw.
$$\text{windexzv} = \text{educzv} + \text{moneyz} + \text{profv}$$

Zur Reduktion möglicher Missing Values wird aus zwei vorhandenen Teilscores der dritte geschätzt, z.B.

$$\text{IF (MISSING(moneyz)) windexzm} = \text{RND}((\text{educzm} + \text{profm})/2 * 3)$$

Der Winkler-Index kann die Werte 3..21 annehmen. Der Wertebereich 3..8 wird als Unterschicht bezeichnet; die Mittelschicht ist über den Wertebereich 9..14 definiert. Davon grenzt sich die Oberschicht ab, die Werte von 15..21 Punkte annimmt. Somit sind Indexwert und Schichtzugehörigkeit für Mutter/Vater bestimmt.

Erreicht die Mutter einen Indexwert der größer bzw. gleich dem des Vaters ist oder der Indexwert für den Vater kann nicht geschätzt werden, so ist der Indexwert = windexzm. Ist der Indexwert für die Mutter kleiner als der des Vaters oder der Indexwert für die Mutter ist nicht zu schätzen, so ist der Indexwert = windexzv. Der Index bzw. die Schichtzugehörigkeit des untersuchten Kindes wird indirekt über die jeweiligen Schätzwerte für die Eltern approximiert. Lebt jedoch das Kind überwiegend nur bei der Mutter bzw. überwiegend nur beim Vater, dann wird unabhängig von der vorhergehenden Regel der Indexwert auf windexzm bzw. windexzv gesetzt.

e) SPSS-Syntax

* Schichtz und Windexz für KiGGS.

* Neuer Winkler-Index (D2003/4) umgesetzt für Mutter.

If (e089m = 8) educzm = 1.

If ((range(e089m,1,4) or range(e089m,6,7)) and (range(e090m,6,7) or e091m = 6)) educzm = 1.

If ((e089m = 1 or range(e089m,6,7)) and (range(e090m,1,3) or e090m = 8)) educzm = 2.

If (e089m = 2 and (range(e090m,1,3) or e090m = 8)) educzm = 3.

If (range(e089m,3,4) and (range(e090m,1,3) or e090m = 8)) educzm = 4.

If (e089m = 5 and (range(e090m,1,3) or range(e090m,6,8))) educzm = 5.

If (e090m = 4) educzm = 6.

If (e090m = 5) educzm = 7.

If (e093 lt 5) moneyz = 1.

If (e093 = 5 or e093 = 6) moneyz = 2.

If (e093 = 7 or e093 = 8) moneyz = 3.

If (e093 = 9 or e093 = 10) moneyz = 4.

If (e093 = 11) moneyz = 5.

If (e093 = 12) moneyz = 6.

If (e093 = 13) moneyz = 7.

IF (e092m = 1 or e092m = 19 or e092m = 20) profm = 1 .

IF (e092m = 2 or e092m = 3 or e092m = 5) profm = 2 .

IF (e092m = 4 or e092m = 11 or e092m = 15 or e092m = 9) profm = 3 .

IF (e092m = 10 or e092m = 12 or e092m = 16) profm = 4 .

IF (e092m = 7) profm = 5 .

IF (e092m = 13 or e092m = 17 or e092m = 6) profm = 6 .

IF (e092m = 14 or e092m = 18 or e092m = 8) profm = 7 .

COMPUTE windexzm = educzm + profm + moneyz .

IF (MISSING(educzm)) windexzm = RND((moneyz + profm)/ 2 * 3) .

IF (MISSING(moneyz)) windexzm = RND((educzm + profm)/ 2 * 3) .

IF (MISSING(profm)) windexzm = RND((educzm + moneyz)/ 2 * 3) .

IF (windexzm < 9) schichtm = 1 .

IF (windexzm > 8 & windexzm < 15) schichtm = 2 .

IF (windexzm > 14) schichtm = 3 .

FORMATS educzm, profm, moneyz, windexzm, schichtm (F2).

VARIABLE LABELS educzm "Winklerscore: Schul- u. Berufsausbildung(Mutter)".

VARIABLE LABELS profm "Winklerscore: Berufl. Stellung(Mutter)".

VARIABLE LABELS moneyz "Winklerscore: Monatl. Haushalts-Nettoeinkommen".

VARIABLE LABELS windexzm "Mutter: Winkler-Indexscore (D2003)".

VARIABLE LABELS schichtm "Mutter: Soziale Schicht n. Winkler (D2003)".

VALUE LABELS moneyz 1 '<1250 €' 2 '1250 - <1750 €' 3 '1750 - <2250 €'

4 '2250 - <3000 €' 5 '3000 - <4000 €' 6 '4000 - <5000 €'

7 '>=5000 €'.

VALUE LABELS schichtm 1 'Unterschicht' 2 'Mittelschicht' 3 'Oberschicht'.

VALUE LABELS educzm 1 "niedrig" 7 "hoch".

VALUE LABELS profm 1 "niedrig" 7 "hoch".

EXECUTE .

* Neuer Winkler-Index (D2003/4) umgesetzt für Vater.

If (e089v = 8) educzv = 1.

If ((range(e089v,1,4) or range(e089v,6,7)) and (range(e090v,6,7) or e091v = 6)) educzv = 1.

If ((e089v = 1 or range(e089v,6,7)) and (range(e090v,1,3) or e090v = 8)) educzv = 2.

If (e089v = 2 and (range(e090v,1,3) or e090v = 8)) educzv = 3.

