

Opbouw en instructie totaalbestand Gezondheidsmonitor Jeugd 2019

17 december 2020 versie 2.0

Auteurs: Marieke Hiemstra (marieke.hiemstra@rivm.nl)
Bianca Sniijders (bianca.sniijders@rivm.nl)

1. Achtergrond
 2. Onderwerpen in het totaalbestand
 3. Verwerking van databestanden tot een totaalbestand
 4. Variabelen in het totaalbestand
 - 4.1 Identificatie-variabelen
 - 4.2 Geografische indicatoren
 - 4.3 Weegfactoren
 - 4.4 Methode van invullen
 - 4.5 Achtergrondkenmerken
 - 4.5.1 Geslacht en leeftijd
 - 4.5.2 Gezinsamenstelling
 - 4.5.3 Opleiding
 - 4.6 Uitkomstvariabelen
 - 4.6.1 Moeite met rondkomen
 - 4.6.2 Ervaren gezondheid
 - 4.6.3 Geluk
 - 4.6.4 Psychosociale gezondheid
 - 4.6.5 Opkomen voor jezelf (weerbaarheid)
 - 4.6.6 Stress
 - 4.6.7 Ingrijpende gebeurtenissen
 - 4.6.8 Bewegen
 - 4.6.9 Voeding (Water drinken, suikerhoudende dranken, energiedrankjes)
 - 4.6.10 Roken actief en e-sigaret gebruik
 - 4.6.11 Alcoholgebruik/Alcoholgebruik binge en aangeschoten
 - 4.6.12 Gebruik wiet of hasj
 - 4.6.13 Pesten op school
 - 4.6.14 Cyberpesten
 - 4.6.15 Sexting
 - 4.6.16 Sociale mediagebruik
 - 4.6.17 Gamen
 - 4.6.18 Seksualiteit
 5. Complex Samples in SPSS
- Bijlagen:
- Bijlage 1 Overzicht steekproefniveaus GGD-vragenlijsten per GGD-regio
 - Bijlage 2 Responsaantallen per GGD-regio
 - Bijlage 3 Notitie Weging Gezondheidsmonitor Jeugd

1. Achtergrond

In dit opbouw en instructie document wordt een toelichting gegeven op het totaalbestand voor derden van de Gezondheidsmonitor Jeugd 2019. Dit document is bestemd voor alle gebruikers van dit bestand. Gebruikers zullen overigens niet altijd toegang hebben tot een totaalbestand met alle variabelen erin. Tot welke variabelen de gebruiker toegang heeft, zal afhangen van de data-aanvraag en toekenning door de registratiecommissie.

Naast dit document is er een codeboek aanwezig voor de basisvragenset. De vragenlijst is beschikbaar via www.monitoregezondheid.nl. Voor vragen over gezondheidsmonitor jeugd kan contact worden opgenomen via gezondheidsmonitor@rivm.nl.

Het landelijke totaalbestand van de Gezondheidsmonitor Jeugd 2019¹ bevat vragenlijst- en bewerkte gegevens over (determinanten van) de gezondheid, sociale situatie en leefstijl van 171.192 personen van 12-18 jaar (klas 2 en Klas 4 VO). Het is een landelijk bestand dat tot stand is gekomen uit een samenwerking van alle GGD'en, GGD GHOR NL en het RIVM. Hierdoor zijn geharmoniseerde cijfers over de gezondheid en gezondheidsbeleving van de Nederlandse Jeugd beschikbaar.

Voor de Gezondheidsmonitor Jeugd 2019 heeft er afstemming plaatsgevonden tussen de GGD'en waarbij vraagstellingen en methode zo veel mogelijk op elkaar zijn afgestemd. Dit heeft geleid tot een basisvragenlijst van onderwerpen waarvoor landelijke en regionaal cijfers beschikbaar zijn voor alle GGD'en (N=25). Voor een groot aantal GGD'en zijn ook cijfers op gemeentelijk niveau beschikbaar (N=19). De gegevens van deze basisvragenlijst zijn opgenomen in het totale databestand. Het RIVM heeft een coördinerende rol in de realisatie van de Gezondheidsmonitor Jeugd.

2. Onderwerpen in het totaalbestand

De geharmoniseerde Gezondheidsmonitor Jeugd 2019 bevat gegevens over een basisvragenlijst van onderwerpen. De onderwerpen in de basisvragenlijst zijn vragenlijsten die door alle GGD'en met dezelfde vraagstellingen nagevraagd. Naast de vraagstellingen in de basisvragenlijst kon elke GGD ervoor kiezen om de basisvragenlijst uit te breiden met extra vraagstellingen, bijvoorbeeld op verzoek van de gemeenten. Dit kunnen door GGD'en samen ontwikkelde standaardvraagstellingen zijn (bij voorkeur) dan wel lokaal ontwikkelde vraagstellingen (indien er geen standaard vraagstelling voorhanden is). De extra vragen buiten de basisvragenlijst zijn niet opgenomen in het totaalbestand van de Gezondheidsmonitor Jeugd 2019.

De vragen zijn ingedeeld in blokken, waarbij na de achtergrondvragen, 'van binnen naar buiten' wordt geredeneerd; van de jongeren zelf, naar een steeds bredere omgeving:

Blok 1 betreft achtergrondvragen (methode van invullen, opleiding, leeftijd, geslacht, geografische indicator, gezinssamenstelling, moeite met rondkomen).

Blok 2 betreft vragen over het gevoel van de jongere zelf (ervaren gezondheid, geluk, psychosociale gezondheid, opkomen voor jezelf (weerbaarheid), stress, ingrijpende gebeurtenissen).

Blok 3 betreft vragen over gedrag van de jongere zelf (bewegen, voeding-water, voeding- suikerhoudende drankjes, voeding-energie drankjes, alcoholgebruik, roken actief, gebruik e-sigaret, gebruik wiet of hasj)

Blok 4 betreft vragen over pesten (pesten op school, cyberpesten en sexting).

Blok 5 betreft vragen over de jongere in relatie tot anderen (sociale mediagebruik, gamen past misschien niet helemaal, maar volgt logischerwijs na sociale media gebruik; seksualiteit).

3. Verwerking van databestanden tot een totaalbestand

¹ De officiële benaming is: Gezondheidsmonitor Jeugd 2019, GGD'en en RIVM.