If (range(e089v,3,4) and (range(e090v,1,3) or e090v = 8)) educzv = 4.

If (e089v = 5 and (range(e090v,1,3) or range(e090v,6,8))) educzv = 5.

If (e090v = 4) educzv = 6.

If (e090v = 5) educzv = 7.

If (e093 lt 5) moneyz = 1.

If (e093 = 5 or e093 = 6) moneyz = 2.


```

If (e093 = 7 or e093 = 8) moneyz = 3.
If (e093 = 9 or e093 = 10) moneyz = 4.
If (e093 = 11) moneyz = 5.
If (e093 = 12) moneyz = 6.
If (e093 = 13) moneyz = 7.
IF (e092v = 1 or e092v = 19 or e092v = 20) profv = 1 .
IF (e092v = 2 or e092v = 3 or e092v = 5) profv = 2 .
IF (e092v = 4 or e092v = 11 or e092v = 15 or e092v = 9) profv = 3 .
IF (e092v = 10 or e092v = 12 or e092v = 16) profv = 4 .
IF (e092v = 7) profv = 5 .
IF (e092v = 13 or e092v = 17 or e092v = 6) profv = 6 .
IF (e092v = 14 or e092v = 18 or e092v = 8) profv = 7 .
COMPUTE windexzv = educzv + profv + moneyz .
IF (MISSING(educzv)) windexzv = RND((moneyz + profv)/ 2 * 3) .
IF (MISSING(moneyz)) windexzv = RND((educzv + profv)/ 2 * 3) .
IF (MISSING(profv)) windexzv = RND((educzv + moneyz)/ 2 * 3) .
IF (windexzv < 9) schichtv = 1 .
IF (windexzv > 8 & windexzv < 15) schichtv = 2 .
IF (windexzv > 14) schichtv = 3 .
FORMATS educzv, profv, moneyz, windexzv, schichtv (F2).
VARIABLE LABELS educzv "Winklerscore: Schul- u. Berufsbildung(Vater)".
VARIABLE LABELS profv "Winklerscore: Berufl. Stellung(Vater)".
VARIABLE LABELS moneyz "Winklerscore: Monatl. Haushalts-Nettoeinkommen".
VARIABLE LABELS windexzv "Vater: Winkler-Indexscore (D2003/4)".
VARIABLE LABELS schichtv "Vater: Soziale Schicht n. Winkler (D2003/4)".
VALUE LABELS moneyz 1 '<1250 €' 2 '1250 - <1750 €' 3 '1750 - <2250 €'
4 '2250 - <3000 €' 5 '3000 - <4000 €'
6 '4000 - <5000 €' 7 '>=5000 €'.
VALUE LABELS schichtv 1 'Unterschicht' 2 'Mittelschicht' 3 'Oberschicht'.
VALUE LABELS educzv 1 "niedrig" 7 "hoch".
VALUE LABELS profv 1 "niedrig" 7 "hoch".
EXECUTE .

```

* Ermittlung des Index-Scores und Schichtbildung.

```

IF ((windexzm ge windexzv) or missing(windexzv)) windexz = windexzm.
IF ((windexzm lt windexzv) or missing(windexzm)) windexz = windexzv.
IF (e001 = 4) windexz = windexzm.
IF (e001 = 5) windexz = windexzv.
FORMATS windexz (F2).
VARIABLE LABELS windexz "Winkler-Indexscore (D2003/4)".
EXECUTE .
IF ((schichtm ge schichtv) or missing(schichtv)) schichtz = schichtm.
IF ((schichtm lt schichtv) or missing(schichtm)) schichtz = schichtv.
IF (e001 = 4) schichtz = schichtm.
IF (e001 = 5) schichtz = schichtv.
FORMATS schichtz (F2).
Variable Level schichtz (Ordinal).
VARIABLE LABELS schichtz "Soziale Schicht n. Winkler (D2003/4)".
VALUE LABELS schichtz 1 'Unterschicht' 2 'Mittelschicht' 3 'Oberschicht'.
EXECUTE .

```

5.10. Impfvariablen

Für die Erstellung der Impfvariablen wurden die zur Untersuchung mitgebrachten Impfdokumente (Impfausweise) der Kinder und Jugendlichen kopiert und anschließend der Präparatename und das Impfdatum jeder Impfung in eine Datenbank eingegeben. War der Name des Impfpräparats nicht dokumentiert, jedoch die Erkrankung/en, gegen die geimpft worden war, wurden diese Erkrankung/en eingegeben; zusätzlich wurde - falls vorhanden - vermerkt, ob die Impfung mit einem Mono- oder Kombinationspräparat verabreicht wurde. Auf der Grundlage dieser Information wurden die Indikatoren Grundimmunisierung mit bzw. ohne Boosterdosis generiert. Dabei wurde für Tetanus, Diphtherie, Poliomyelitis, Haemophilus influenzae Typ B (Hib) und Hepatitis B berücksichtigt, ob die Impfung mit einem Kombinationspräparat mit Pertussiskomponente durchgeführt wurde.

Es galten die folgenden Kriterien für die Definition der vollständigen Grundimmunisierung (mit Boosterdosis):

- Vier Impfdosen bei Impfstoff mit Pertussiskomponente
- Drei Impfdosen bei Impfstoff ohne Pertussiskomponente
- Ausnahmen/Zusatzregeln
 - Hib: eine Dosis im Alter > 14 Monate
 - Polio: zwei Dosen IPV-VirelonR oder drei (DDR: vier) Dosen orale Polio-Vakzine (OPV)

Für die erste Auffrischungsimpfung Tetanus und Diphtherie (empfohlen für 6.-7. Lebensjahr) zählte

- jede weitere Dosis nach vollständiger Grundimmunisierung.

Der Impfstatus wurde ermittelt unter Beachtung der von der STIKO für das jeweilige Alter des Kindes geforderten Mindestimpfungen. Die Gruppe der Teilnehmer, für die eine Auswertung der Impfvariablen möglich ist, ist beschränkt auf diejenigen, die einen vollständigen Impfausweis vorgelegt haben oder die nach Angaben der Eltern noch ungeimpft sind (**Impfpop**). Der spezielle Impffragebogen ist im Ordner FB gespeichert.

5.11. Neuvariablen zur Arzneimittelanwendung

Die Erfassung des Arzneimittelgebrauchs erfolgte in einem standardisierten Laptop-gestützten persönlichen Interview durch die untersuchende Ärztin bzw. den untersuchenden Arzt. Interviewt wurden die Eltern aller Studienteilnehmer/innen. Jugendliche ab 14 Jahre konnten die Angaben zum Medikamentengebrauch auch selbst machen.

Mit der Frage „*Hat Ihr Kind innerhalb der letzten 7 Tage Medikamente angewendet? Bitte denken Sie auch an Salben, Einreibungen, Empfängnisverhütung (z. B. Pille), Vitamine (z. B. Vit C, Vit D, Vit E, Multivit.) und Mineralstoffe (z. B. Kalzium, Magnesium, Silicea, Selen), med. Tees, pflanzliche Arzneimittel und Homöopathika!*“ wurde der gesamte Arzneimittelgebrauch in den letzten 7 Tagen vor der Untersuchung erhoben. Es wurden nicht nur die vom Arzt oder vom Heilpraktiker verordneten Arzneimittel erfasst, sondern auch auf Präparate der Selbstmedikation, worin sowohl die freiverkäuflichen OTC-Produkte (Over-The-Counter-Produkte) als auch auf anderem Wege (z. B. Hausapotheke) beschaffte Medikamente eingingen.

Die Freitextangaben zum Arzneimittelnamen wurden unter Berücksichtigung der Indikation nach dem ATC-Code der WHO verschlüsselt (LIT2: Knopf_Arzneimittelanwendung.pdf). Als Referenz für die Kodierung wurde die Gelbe Liste (4/2005) genutzt. Konnte die Arzneimittelnennung hier nicht identifiziert werden, wurde das Arzneimittelinformationssystem (AMIS) bei DIMDI herangezogen. Ließen sich auch über AMIS keine Informationen zum

Arzneimittel ermitteln, wurde im Internet mit Hilfe von Google recherchiert und der ATC-Kode unter Berücksichtigung von Arzneimittelindikation und ermittelten Wirkstoffen vergeben. Für Homöopathika, die weder in der Gelben Liste noch in AMIS mit einer ATC-Kodierung versehen waren, wurde der ATC-Kode Z „Homöopathika, nicht anders zuordenbar“ zugeordnet.

Im Public Use File liegen Angaben zur Arzneimittelanwendung nach allen ATC-Klassen (A00 bis Z00) und nach ausgewählten ATC-Gruppen (A01 bis V07) vor.

5.12. Variablen des Ernährungsfragebogens

Mit Hilfe eines standardisierten Ernährungsfragebogens wurden die Verzehrshäufigkeiten und Portionsmengen verschiedener Lebensmittelgruppen „in den letzten Wochen“ ermittelt. Es wurden hierbei zwei inhaltlich identische Varianten dieses Erhebungsinstruments eingesetzt: Einer für die Eltern der 1- bis 10-Jährigen sowie einer für die Kinder und Jugendlichen im Alter von 11-17 Jahren (LIT2: Mensink_Lebensmittelverzehr.pdf).

Berechnung von Verzehrsmengen:

Um aus den vorliegenden Angaben mittlere Tagesmengen zu berechnen, müssen die Portionsmengen mit der Verzehrshäufigkeit kombiniert werden. Dazu wird empfohlen, die Häufigkeiten in Tage bzw. Gelegenheiten pro 4 Wochen (28 Tage) nach dem folgenden Schema umzurechnen:

Tabelle 1: Umrechnung Verzehrshäufigkeiten

Kategorie 1	(„nie“)	Wert : 0
Kategorie 2	(1mal im Monat)	Wert: 1
Kategorie 3	(2-3mal im Monat)	Wert: 2,5
Kategorie 4	(1-2mal pro Woche)	Wert: 6
Kategorie 5	(3-4mal pro Woche)	Wert : 14
Kategorie 6	(5-6mal pro Woche)	Wert: 22
Kategorie 7	(1mal am Tag)	Wert 28
Kategorie 8	(2-3mal am Tag)	Wert: 70
Kategorie 9	(4-5mal am Tag)	Wert: 126
Kategorie 10	(öfter als 5mal am Tag)	Wert: 168

Es wird empfohlen, für eine Umrechnung der Portionskategorien in verzehrte Mengen die in der Tabelle 2 dargestellten Mengen für die jeweils mittlere Portionskategorie (im Datensatz mit 3 kodiert) zu verwenden. Anhand der Wertelabels der Portionen bzw. direkt aus dem Fragebogen lassen sich hieraus die Mengen für die anderen Portionskategorien ableiten. (Diese sind nicht bei jeder Frage einheitlich!)