De benaming in het Engels is: Public Health Monitor Youth 2019 of the Community Health Services, and the National Institute for Public Health and the Environment

Het RIVM heeft de verwerking van de databestanden gedaan tot één totaalbestand. Hiervoor zijn de volgende stappen ondernomen: aan het eind van de dataverzameling zijn alle ruwe regionale bestanden gedownload. Vervolgens zijn alle respondenten buiten de doelgroep van de GM Jeugd uit het databestand gehaald. Daarna zijn voor de weegfactoren (klas, onderwijsniveau en gemeentecode) de variabelen opgeschoond en waar mogelijk aangevuld met informatie die door de GGD'en aangeleverd was. Tot slot zijn alle syntaxen om data te schonen en bewerken op vraagstelling- en respondent-niveau gerund. Hiervoor was er per onderwerp een aparte syntax. Met deze syntaxen zijn de data opgeschoond en zijn afgeleide variabelen aangemaakt. Van te voren zijn deze syntaxen afgestemd met de GGD'en. De syntaxen zijn voor derden (per onderwerp) aan te vragen via gezondheidsmonitor@rivm.nl².

Na het opschonen voegt het RIVM alle regionale bestanden samen tot het totale landelijke databestand

Daarna wordt de weegfactor berekend. Deze is berekend op basis van achtergrondkenmerken om te zorgen dat de prevalentiecijfers zo goed mogelijk de werkelijkheid weergeven. Leerlingen die in de dataset zijn ondervertegenwoordigd tellen zwaarder mee dan andere leerlingen. Elke leerling die de monitor-vragenlijst heeft ingevuld krijgt daarom een weegfactor die aangeeft hoe zwaar de betreffende leerling moet meetellen in de regionale en landelijke cijfers. In paragraaf 4.3 wordt verder ingegaan op de weegfactoren. Voor een uitgebreide beschrijving van de manier waarop de wegingen door het RIVM zijn uitgevoerd, wordt verwezen naar Bijlage 3 Notitie Weging Gezondheidsmonitor Jeugd.

Er bestaan 3 verschillende versies van het databestand:

- 1) Landelijk databestand voor GGD'en: Totaalbestand GM Jeugd 2019 - naar GGDen - versie 6.0 (N=171.192).sav
- 2) Landelijk databestand voor RIVM,GGD-GHOR en derden: Totaalbestand GM Jeugd 2019 – RIVM-GGD-GHOR en derden- versie 6.0 (N=171.192).sav
Voor derden wordt uit dit bestand een selectie gemaakt van variabelen die ze nodig hebben.
- 3) Regionale databestand voor elke GGD: GM Jeugd 2019 - 01 - GGD NAAM - versie 6.0.sav

Wat betreft de steekproefopzet dient opgemerkt te worden dat deze verschilde tussen GGD-regio's. De GGD'en konden hierbij kiezen uit 2 opties. Bij de eerste optie zijn alle scholen benaderd met het verzoek om mee te doen welke heeft geleid tot het genereren van gemeentelijke cijfers bij deze GGD'en (n=19). Bij de tweede optie werd de vragenlijst afgenomen bij een steekproef van tweede en vierde klassen welke heeft geleid tot regionale cijfers (n=6). Bij het berekenen van de weegfactoren is rekening gehouden met beide opties, d.w.z. dat alleen de GGD'en die meededen aan de eerste optie ook gemeentelijke weegfactoren krijgen. In bijlage 1 wordt een overzicht gegeven van de steekproefniveaus per GGD-regio. In bijlage 2 staan de uiteindelijke responsaantallen per GGD-regio weergegeven.

Publicaties

Indien cijfers van de Gezondheidsmonitor Jeugd 2019 gebruikt worden bij publicaties, dient de volgende bronvermelding te worden opgenomen: Gezondheidsmonitor Jeugd 2019, GGD'en en RIVM.

² Tevens is er per onderwerp een achtergronddocument beschikbaar met achtergrondinformatie over de vraagstelling (voor GGD'en via monitoregezondheid.nl en voor derden aan te vragen via gezondheidsmonitor@rivm.nl).

4. Variabelen in het totaalbestand

De variabelen in het totaalbestand zijn onderverdeeld in identificatie-variabelen, weegfactoren, geografische indicatoren, variabelen die betrekking hebben op de manier van invullen en de taal van de vragenlijst en ten slotte in achtergrondkenmerken en uitkomstvariabelen.

4.1 Identificatie-variabelen

Overzicht identificatievariabelen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
Onderzoeksjaar	Jaar waarin de vragenlijst is afgenomen	2019
@_fd_usrnm	Uniek respondentnummer (koppelsleutel)	2 letters van GGD regio, 3 cijfers schoolcode en 4 cijfers persoonlijke ID
@_fd_add	Datum invoer	
@_fd_edit	Datum laatste wijziging	
@_fd_compl	Datum gereed	Gereed Geannuleerd Te verzenden
@_fd_status	Verzend status vragenlijst	
aantal_missings	Aantal missings op de eerste vragen van de 22 standaarden (respondenten met 22 of meer missings zijn verwijderd)	

Het totaalbestand bevat een aantal variabelen dat betrekking heeft op de identificatie van respondenten. Let wel, hiermee wordt niet bedoeld dat respondenten herleidbaar zijn. Het is voor gebruikers van het totaalbestand niet mogelijk om specifieke individuen te herleiden.

'@_fd_usrnm' is de koppelsleutel, unieke respondentnummer van alle respondenten in het bestand. Uitzondering hierop zijn in Twente. Hierin zitten een aantal dubbele ID codes omdat per abuis een aantal klassen allemaal dezelfde inlogcode kregen van de docent: TW2155469; TW2185623; TW2186668.

'@_fd_add' geeft datum en tijdstip aan op welk moment de vragenlijst voor het eerst opgeslagen is op de server van Formdesk.

'@_fd_edit' geeft datum en tijdstip van wanneer de vragenlijst voor de laatste keer naar de Formdesk server gezonden is.

'@_fd_compl' zijn de datum en tijdstip waarop de scholier de vragenlijst met de verzendknop definitief heeft ingezonden.

'@_fd_status' is de verzendstatus van de vragenlijst. Deze kan zijn "Gereed" (op de verzendknop gedrukt om de gehele vragenlijst in te zenden), "Geannuleerd" (Op de annuleren knop gedrukt of op kruisje van de browser) of "Te verzenden" zijn (de vragenlijst eerder tussentijds heeft bewaard en de vragenlijst niet heeft verzonden met de verzendknop).

'aantal_missings' geeft het aantal missing weer op de eerste vragen van 22 basisstandaarden. Respondenten met 22 missings zijn verwijderd uit het data bestand.