Für die Portionsmengen von Fleisch würde dies z.B. bedeuten, 60 Gramm für die Kategorie 3 (mittlere Kategorie = ½ Portion) und für Kategorie 1 (1/8 Portion oder weniger): 15 g, Kategorie 2 (1/4 Portion): 30 g, Kategorie 4 (1 Portion): 120 g, und Kategorie 5 (2 Portionen oder mehr): 240 g. Bei Kategorien wie „¼ Portion (oder weniger)“ bzw. „2 Portionen (oder mehr)“ wird immer mit der angegebenen Portionsmenge gerechnet, d.h. die Angabe „oder weniger“ bzw. „oder mehr“ bleibt unberücksichtigt.

Die mittlere Tagesmenge errechnet sich anhand folgender Formel:

$$\text{Mittlere Tagesmenge} = \text{Portionsmenge} * \text{Häufigkeit} / 28$$

Tabelle 2: Portionsmengen in Gramm für die Berechnung von Verzehrsmengen

FQ - Frage	Lebensmittelgruppe	Mittlere Angabe (Kategorie 3)	Grammmenge für mittlere Angabe
1a	Milch	1 Glas (200 ml)	200
2a	Erfrischungsgetränke	1 Glas (200 ml)	200
3a	Sportler-/Energiegetränke	1 Glas (200 ml)	200
4a	Fruchtsaft, -nektar, Gemüsesaft	1 Glas (200 ml)	200
5a	Leitungswasser	1 Glas (200 ml)	200
6a	Mineralwasser	1 Glas (200 ml)	200
7a	Früchte-, Kräutertee	1 Tasse (150 ml)	150
8a	Schwarzer, grüner Tee	1 Tasse (150 ml)	150
9a	Kaffee	1 Tasse (150 ml)	150
10a	Müsli, Cornflakes, Snacks	1 Schale	50
11a	Vollkornbrot, -brötchen, Schwarzbrot	1 Scheibe/ Brötchen	50
12a	Weißbrot, Graubrot, Brötchen	1 Scheibe/ Brötchen	50
13a	Butter	2 Teelöffel	10
14a	Margarine	2 Teelöffel	10
15a	Käse	1 Scheibe/ Portion	30
16a	Quark, Joghurt, Dickmilch	1 Becher	200
17a	Frischkäse	2 Esslöffel	30
18a	Eier	1 Ei	60
19a	Suppe	1 Teller	400
20a	Fleisch	½ Portion	60
21a	Geflügel	½ Portion	75
22a	Wurst, Schinken	1 Scheibe	20
23a	Fisch	1 Portion	90
24a	Frisches Obst	1 Stück/ Schale	150
25a	Gekochtes Obst, Konservenobst	1 Schale	150
26a	Gekochtes, frisches Gemüse	1 Portion	150
27a	Tiefkühlgemüse	1 Portion	150
28a	Konservengemüse	1 Portion	150
29a	Blattsalat, Rohkost, rohes Gemüse	1 Portion	100
29b	Salatdressing		1 Esslöffel = 12 g
30a	Nudeln, Reis	1 Teller	125
31a	Gekochte Kartoffeln	1 Portion/2 Kartoffeln	175
32a	Frittierte, gebratene Kartoffeln	1 Portion	150
33a	Bratwurst, Currywurst, Hamburger, Döner Kebab	1 Portion Fleisch aus 1 Portion	195 85
34a	Ketchup, Majonäse	2 Esslöffel	20
35a	Kuchen, Gebäck	1 Stück	100
36a	Kekse	3 Kekse	15
37a	Schokolade, Schokoriegel	½ Tafel/ 1 Riegel	50
38a	Süßigkeiten (z.B. Bonbons)	6-10 Stück	16
39a	Eis	2 Kugeln	150
40a	Pudding, Milchreis	1 Schale	150
41a	Pfannkuchen	1 Stück	100
42a	Honig, Marmelade	2 Teelöffel	20
43a	Nuss-Nougatcreme	2 Teelöffel	20
42a	Knabberartikel	1 Schale	50
45a	Nüsse	1 Portion	25

5.13. Die Laborvariablen

(LIT2: Thierfelder_biochem.Messparameter.pdf)

Tabelle 3: Labor - Untersuchungsmethodik

Analyt (Einheit)	Methode	Gerät	VK %	K
Vitamin B12 [ng/l]	Elektro-Chemilumineszenz-Bindungsassay	Elecsys E 2010	9,6	4
Eisen [µmol/l]	Ferrozin- Methode ohne Enteiweißung (Roche, Mannheim)	Hitachi 917	1,7	2
s-TfR [mg/l]	latexverstärkter Agglutinationstest (Behring, Marburg)	BNA- Nephelometer (Behring)	6,8	5
Ferritin [µg/l]	Elektro-Chemilumineszenz-Bindungsassay	Elecsys E 2010	8,3	5
Erythrozytenzahl (RBC) [Mio/ml]	Impedanzprinzip	Cell-Dyn 3500 (Abbott, Wiesbaden),	1,8	1
Hämoglobin im Blut (Hb) [g/dl]	Fotometrie	Cell-Dyn 3500	1,4	1
Hämatokrit (PCV) [%]	Rechenparameter	Cell-Dyn 3500	1,8	2
Mittlerer Hämoglobingehalt des Erythrozyten (MCH) [pg]	Rechenparameter	Cell-Dyn 3500	1,4	2
Mittlere Hämoglobinkonzentration in den Erythrozyten (MCHC) [g/dl]	Rechenparameter	Cell-Dyn 3500	1,4	2
Erythrozytenvolumen (MCV) [fl]	Volumenmessung	Cell-Dyn 3500	1,4	2
Cholesterin-gesamt [mg/dl]	Vollezymatische Cholesterin-Oxidase-PAP Methode (Roche, Mannheim)	Hitachi 917	2,1	2
HDL- Cholesterin [mg/dl]	Homogener enzymatischer Farbstest (Roche, Mannheim)	Hitachi 917	2,9	2
LDL- Cholesterin [mg/dl]	Homogener enzymatischer Farbstest (Roche, Mannheim)	Hitachi 917	2,4	2
Triglyzeride [mg/dl]	Gesamt- Glycerin mit der GPO-PAP- Reaktion (Roche, Mannheim)	Hitachi 917	2,0	3
Homocystein [µmol/l]	Fluoreszenzpartikel-Immunoassay (Abbott)	AxSYM (Abbott)	5,2	4
Harnsäure [mg/dl]	Uricase-PAP- Reaktion (Roche, Mannheim)	Hitachi 917	1,4	2
Glukose [mg/dl]	Hexokinase- Methode (Roche, Mannheim)	Hitachi 917	1,7	2
HbA1c [%]	high performance liquid chromatography (HPLC)	Diastat (Bio-Rad, München)	3,5	5
25(OH)Vitamin D [nmol/l]	Lumineszimmunoassay (LIA) (DiaSorin, Diezenbach)	Liaison (DiaSorin, Diezenbach)	11,7	4

Analyt (Einheit)	Methode	Gerät	VK %	K
Magnesium [mmol/l]	Xylidylblaureaktion, photometr. Farbtest (Roche, Mannheim)	Hitachi 917	2,2	2
Kalzium [mmol/l]	ortho-Kresolphthalein-Komplexon Methode (Roche, Mannheim)	Hitachi 917	2,3	2
Anorganisches Phosphat [mmol/l]	Ammonium-Phosphomolybdat- Methode (Roche, Mannheim)	Hitachi 917	1,7	2
Alkalische Phosphatase [U/l]	Optimierte Standard-Methode der IFCC (Roche, Mannheim)	Hitachi 917	2,8	2
Gamma-Glutamyltransferase (GGT) [U/l]	optimierte IFCC- Methode (Roche, Mannheim)	Hitachi 917	2,2	2
Kalium [mmol/l]	Potentiometrie mit ionenselektiver Elektrode (Roche, Mannheim)	Hitachi 917	1,2	2
Gesamtprotein [g/dl]	Biuret- Reaktion (Roche, Mannheim)	Hitachi 917	2,0	2
Kreatinin im Urin (Spontanurin) [mg/dl]	Jaffé-Methode ohne Enteiweißung (Roche, Mannheim)	Hitachi 917	2,7	5
Spezifisches IgE ^{*)} [kU/l]	Fluorezenz-Immunoassay (CAP, Phadia)	UNICAP 1000	7,8	1

^{*)} Allergene: Sensibilisierung gemäß CAP-Klasse 0 und ≥ 1 .

VK: interserieller Variationskoeffizient;

K: Kategorie der Bewertungsgrundlage zur Interpretation von Messergebnissen (siehe dazu LIT2: Thierfelder_biochem.Messparameter.pdf

6. Auswertung von KiGGS-Daten

In diesem Abschnitt werden einige wichtige Grundsätze zur Auswertung von KiGGS-Daten beschrieben.

6.1. Altersspezifische Erhebungsinstrumente

Bei allen Auswertungen muss darauf geachtet werden, dass die Erhebungsinstrumente, insbesondere die Fragebögen, den Altersgruppen entsprechend variieren können. Die dadurch entstehenden Missings sind durch –7 bzw. .N gekennzeichnet (s.o.). Ausländer, die den Ausländerfragebogen oder nur den Ausländerkurzfragebogen ausgefüllt haben, haben darüber hinaus bei einzelnen Variablen systematisch fehlende Werte.

6.2. Stichprobengewichtung

6.2.1. Beschreibung der Gewichtungsvariablen

Die KiGGS-Teilnehmer wurden über ein komplexes zweistufiges Stichprobenverfahren mit ungleichen Auswahlwahrscheinlichkeiten gezogen. Auf der ersten Stufe wurden 167 Sample Points (Untersuchungsorte) gezogen, wobei Gemeinden im Osten Deutschlands überproportional vertreten waren, um die Power für Ost-West-Vergleiche zu erhöhen. Vom diesem Oversampling Ost abgesehen, bilden die Sample Points nach Gemeindetypen (BIK-Klassifikation) und räumlicher Verteilung die Struktur der Bundesrepublik ab. Die Ziehung der Gemeinden geschah proportional zur Zahl der Kinder und Jugendlichen von 0–17 Jahren. Auf der zweiten Stufe wurde innerhalb der Sample Points eine feste Zahl von Kindern und Jugendlichen pro Jahrgang gezogen.