4.2 Geografische indicatoren

Overzicht geografische variabelen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
mbgik401**	Wat zijn de 4 postcodecijfers van je adres?	0 = onbekend
POSTCODE4**	Wat zijn de 4 postcodecijfers van je adres? opgeschoond	0 onbekend, 1000-9999
MBGIK402**	Wat is je woonplaats?	Opklapscherm met keuze uit woonplaatsen
GGDregio_school	GGD-regio waarin leerling naar school gaat	1 = GGD Groningen 2 = GGD Fryslan 3 = GGD Drenthe 4 = GGD IJsselland 5 = GGD Twente 6 = GGD Noord-en Oost-Gelderland 7 = VG Gelderland-Midden 8 = GGD Gelderland Zuid 9 = GGD regio Utrecht 10 = GGD Hollands Noorden 11 = GGD Zaanstreek-Waterland 12 = GGD Kennemerland 13 = GGD Amsterdam 14 = GGD Gooi en Vechtstreek 15 = GGD Haaglanden 16 = GGD Hollands Midden 17 = GGD Rotterdam-Rijnmond 18 = DGJ Zuid-Holland Zuid 19 = GGD Zeeland 20 = GGD West-Brabant 21 = GGD Hart voor Brabant 22 = GGD Brabant-Zuidoost 23 = GGD Limburg-Noord 24 = GGD Zuid-Limburg 25 = GGD Flevoland.
GGDregio_leerling_2018	GGD-regio waarin leerling woont 2018	Zie GGDregio_school
GGDregio_leerling_2019	GGD-regio waarin leerling woont 2019	Zie GGDregio_school
Schoolcode_uniek	Schoolcode, versleuteld	1-708
Schoolcode_BRIN**	Schoolcode omgezet naar BRIN	None
Schoolcode_BRIN2**	Schoolcode omgezet naar BRIN indien 2 BRIN locaties	None
Schoolcode_BRIN3**	Schoolcode omgezet naar BRIN indien 3 BRIN locaties	None

GEMEENTECODE***	Gemeentecode 2018 leerling obv ingevulde postcode	
GEMEENTECODE2018***	Gemeentecode 2018 leerling obv ingevulde postcode, opgeschoond	
GEMEENTECODE2019***	Gemeentecode 2019 leerling obv ingevulde postcode, opgeschoond	
Stedelijkheid2018	Stedelijkheid 2018 gemeente leerling obv ingevulde postcode	1 = Zeer sterk 2 = Sterk 3 = Matig 4 = Weinig 5 = Niet
Stedelijkheid2019	Stedelijkheid 2019 gemeente leerling obv ingevulde postcode	1 = Zeer sterk 2 = Sterk 3 = Matig 4 = Weinig 5 = Niet

*Niet opgenomen in totaalbestand voor externe gebruikers. **Alleen in GGD regio bestand *** alleen in landelijk bestand GGDGHOR, RIVM, derden en regio bestand per GGD.
#afgeleide

‘MBGIK401’ is de originele vraag. POSTCODE4 zijn mensen met postcodes met minder of meer dan 4 cijfers wordt op missing gezet. Dit is de opgeschoonde variabele. ‘MBGIK402’ is een samen genomen variabele van woonplaats gebaseerd op de woonplaats variabele voor afname op mobiele telefoon en computer (anders waren dit 27 losse woonplaats vragen (26 mobiel (alle letters van het alfabet) + 1 voor computer dropdown menu)

De variabele ‘GGDregio_schoonl’ geeft aan in welke van de 25 GGD-regio’s een leerling naar school gaat.

De variabele ‘GGDregio_leerling_2018’ geeft aan in welke van de 25 GGD-regio’s een leerling woont. Deze is gebaseerd op gemeentecode van 2018. Deze variabele is niet geschikt voor het presenteren van regio cijfers.

Deze variabele ‘GGDregio_leerling_2019’ geeft aan in welke van de 25 GGD-regio’s een leerling woont. Deze is gebaseerd op gemeentecode van 2019 welke rekening houdt met een aantal herindelingen. Deze variabele is geschikt voor het presenteren van regio cijfers.

‘Schoolcode_uniek’ geeft aan welke leerlingen geclusterd zijn op een school. In totaal hebben er 707 schoollocaties meegedaan met GMJ 2019. In het databestand zie je 708 school locaties. Voor GGD Limburg-Noord zijn echter 2 schoollocaties dezelfde school, het zijn dan in totaal 707 schoollocaties.

‘Schoolcode_BRIN’ geeft de BRINcode + vestigingsnummer per schoollocaties weer. Sommige schoollocaties hebben meerdere BRINcodes+ vestigingsnummer deze worden weergegeven in ‘Schoolcode_BRIN2’, ‘Schoolcode_BRIN3’. Deze variabele is alleen beschikbaar in de GGD regio bestanden en alleen voor deelnemers uit eigen regio. Voor grensoverschrijders is dit omgezet naar ‘XXXXXX’.

‘Gemeentecode’ is de gemeentecode leerling op basis van ingevulde postcode en woonplaats. Deze is gebaseerd op de gemeentelijke indeling van 2018. De variabele ‘Gemeentecode2018’ is op geschoond en aangevuld met gemeentecodes die door de GGD’en per schoolcode zijn aangeleverd. ‘Gemeentecode 2019’ is gebaseerd op ‘ GEMEENTECODE2018’ waarbij gecorrigeerd is voor de

gemeentelijke herindeling van 2019 (zie <https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-jaar/indeling-per-jaar/gemeentelijke-indeling-op-1-januari-2019>).

‘Stedelijkheid2018’ geeft de stedelijkheid van de gemeente aan waar de leerling woont gebaseerd op de ingevulde postcode/plaats met bijbehorende gemeentecode. Dit is de stedelijkheid gebaseerd op gemeentecode van 2018.

‘Stedelijkheid2019’ geeft de stedelijkheid van de gemeente aan waar de leerling woont gebaseerd op de ingevulde postcode/plaats met bijbehorende gemeentecode. Dit is de stedelijkheid gebaseerd op gemeentecode van 2019 rekening houdend met een aantal gemeentelijke herindeling. Stedelijkheid is gebaseerd op informatie van het CBS over de stedelijkheid per gemeentecode (<https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83859NED/table?dl=34AC5>).

Voor 6 GGD regio’s (die meededen op regionaal/landelijk niveau) kunnen geen gemeentecijfers worden gepresenteerd. Het kan zijn dat een gemeente wel voldoende leerlingen voor gemeentecijfers hadden, maar deze cijfers zijn niet representatief.

4.3 Weegfactoren

Overzicht van de weegfactoren in het totaalbestand

Variabele naam	Label/Toelichting	Codering
Weefactor_landelijk	Weefactor, voor landelijke en regionale cijfers	None
Weefactor_gemeentelijk	Weefactor, voor gemeentecijfers	None
Weefactor_gemeentelijk_samengesteld*	Weefactor, voor gemeentecijfers van samengestelde gemeenten op verzoek van GGD bij te lage n voor weefactor	None
Stratum	Eenheid waarmee weefactoren worden berekend, gebruiken bij Complex Samples in SPSS	None

* deze variabele is niet beschikbaar voor derden.

Om voor verschillen tussen de samenstelling van de respondentengroep en de totale bevolking te corrigeren, heeft het RIVM, afdeling Statistiek Informatica en Modellerie (SIM) weefactoren aangemaakt voor de Gezondheidsmonitor Jeugd 2019. Deze weefactoren zijn gemaakt voor het berekenen van de regionale en landelijke cijfers. Voor een uitgebreide beschrijving van de berekening van de weefactor, zie bijlage 3.

Variabele 'Weefactor_landelijk' is de weefactor voor de landelijke en regionale cijfers. Deze variabele zit in de landelijke databestanden voor GGD'en en landelijk databestand RIVM, GGD-GHOR en derden.

Variabele 'Weefactor_gemeentelijk' is de weefactor voor gemeentecijfers. Deze variabele zit in het regio bestand voor de GGD'en in het landelijk bestand voor RIVM, GGD-GHOR en derden.

Voor gemeenten met te weinig respondenten is het niet mogelijk om te wegen. Bij de variabele 'Weefactor_gemeentelijk_samengesteld' is er voor gemeenten die niet genoeg leerlingen hadden voor representatieve gemeentecijfers gebruik gemaakt van gecombineerde gemeenten. De keuze voor het wel of niet combineren en welke gemeenten onderling is een keuze van de betreffende GGD bijvoorbeeld gebaseerd op al bestaande samenwerking tussen gemeenten en/of geografische ligging of een vergelijkbare populatie op basis van leeftijd en opleidingsniveau. De volgende gecombineerde gemeentecijfers kunnen gemaakt worden: Laren-Blaricum, Bergen-Gennep, Appingedam-Delftziel-Loppersum, Pekela-Veendam, Oldambt-Westerwolde, Vaals-Sittard-Geleen. Deze samengestelde gemeentelijke weefactoren zit in het regio bestand voor GGD'en in data bestand voor RIVM, GGD-GHOR en derden. Echter is deze niet voor derden beschikbaar.

Voor enkele gemeenten met een te lage n is door de betreffende GGD gekozen om geen combinatie te maken met een andere gemeente. Dit betreft Renswoude, Staphorst, Raalte, Haaksbergen, en Gorinchem. Hiervan kunnen dus geen gewogen gemeentecijfers worden gepresenteerd.

'Stratum' geeft de steekproef stratificatie weer. Er is voor de GGD regio's die mee doen op regionaal niveau per GGD een steekproef van scholen getrokken. GGD-regio is de eerste variabele waarop gestratificeerd wordt. Vervolgens zijn er binnen die scholen klassen (= leerlingen) getrokken/aangewezen. Van elk schoolniveau en leerjaar. Dit laatste geldt overigens alleen voor de regio's waar steekproeven van scholen zijn getrokken. Voor de regio's waar alle scholen zijn

benaderd (=gemeente niveau) is er maar 1 stratum, namelijk de GGD-regio. In SPSS Complex Samples wordt gevraagd om rekening te houden met de manier waarop de steekproef is getrokken. Hiervoor kan deze variabele gebruikt worden (zie hoofdstuk 5: Complex Samples in SPSS).

4.4 Manier van invullen

Overzicht van de variabelen m.b.t. de manier van invullen

Variabele naam	Label/Toelichting	Codering
MIMIK306	Methode van invullen vd vragenlijst	1= op een pc of laptop 2 = op een tablet 3 = op een smartphone 4 = op papier

* = aangemaakte variabele, indicator

4.5 Achtergrondkenmerken

4.5.1 Geslacht en leeftijd

Overzicht van de variabelen m.b.t. geslacht en leeftijd in het totaalbestand

Variabele naam	Label/Toelichting	Codering
BBLFK321	Hoe oud ben je?	1 = 12 jaar of jonger 2 = 13 jaar 3 = 14 jaar 4 = 15 jaar 5 = 16 jaar 6 = 17 jaar 7 = 18 jaar en ouder
BBLFA321*	Leeftijd in jaren	12 = 12 jaar of jonger 13 = 13 jaar 14 = 14 jaar 15 = 15 jaar 16 = 16 jaar 17 = 17 jaar 18 = 18 jaar en ouder
BBGSK401	Ben je een jongen of een meisje?	1 = Jongen 2 = Meisje
GESLACHT*	Geslacht	1 = Jongen 2 = Meisje

* = aangemaakte variabele, indicator

'BBLFK321' en 'BBGSK401' zijn de ruwe variabelen. 'BBLFA321' en 'GESLACHT' zijn de indicatoren.

4.5.2 Gezinsamenstelling

Overzicht van variabelen m.b.t. gezinsamenstelling in het totaalbestand

Variabele naam	Label/Toelichting	Codering
MBGSK321	Bij wie woon je de meeste dagen vd week?	1 = Bij mijn ouders (samen) 2 = Ongeveer de helft van de tijd bij mijn ene ouder en de helft van de tijd bij mijn andere ouder (co-ouders) 3 = Bij mijn moeder en haar partner 4 = Bij mijn vader en zijn partner 5 = Alleen bij mijn moeder 6 = Alleen bij mijn vader 7 = Bij anderen (bijv. pleegouders, andere familie, internaat) 8 = Ik woon op mezelf
MBGSK3S2*	Gezinsamenstelling aangepast	1 = Bij beide ouders 2 = Co-ouders 3 = Vader of moeder en partner 4 = Eenoudergezin 5 = Bij anderen of op zichzelf

* = aangemaakte variabele, indicator

'MBGSK321' is gehercodeerd van 8 naar 5 categorieën in variabele 'MBGSK3S2'