Die Stichprobengewichte, die in der Variable **wKiGGS** enthalten sind, gleichen die durch die Art der Stichprobenziehung und das Studiendesign bedingte unterschiedlich hohe Auswahlwahrscheinlichkeit der Probanden aus. Darüber hinaus gleichen sie die Verteilung von Alter, Geschlecht und Staatsangehörigkeit (deutsch vs. nicht-deutsch) in der KiGGS-Stichprobe an die entsprechenden Bevölkerungszahlen in Deutschland mit Stand vom 31.12.2004 an. Schließlich leisten sie den Ausgleich des Oversamplings Ost.

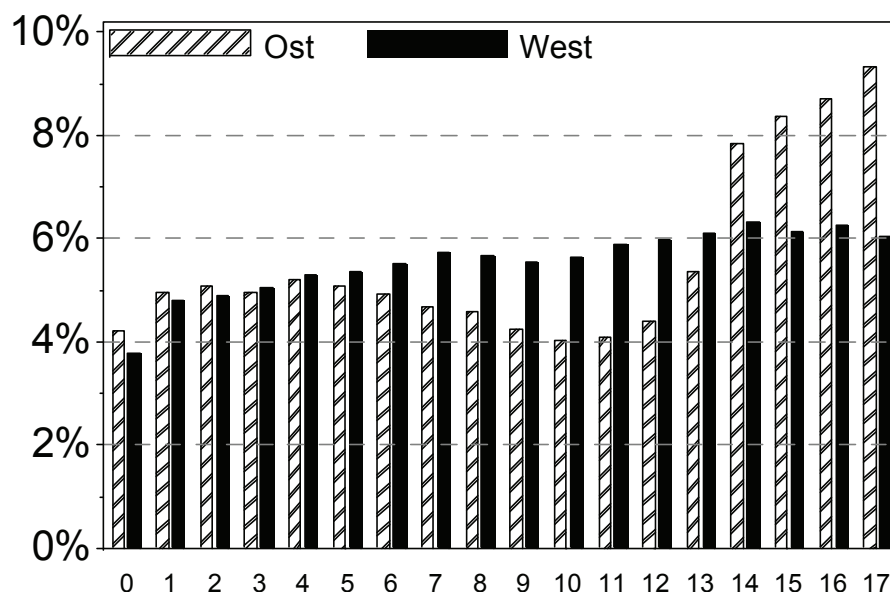
Non-Response wird durch die Stichprobengewichte nur bedingt ausgeglichen. Lediglich die Non-Response, die sich durch die Altersgruppe (0-2, 3-6, 7-10, 11-13 und 14-17 Jahre) und das Geschlecht innerhalb der Sample Points erklären lässt, oder durch Altersjahrgang, Geschlecht und Staatsangehörigkeit (deutsch vs. nicht-deutsch) in der gesamten Studienpopulation (über alle Sample Points hinweg) und/oder durch Unterschiede zwischen den drei Regionen Westdeutschland, Berlin und Ostdeutschland erklären lässt, wird durch die Gewichtung ausgeglichen. Non-Response, die auf anderen Faktoren beruht, wird durch die Gewichtung nicht ausgeglichen.

Den größten Einfluss auf die Gewichte haben das Oversampling Ost und die Altersverteilung der Kinder und Jugendlichen in der Gesamtbevölkerung, die von einer Gleichverteilung abweicht und sich zudem zwischen West- und Ostdeutschland unterscheidet. In ganz Deutschland haben die Geburtenzahlen in den jüngeren, später geborenen Jahrgängen abgenommen. In Ost-Deutschland ist zusätzlich der Geburtenknick nach der politischen Wende deutlich erkennbar.

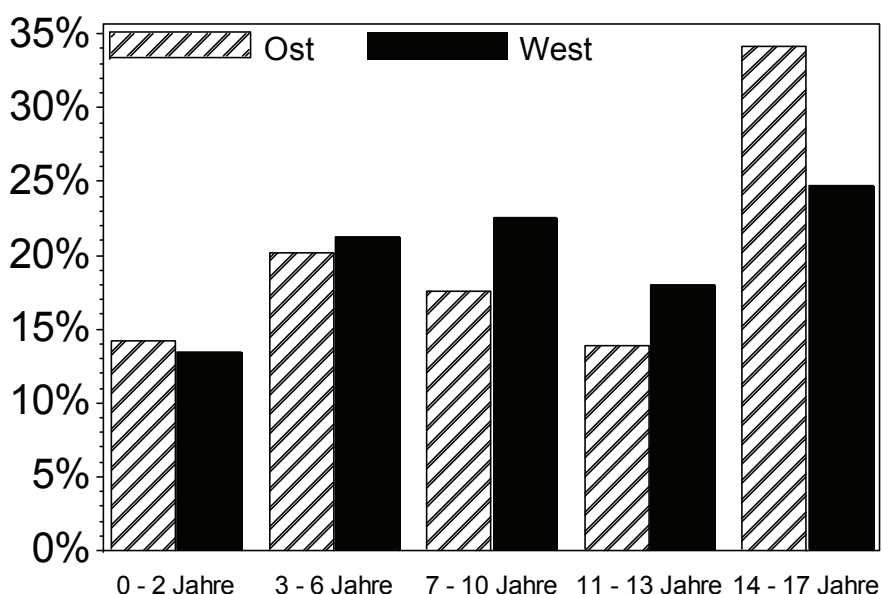
Den deutlichen Unterschied im Altersaufbau der jugendlichen Bevölkerung zwischen Ost und West zeigt auch die folgende Abbildung.