4.5.3 Opleiding

Overzicht variabelen m.b.t. opleiding in het totaalbestand

Variabele naam	Label/Toelichting	Codering
MBOKK331	In welke klas/leerjaar zit je?	1 t/m 6
KLAS*	Klas, opgeschoond	1 t/m 6 (in de praktijk 2 of 4)
mbokk341	Welk soort onderwijs volg je?	1 t/m 7 meerdere antwoorden mogelijk
MBOKK332	Ik volg vmbo-b (basis)	1 = Aangekruist 2 = Niet aangekruist
MBOKK333	Ik volg vmbo-k (kader)	1 = Aangekruist 2 = Niet aangekruist
MBOKK334	Ik volg vmbo-g (gemengd)	1 = Aangekruist 2 = Niet aangekruist
MBOKK335	Ik volg vmbo-t (theoretisch, mavo)	1 = Aangekruist 2 = Niet aangekruist
MBOKK336	Ik volg havo	1 = Aangekruist 2 = Niet aangekruist
MBOKK337	Ik volg vwo (atheneum, gymnasium)	1 = Aangekruist 2 = Niet aangekruist
MBOKK338	Ik volg ander soort onderwijs	1 = Aangekruist 2 = Niet aangekruist
MBOKA332	Ik volg vmbo-b (basis) (opgeschoond)	1 = Aangekruist 2 = Niet aangekruist
MBOKA333	Ik volg vmbo-k (kader) (opgeschoond)	1 = Aangekruist 2 = Niet aangekruist
MBOKA334	Ik volg vmbo-g (gemengd) (opgeschoond)	1 = Aangekruist 2 = Niet aangekruist
MBOKA335	Ik volg vmbo-t (theoretisch, mavo) (opgeschoond)	1 = Aangekruist 2 = Niet aangekruist
MBOKA336	Ik volg havo (opgeschoond)	1 = Aangekruist 2 = Niet aangekruist
MBOKK3S31*	Onderwijssoort in 2 groepen	1 = Vmbo 2 = Havo/Vwo
MBOKK3S32*	Onderwijssoort in 3 groepen	1 = Vmbo 2 = Havo 3 = Vwo
MBOKK3S33*	Onderwijssoort in 4 groepen	1 = Vmbo-bk 2 = Vmbo-gt 3 = Havo 4 = Vwo
MBOKA3S31*	Onderwijssoort in 2 groepen, opgeschoond	1 = Vmbo 2 = Havo/Vwo
MBOKA3S32*	Onderwijssoort in 3 groepen, opgeschoond	1 = Vmbo 2 = Havo 3 = Vwo
MBOKA3S33*	Onderwijssoort in 4 groepen, opgeschoond	1 = Vmbo-bk 2 = Vmbo-gt 3 = Havo 4 = Vwo

* = aangemaakte variabele, indicator

De variabele 'MBOKK331' geeft aan in welke klas/leerjaar een leerling zit. 'KLAS' is de opgeschoonde variant hiervan. 'mbokk341' is de ruwe variabelen van het opleidingsniveau die ze volgen. Deze is omgezet naar aparte variabelen per onderwijs niveau 'MBOKK332 t/m MBOKK338. Voor 'MBOKA332 t/m MBOKA336' zijn deze opgeschoond. Vervolgens is onderwijs samengenomen in 2, 3 of 4 onderwijssoort groepen in de variabelen 'MBOKK3S31' 'MBOKK3S32' 'MBOKK3S33'. Waarbij 'MBOKA3S31, MBOKA3S32 en MBOKA3S33' de opgeschoonde variabelen zijn.

4.6 Uitkomstvariabelen

4.6.1. Moeite met rondkomen

Overzicht van de variabelen m.b.t. moeite met rondkomen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
mbmrk301	Thuis moeite om rond te komen met geld?	1 = Nee, geen enkele moeite 2 = Nee, geen moeite maar we moeten wel opletten 3 = Ja, enige moeite 4 = Ja, grote moeite
LBARK3S1*	Moeite met rondkomen in 2 categorieën	1 = Ervaart geen enkele moeite of geen moeite maar wel opletten 2 = Ervaart enige of grote moeite
LBARK3S2*	Moeite met rondkomen in 3 categorieën	1= Ervaart geen enkele moeite 2= Ervaart geen moeite maar wel opletten 3= Ervaart enige of grote moeite
LBARK3S3*	Ervaart grote moeite met rondkomen	0 = Nee (ervaart geen (enkele) of enige moeite) 1 = Ja (ervaart grote moeite)

* = aangemaakte variabele, indicator

'mbmrk301' is de ruwe variabele. 'LBARK3S1' en 'LBARK3S2' is moeite met rondkomen gehercodeerd in 2 en 3 categorieën. 'LBARK3S3' is wel of geen grote moeite met rondkomen.

4.6.2 Ervaren gezondheid

Overzicht van de variabelen m.b.t. ervaren gezondheid in het totaalbestand

Variabele naam	Label/Toelichting	Codering
EBEGK311	Hoe is over het algemeen je gezondheid?	1 = Zeer goed 2 = Goed 3 = Gaat wel 4 = Slecht 5 = Zeer slecht
EBEGK3S2*	Ervaren gezondheid, in 2 categorieën	1 = (Zeer) goed 2 = Gaat wel of (zeer) slecht
EBEGK3S3*	Ervaren gezondheid, in 3 categorieën	1 = (Zeer) goed 2 = Gaat wel 3 = (Zeer) slecht

* = aangemaakte variabele, indicator

Variabele 'EBEGK311' is de ruwe variabele. 'EBEGK3S2' en 'EBEGK3S3' is ervaren gezondheid gehercodeerd in 2 en 3 categorieën.

4.6.3 Geluk

Overzicht van de variabelen m.b.t. geluk in het totaalbestand

Variabele naam	Label/Toelichting	Codering
EBGLK301	Hoe gelukkig voel je je meestal?	1 = score 1 2 = score 2 3 = score 3 4 = score 4 5 = score 5
EBGLK3S1*	Voelt zich meestal (zeer) gelukkig	0 = Nee, voelt zich (zeer) ongelukkig of neutraal (score 1, 2 of 3) 1 = Ja (score 4 of 5)
EBGLK3S2*	Voelt zich meestal (zeer) ongelukkig	0 = Nee, voelt zich neutraal of (zeer) gelukkig (score 3, 4 of 5) 1 = Ja (score 1 of 2)
EBGLK3S3*	Geluk in 3 categorieën	1 = Voelt zich meestal (zeer) ongelukkig (score 1 of 2) 2 = Voelt zich meestal neutraal (score 3) 3 = Voelt zich meestal (zeer) gelukkig (score 4 of 5)

* = aangemaakte variabele, indicator

Variabele 'EBGLK301' is de ruwe variabele. 'EBGLK3S1' en 'EBGLK3S2' is geluk opgedeeld in gelukkig en ongelukkig. 'EBGLK3S3' is geluk gehercodeerd in 3 categorieën.