Abbildung 1: Altersaufbau in Ost und West zum Vergleich
(Die Anteile summieren sich jeweils in Ost und West zu 100% auf.)

a) nach Altersjahrgängen



b) nach Altersgruppen



Bei der Zahl der Unter-1-Jährigen ist zu beachten, dass dieser Jahrgang nur zu 80% besetzt ist, da bedingt durch das Studiendesign erst Kinder ab einem Alter von etwa 10 Wochen in die Studie aufgenommen wurden.

Die Stichprobengewichte nehmen Werte von 0,13 bis 4,65 an und weisen einen Variationskoeffizienten von 53% auf, was einem Effizienzmaß von 78% entspricht. Ausländer werden im Mittel um 9% hoch gewichtet (mittleres Gewicht 1,09), Probanden mit deutscher Staatsangehörigkeit um 1% herunter gewichtet (mittleres Gewicht 0,99). Probanden im Osten werden zum Ausgleich des Oversamplings um mehr als die Hälfte herunter gewichtet (mittleres Gewicht 0,44, im Westen 1,25, in Berlin 1,15).

Eine detaillierte Beschreibung der GewichtungsvARIABLEN ist auf Anfrage erhältlich.

6.2.2. Gewichtungsvariablen für Laborparameter

Für die Analyse von Laborwerten stehen die Gewichtungsvariablen **wKGS_{Lab}** und **wKGS_{VitD}** zur Verfügung. Da der Anteil der Kleinkinder im Alter von 1 und 2 Jahren, die keine Blutprobe abgegeben haben, recht hoch ist, wurde eine gesonderte Gewichtungsvariable zur Auswertung von Laborparametern bestimmt. Diese Variable (**wKGS_{Lab}**) berücksichtigt bei der Herleitung der Gewichte nur die Probanden, für die eine Blutprobe vorliegt. Bei Vitamin D können wegen eines Methodenwechsels nicht alle Probanden ausgewertet werden, so dass hier eine gesonderte Gewichtungsvariable (**wKGS_{VitD}**) gebildet wurde.

6.2.3. Verwendung der Gewichte in den Auswertungen

Bei allen deskriptiven KiGGS-Auswertungen müssen die Stichprobengewichte **wKiGGS** verwendet werden, um deutschlandweit repräsentative Aussagen zu erzielen.

Es empfiehlt sich bei der Analyse von kategoriellen Variablen, keine gewichteten Absolutzahlen zu veröffentlichen, da diese Zahlen eigentlich nicht existieren und in gewissem Sinne rein fiktiv sind. Stattdessen sollte die ungewichtete Zahl der Probanden angegeben werden, die in die jeweilige Analyse (ggf. auch Subgruppe) eingehen, sowie die gewichteten Prozente.

Auch für Zusammenhangsanalysen (z.B. Regressionsmodelle) sollte in der Regel gewichtet gerechnet werden. Hier kann zusätzlich eine ungewichtete Analyse als Sensitivitätsanalyse durchgeführt werden. Sollten sich wesentliche Unterschiede zwischen der gewichteten und der ungewichteten Analyse ergeben, so ist dies in erster Linie ein Hinweis darauf, dass Wechselwirkungen der untersuchten Variablen mit Ost-West oder mit dem Alter bestehen, die im Modell nicht ausreichend berücksichtigt sind.

Bei einer Variablen mit einem hohen Anteil von fehlenden Angaben empfiehlt es sich, die Verteilung der Gewichtsvariablen zwischen den Probanden mit vorhandenen und fehlenden Angaben zu vergleichen. Falls sich hier Unterschiede ergeben, sind Verzerrungen hinsichtlich der Gewichtungsvariablen in Analysen zu erwarten, die sich auf die Probanden mit vollständigen Angaben beschränken.

Statistische Programmpakete sind in der Regel in der Lage, Gewichtungsvariablen bei der Auswertung zu berücksichtigen. Allerdings ist zu beachten, dass manche Prozeduren nur ganzzahligen Gewichtungsvariablen verwenden; dies führt zu falschen Ergebnissen. Die folgenden Anmerkungen beziehen sich auf zwei spezielle Programmpakete und erheben keinen Anspruch auf Vollständigkeit.

In SPSS wird mit der Prozedur „Fälle gewichten“ (WEIGHT = wKiGGS) der Gewichtungsfaktor für alle nachfolgenden Berechnungen wirksam. Bei speziellen SPSS-Prozeduren (GLM, MIXED, REGRESSION, UNIANOVA, VARCOMP) führt diese globale Einstellung allerdings zu fehlerhaften Ergebnissen; hier muss anstelle des globalen WEIGHT-Befehls das Subkommando /REGWGT wKiGGS gesetzt werden. Bei der Berechnung von gewichteten Perzentilen in SPSS ist zu beachten, dass unterschiedliche Prozeduren teilweise unterschiedliche Algorithmen zur Berechnung gewichteter Perzentile verwenden.

In SAS lassen sich gewichtete Analysen mit den meisten statistischen Prozeduren durchführen. Man verwendet dazu das WEIGHT-Statement. Bei der Berechnung von Standardabweichungen oder Varianzen ist außerdem die Option VARDEF=WGT oder VARDEF=WDF anzugeben. Die Prozedur PROC TABULATE macht keine gewichteten Analysen.