4.6.4 Psychosociale gezondheid

Overzicht van de variabelen m.b.t. psychosociale gezondheid in het totaalbestand

Variabele naam	Label/Toelichting	Codering
GGPGK301	Ik probeer aardig te zijn tegen anderen	0 = Niet waar 1 = Beetje waar 2 = Zeker waar
GGPGK302	Ik ben rusteloos, kan niet lang stil zitten	Zie GGPGK301
GGPGK303	Ik heb vaak hoofdpijn, buikpijn of ben misselijk	Zie GGPGK301
GGPGK304	Ik deel makkelijk met anderen	Zie GGPGK301
GGPGK305	Ik word erg boos en ben vaak driftig	Zie GGPGK301
GGPGK306	Ik ben nogal op mijzelf	Zie GGPGK301
GGPGK307	Ik doe meestal wat me wordt opgedragen	Zie GGPGK301
GGPGK308	Ik pieker veel	Zie GGPGK301
GGPGK309	Ik help iemand die zich heeft bezeerd, van streek is of ziek voelt	Zie GGPGK301
GGPGK310	Ik zit constant te wiebelen of friemelen	Zie GGPGK301
GGPGK311	Ik heb minstens één goede vriend(in)	Zie GGPGK301
GGPGK312	Ik vecht vaak	Zie GGPGK301
GGPGK313	Ik ben vaak ongelukkig, in tranen of zit in de put	Zie GGPGK301
GGPGK314	Andere jongeren vinden mij over het algemeen aardig	Zie GGPGK301
GGPGK315	Ik ben snel afgeleid, ik vind het moeilijk me te concentreren	Zie GGPGK301
GGPGK316	Ik ben zenuwachtig in nieuwe situaties	Zie GGPGK301
GGPGK317	Ik ben aardig tegen jongere kinderen	Zie GGPGK301
GGPGK318	Ik word er vaak van beschuldigd dat ik lieg of bedrieg	Zie GGPGK301
GGPGK319	Andere kinderen of jongeren pesten of treiteren mij	Zie GGPGK301
GGPGK320	Ik bied anderen vaak aan hen te helpen	Zie GGPGK301
GGPGK321	Ik denk na voor ik iets doe	Zie GGPGK301
GGPGK322	Ik neem dingen weg die niet van mij zijn	Zie GGPGK301
GGPGK323	Ik kan beter met volwassenen opschieten dan met jongeren	Zie GGPGK301
GGPGK324	Ik ben voor veel dingen bang, ik ben snel angstig	Zie GGPGK301
GGPGK325	Ik maak af waar ik mee bezig ben	Zie GGPGK301
GGPGK3S1*	Emotionele problemen, totaalscore	
GGPGK3S9**	Emotionele problemen in 3 klassen (oude afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend
GGPK3S19**	Emotionele problemen in 3 klassen (nieuwe afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend
GGPGK3S2*	Gedragsproblemen, totaalscore	
GGPK3S10**	Gedragsproblemen in 3 klassen (oude afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend

GGPK3S20**	Gedragsproblemen in 3 klassen (nieuwe afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend
GGPGK3S6*	Totale probleemscore, totaalscore	
GGPGK3S8**	Totale probleemscore in 3 klassen (oude afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend
GGPK3S18**	Totale probleemscore in 3 klassen (nieuwe afkappunten)	1 = Normaal 2 = Grensgebied 3 = Afwijkend

* = aangemaakte variabele, ** = aangemaakte indicator

De SDQ vragenlijst (Goodman, 1997) is afgenomen voor het meten van psychosociale problemen (25 items). Deze vragenlijst bestaat uit verschillende schalen: pro-sociaal gedrag, hyperactiviteit, gedragsproblemen, emotionele problemen, problemen met leeftijdsgenoten.

Variabelen GGPGK301 tot en met GGPGK325 de ruwe variabelen. Variabelen met een sterretje (*) zijn indicatoren van de (sub)schalen van de SDQ, aangemaakt in de bijbehorende syntax.

Er worden 2 soorten afkapwaarden getoond: oude en nieuwe afkapwaarden:

- Oude afkapwaarden gebaseerd op <http://www.sdqinfo.com/ScoreSheets/e2.pdf>.

Afkapwaarden (oude afkappunten):

- Emotionele problemen: 0-5 is normaal / 6 is grensgebied / 7-10 is afwijkend.
- Gedragsproblemen: 0-3 is normaal / 4 is grensgebied / 5-10 is afwijkend.
- Totale probleemscore: 0-15 is normaal / 16-19 is grensgebied / 20-40 is afwijkend.

- De nieuwe afkapwaarden zijn gebaseerd op 80/10/10 regel om hiermee normaal/grensgebied/afwijkend te definiëren. Deze zijn berekend op de GMJ 2015 data:

Afkapwaarden (nieuwe afkappunten):

- Emotionele problemen: 0-3 is normaal / 4 is grensgebied / 5-10 is afwijkend.
- Gedragsproblemen: 0-2 is normaal / 3 is grensgebied / 4-10 is afwijkend.
- Totale probleemscore: 0-13 is normaal / 14-16 is grensgebied / 17-40 is afwijkend.

Voor een subschalen hyperactiviteit, problemen met leeftijdsgenoten en pro-sociaal gedrag zijn geen afkappunt (oud en nieuw) gemaakt i.v.m. onbetrouwbaarheid van de schaal.

4.6.5 Opkomen voor jezelf (weerbaarheid)

Overzicht van de variabelen m.b.t. weerbaarheid in het totaalbestand

Variabele naam	Label/Toelichting	Codering
PBWBK301	Ik weet wat ik wel en niet wil	1 = Helemaal mee eens 2 = Mee eens 3 = Soms wel, soms niet 4 = Mee oneens 5 = Helemaal mee oneens
PBWBK302	Ik zeg het als iemand iets doet wat ik vervelend vind	Zie PBWBK301
PBWBK303	Ik laat mij makkelijk overhalen	Zie PBWBK301
PBWBK304	Ik kan heel goed nee zeggen	Zie PBWBK301
PBWBK305	Ik vind het moeilijk om hulp te vragen	Zie PBWBK301
PBWBK306	Ik doe dingen die ik zelf echt wil	Zie PBWBK301
PBWBK307	Ik kom voor mezelf op	Zie PBWBK301
PBWBK308	Als vrienden iets doen wat ik niet wil doe ik toch mee	Zie PBWBK301
PBWBK3S1*	Weerbaarheid, totaalscore	None
PBWBK3S2*	Weerbaarheid in 2 klassen	0 = Onvoldoende weerbaar 1 = Voldoende weerbaar

* = aangemaakte variabele, indicator

Variabelen 'PBWBK301 tot en met PBWBK308' zijn de ruwe variabelen. 'PBWBK3S1' is weerbaarheid totaalscore en 'PBWBK3S2' weerbaarheid opgedeeld in 2 klassen, aangemaakt in de bijbehorende syntax. Afkapwaarden zijn gebaseerd op syntax GGD Hart voor Brabant 03-03-2012. Afkapwaarden: 1 t/m 3 is onvoldoende weerbaar / > 3 t/m 5 is voldoende weerbaar (vaker dan 'soms wel, soms niet')

4.6.6. Stress

Overzicht van de variabelen m.b.t. stress in het totaalbestand

Variabele naam	Label/Toelichting	Codering
sbssk301	Ik voel me gestrest door school of huiswerk	1= Nooit 2 = Bijna nooit 3 = Soms 4 = Vaak 5 = Zeer vaak
sbssk302	Ik voel me gestrest door mijn situatie thuis	Zie sbssk301
sbssk303	Ik voel me gestrest door eigen problemen	Zie sbssk301
sbssk304	Ik voel me gestrest over wat anderen van me vinden	Zie sbssk301
sbssk305	Ik voel me gestrest door alles wat ik moet doen	Zie sbssk301
SBSSK3S1*	Voelt zich (zeer) vaak gestrest door school of huiswerk	0 = Nee, (bijna) nooit of soms 1 = Ja, (zeer) vaak
SBSSK3S2*	Voelt zich (zeer) vaak gestrest door situatie thuis	Zie SBSSK3S1
SBSSK3S3*	Voelt zich (zeer) vaak gestrest door eigen problemen	Zie SBSSK3S1
SBSSK3S4*	Voelt zich (zeer) vaak gestrest over wat anderen van hem/haar vinden	Zie SBSSK3S1
SBSSK3S5*	Voelt zich (zeer) vaak gestrest door alles wat ze moeten doen	Zie SBSSK3S1
SBSSK3S6*	Voelt zich (zeer) vaak gestrest door één of meer factoren	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

'sbssk301't/m 'sbssk305'zijn de ruwe variabelen. SBSSK3S1 t/m SBSSK3S6 zijn gehercodeerd in 2 categorieën.

4.6.7 Ingrijpende gebeurtenissen

Overzicht van de variabelen m.b.t. ingrijpende gebeurtenissen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
mbigk371	Overlijden van iemand in je gezin	1 = Nooit meegemaakt 2 = Heb ik meegemaakt 3 = Maak ik nu mee
mbigk372	Ernstige lichamelijke ziekte of beperking van iemand in je gezin	Zie mbigk371
mbigk373	Ernstige lichamelijke ziekte of beperking van jezelf	Zie mbigk371
mbigk374	Psychische ziekte van iemand in je gezin	Zie mbigk371
mbigk375	Psychische ziekte van jezelf	Zie mbigk371
mbigk376	Verslaving aan alcohol, drugs of gokken van iemand in je gezin (niet roken)	Zie mbigk371
mbigk377	Scheiding van je ouders	Zie mbigk371
MBIGK3S2*	Heeft nu een gezinslid met een ernstige ziekte, beperking of verslaving (jonge mantelzorger)	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

‘mbigk371’ t/m ‘mbigk377’ zijn de ruwe variabelen. Hierbij lijkt het er op dat bij vraag ‘mbigk371’ over overlijden van iemand in je gezin dit niet altijd goed geïnterpreteerd als gezin maar als familie. Hiervoor moet rekening worden gehouden bij de interpretatie van dit item. ‘MBIGK3S2’ is een indicator die jonge mantelzorger weergeeft.

4.6.8 Bewegen

Overzicht van de variabelen m.b.t. bewegen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
lblbk321	Hoeveel dgn pw lopend of fietsend naar school of stageplek?	1 = (Bijna) nooit 2 = 1 dag per week 3 = 2 dagen per week 4 = 3 dagen per week 5 = 4 dagen per week 6 = 5 dagen per week
lblbk322	Hoeveel dgn pw sport je bij club etc?	1 = (Bijna) nooit 2 = Minder dan 1 dag per week 3 = 1 dag per week 4 = Meerdere dagen per week 5 = Elke dag
lblbk323	Hoeveel dgn pw sport of beweeg je in vrije tijd zonder club etc?	1 = (Bijna) nooit 2 = Minder dan 1 dag per week 3 = 1 dag per week 4 = Meerdere dagen per week 5 = Elke dag
lblbk324	Hoeveel dgn pw sport of beweeg je minstens 1 uur?	1 = (Bijna) nooit 2 = 1 dag per week 3 = 2 dagen per week 4 = 3 dagen per week 5 = 4 dagen per week 6 = 5 dagen per week 7 = 6 dagen per week 8 = Elke dag
LBLBK3S10*	Loopt of fietst 5 dagen per week naar school of stageplek	0 = Nee 1 = Ja
LBLBK3S11*	Sport wekelijks bij club, vereniging of sportschool	0 = Nee 1 = Ja
LBLBK3S12*	Beweegt dagelijks minstens 1 uur	0 = Nee 1 = Ja
LBLBK3S13*	Beweegt 5 of meer dagen per week minstens 1 uur	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

Variabele 'lblbk321 t/m lblbk324' zijn de ruwe variabelen. 'LBLBK3S10' t/m 'LBLBK3S13' zijn de indicatoren voor bewegen.

4.6.9 Voeding (water drinken, suikerhoudende dranken, energiedrankjes)

Overzicht van de variabelen m.b.t. voeding in het totaalbestand

Variabele naam	Label/Toelichting	Codering
lbvok331	Hoeveel dgn pw drink je water (minimaal een glas)?	1 = (Bijna) nooit 2 = 1 dag per week 3 = 2 dagen per week 4 = 3 dagen per week 5 = 4 dagen per week 6 = 5 dagen per week 7 = 6 dagen per week 8 = Elke dag
lbvok332	Hoeveel dgn pw drink je drankjes met suiker?	Zie lbvok331
lbvok333	Als drankjes met suiker, hoeveel glazen drink je dan?	1 = 1 glas per dag, of minder 2 = 2 glazen per dag 3 = 3 glazen per dag 4 = 4 glazen per dag 5 = 5 glazen per dag 6 = 6 glazen per dag 7 = Meer dan 6 glazen per dag
lbvok334	Hoeveel dgn pw drink je energiedrankjes (energy drinks)?	Zie lbvok331
lbvok335	Als energiedrankjes, hoeveel blikjes drink je dan?	1 = 1 blikje per dag 2 = 2 blikjes per dag 3 = 3 blikjes per dag 4 = 4 blikjes per dag 5 = 5 blikjes per dag 6 = 6 blikjes per dag 7 = Meer dan 6 blikjes per dag
LBVOA333	Als drankjes met suiker, hoeveel glazen drink je dan?	Zie lbvok333
LBVOK3S33*	Drinkt (bijna) nooit drankjes met suiker	0 = Nee 1 = Ja
LBVOK3S34*	Drinkt dagelijks drankjes met suiker	0 = Nee 1 = Ja
LBVOK3S35*	Drinkt elke dag meer dan 3 glazen drankjes met suiker	0 = Nee 1 = Ja
LBVOA335	Als energiedrankjes, hoeveel blikjes drink je dan? (opgeschoond)	Zie lbvok335
LBVOK3S36*	Drinkt (bijna) nooit energiedrankjes (is richtlijn voor kinderen tm 13 jr)	0 = Nee 1 = Ja
LBVOK3S37*	Drinkt maximaal 1 blikje energiedrank per dag (is richtlijn voor kinderen va 13 jr)	0 = Nee 1 = Ja
LBVOK3S31*	Drinkt dagelijks minimaal een glas water	0 = Nee 1 = Ja
LBVOK3S32*	Drinkt (bijna) nooit water	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

'lbvok331't/m 'lbvok335' zijn de ruwe variabelen. 'LBVOA333' en 'LBVOA335' zijn opgeschoonde variabelen. De overige zijn aangemaakt variabelen.