6.3. Ost-West-Vergleiche

Das Oversampling Ost wird durch die Gewichtung insofern ausgeglichen, als Probanden im Osten entsprechend heruntergewichtet werden. Bei einer Varianzschätzung, die wie im folgenden Absatz beschrieben auf den Sample Points als Basiseinheiten beruht, bleibt der beim Oversampling erwünschte Anstieg der Power für Ost-West-Vergleiche erhalten, da das Oversampling erreicht wurde, indem im Osten mehr Sample Points gezogen wurden (und nicht etwa mehr Probanden innerhalb der Sample Points). Es ist in diesem Fall also keine eigene Gewichtung für Ost-West-Vergleiche erforderlich. Bei der Berechnung von p -Werten und Konfidenzintervallen auf konventionelle Art und Weise kommt der Gewinn durch das Oversampling Ost in gewichteten Analysen nicht zum Tragen.

Bei Ost-West-Vergleichen ist ferner zu beachten, dass bei diesen Vergleichen das Alter in der Analyse berücksichtigt werden muss, da sich die Altersstruktur der Kinder und Jugendlichen in Ost und West unterscheidet, wie oben bereits beschrieben.

6.4. Varianzschätzung, Berechnung von p -Werten und Konfidenzintervallen

Die KiGGS-Daten wurden in einer zweistufigen Klumpenstichprobe erhoben, bei der Probanden innerhalb von Sample Points (Untersuchungsorten) gezogen wurden. Die daraus resultierende Clusterung der Probanden ist bei der Varianzschätzung und damit bei der Berechnung von p -Werten und Konfidenzintervallen zu berücksichtigen, da eine der Grundannahmen der standardmäßig verwendeten statistischen Verfahren, nämlich die Unabhängigkeit der Beobachtungen, nicht gegeben ist. Mit Standardverfahren erhält man i.A. anti-konservative Ergebnisse, d.h. die Konfidenzintervalle sind zu schmal und die p -Werte zu klein, so dass fälschlicherweise Signifikanzen angezeigt werden können, die in Wahrheit gar nicht bestehen.

Eine Möglichkeit, diese Clusterung zu berücksichtigen, stellen sogenannte Survey-Prozeduren dar, die auf der Stichprobentheorie beruhen und mittlerweile in viele Statistik-Pakete integriert sind (z.B. in SAS, SPSS, STATA oder dem R-Package *survey*). Diese Verfahren können sowohl für deskriptive Analysen als auch für Regressionsmodelle (lineare und logistische Regression) angewendet werden. Sie berücksichtigen die Clusterung der Probanden in den Sample Points und die daraus resultierende Korrelation der Probanden innerhalb eines Sample Points. Die Varianz wird dabei nicht aus der Variabilität der Merkmalswerte zwischen den Probanden geschätzt, sondern aus der Variabilität der Sample-Point-spezifischen Mittelwerte bzw. Häufigkeiten. Basis der Varianzschätzung ist also nicht der Proband, sondern der Sample Point.

Diese Verfahren berücksichtigen nicht die bei der Ziehung der Sample Points hinterlegte Verteilung nach Gemeindetypen und geografischer Region sowie die Ziehung ohne Zurücklegen und dürften von daher leicht konservativ sein.

Für Regressionsmodelle sind alternativ auch andere Verfahren zur Berücksichtigung der Korrelation innerhalb der Sample Points nutzbar, z.B. Modelle mit zufälligen Effekten oder Multi-Level-Modelle.

Im Folgenden wird als Beispiel die Verwendung der Survey-Prozeduren in SPSS und SAS zur Auswertung von KiGGS-Daten kurz beschrieben.

6.4.1. Syntax in SPSS

In SPSS erfolgen surveyspezifische Auswertungen mit dem Modul Complex Samples, für das eine eigene Lizenz benötigt wird. Die Prozeduren aus diesem Modul benötigen eine sog. Plandatei, die die notwendigen Informationen zum Stichprobendesign erhält. Diese Plandatei kann mit folgender Syntax erzeugt werden, wobei unter FILE= der Pfad und Dateiname angegeben werden, unter dem die Plandatei gespeichert werden soll.

```

CSPLAN ANALYSIS
/PLAN FILE='Pfad_und_Dateiname'
/PLANVARS ANALYSISWEIGHT=wKiGGS
/DESIGN CLUSTER=Ppoint
/ESTIMATOR TYPE=WR.

```

Falls WEIGHT BY wKiGGS eingestellt ist und Complex Samples angewendet wird, erscheint eine Warnung, dass die Gewichtungvariable nicht in der Analyse berücksichtigt wird. Dies ist jedoch unerheblich, da anstelle des WEIGHT-Befehls die Plandatei in der Analyse berücksichtigt wird und in der Plandatei die Information zur Gewichtungsvariablen enthalten ist.

In den Complex Samples-Prozeduren gibt man dann mit der Option

```
/PLAN FILE='Pfad_und_Dateiname_der_Plandatei'
```

den Namen und den Pfad der einmal erzeugten Plandatei an. Die Plandatei kann auch bei einer menügesteuerten Analyse über Analysieren – Komplexe Stichproben angegeben werden.

6.4.2. Syntax in SAS

In SAS stehen zur Analyse von Surveydaten die Prozeduren PROC SURVEYMEANS, PROC SURVEYFREQ, PROC SURVEYREG und PROC SURVEYLOGISTIC zur Verfügung.

Bei allen diesen Prozeduren werden die Informationen zum Stichprobendesign durch die folgenden beiden Statements übergeben:

```

CLUSTER Ppoint;
WEIGHT wKiGGS;

```