4.6.10 Roken actief en e-sigaret gebruik

Overzicht van de variabelen m.b.t. roken actief in het totaalbestand

Variabele naam	Label/Toelichting	Codering
lbrak311	Heb je ooit gerookt?	1 = Ja, een hele sigaret of meer 2 = Ja, alleen een paar trekjes 3 = Nee
lbrak312	Hoe vaak rook je nu?	1 = Elke dag 2 = Minstens 1 keer per week maar niet elke dag 3 = Minder dan 1x per week 4 = Ik rook niet
LBRAA312	Hoe vaak rook je nu? (opgeschoond)	1 = Elke dag 2 = Minstens 1 keer per week maar niet elke dag 3 = Minder dan 1x per week 4 = Ik rook niet
LBRAK3S11*	Ooit gerookt (hele sigaret of meer)	0 = Nee 1 = Ja
LBRAK3S12*	Ooit gerookt (alleen trekjes of hele sigaret of meer)	0 = Nee 1 = Ja
LBRAK3S18*	Rookt nu dagelijks	0 = Nee 1 = Ja
LBRAK3S19*	Rookt nu \geq wekelijks	0 = Nee 1 = Ja
lbgek301	Heb je ooit een e-sigaret gebruikt?	1 = Ja, vaker dan 1 keer 2 = Ja, 1 keer 3 = Nee, nooit 4 = Ik weet niet wat dat is
lbgek302	Hoe vaak gebruik je nu een e-sigaret?	1 = Elke dag 2 = Minstens 1 keer per week, maar niet elke dag 3 = Minder dan 1 keer per week 4 = Ik gebruik geen elektronische sigaret
LBGEA302	Hoe vaak gebruik je nu een e-sigaret? (opgeschoond)	1 = Elke dag 2 = Minstens 1 keer per week, maar niet elke dag 3 = Minder dan 1 keer per week 4 = Ik gebruik geen elektronische sigaret
LBGEK3S1*	Ooit een e-sigaret gebruikt	0 = Nee

LBGEK3S2*	Gebruikt nu dagelijks een e-sigaret	1 = Ja 0 = Nee
LBGEK3S3*	Gebruikt nu \geq wekelijks een e-sigaret	1 = Ja 0 = Nee
		1 = Ja

* = aangemaakte variabele, indicator

Variabelen 'lbrak311, lbrak312' 'lbgek301' en 'lbgek302' zijn de ruwe variabelen. LBRAA312 is combi 'LBRAK311' en 'LBRAK312' (niets ingevuld op LBRAK312 dan missing bij LBRAA312; niets ingevuld op LBRAK311 \rightarrow dan missing bij LBRAA312; nee op LBRAK311 dan LBRAA312 \rightarrow nvt). Deze manier van opschonen geldt ook voor de e-sigaret: LBGEA302. Variabelen met een sterretje (*) zijn indicatoren, aangemaakt in de bijbehorende syntax.

4.6.11 Alcoholgebruik/ Alcoholgebruik binge en aangeschoten

Overzicht van de variabelen m.b.t. alcoholgebruik in het totaalbestand

Variabele naam	Label/Toelichting	Codering
lbagk391	Heb je ooit alcohol gedronken?	1 = Ja, een heel glas of meer 2 = Ja, alleen een paar slokjes 3 = Nee
lbagk398	Hoeveel dgn alcohol in hele leven?	1= Nooit 2 = 1 of 2 dagen 3 = 3 t/m 5 dagen 4 = 6 t/m 9 dagen 5 = 10 dagen of meer
lbagk399	Hoeveel dgn alcohol in laatste 4 wkn?	1= Nooit 2 = 1 of 2 dagen 3 = 3 t/m 5 dagen 4 = 6 t/m 9 dagen 5 = 10 dagen of meer
lbagk307	Hoe vaak laatste 4 wkn binge drinken?	1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 of 4 keer 5 = 5 of 6 keer 6 = 7 of 8 keer 7 = 9 keer of vaker
lbagk394	Hoe vaak dronken of aangeschoten in hele leven?	1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 keer 5 = 4 t/m 10 keer 6 = 11 keer of vaker
Lbagk395	Hoe vaak dronken of aangeschoten in laatste 4 wkn?	1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 keer 5 = 4 t/m 10 keer 6 = 11 keer of vaker
LBAGK3S11*	Ooit alcohol gedronken (heel glas of meer)	0 = Nee 1 = Ja
LBAGK3S12*	Ooit alcohol gedronken (alleen slokjes of heel glas of meer)	0 = Nee 1 = Ja
LBAGK3S20*	Laatste 4 weken alcohol gedronken	0 = Nee 1 = Ja
LBAGA307	Hoe vaak laatste 4 wkn binge drinken? (opgeschoond)	1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 of 4 keer 5 = 5 of 6 keer 6 = 7 of 8 keer 7 = 9 keer of vaker

LBAGA394	Hoe vaak dronken of aangeschoten in hele leven? (opgeschoond)	8 = nvt 1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 keer 5 = 4 t/m 10 keer 6 = 11 keer of vaker 8 = nvt
LBAGA395	Hoe vaak dronken of aangeschoten in laatste 4 wkn? (opgeschoond)	1 = Nooit 2 = 1 keer 3 = 2 keer 4 = 3 keer 5 = 4 t/m 10 keer 6 = 11 keer of vaker 8 = nvt
LBAGK3S15*	Laatste 4 weken 5 of meer drankjes bij 1 gelegenheid (binge drinken)	0 = Nee 1 = Ja
LBAGK3S16*	Ooit dronken of aangeschoten geweest	0 = Nee 1 = Ja
LBAGK3S17*	Laatste 4 weken dronken of aangeschoten geweest	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

Variabelen 'lbagk391', 'lbagk398', 'lbagk399', 'lbagk307', lbagk394, lbagk395 zijn de ruwe variabelen. LBAGA307, LBAGA394, LBAGA395 zijn opgeschoonde variabelen aan de hand van de voorloper vraag. De overige variabelen zijn indicatoren.

4.6.12 Gebruik wiet of hasj

Overzicht van de variabelen m.b.t. gebruik wiet of hasj in het totaalbestand

Variabele naam	Label/Toelichting	Codering
LBSDK301	Heb je ooit wiet of hasj gebruikt?	1 = Ja 2 = Nee
LBSDK308	Hoeveel dgn wiet of hasj in hele leven?	1 = Nooit 2 = 1 of 2 dagen 3 = 3 t/m 5 dagen 4 = 6 t/m 9 dagen 5 = 10 dagen of meer
LBSDK309	Hoeveel dgn wiet of hasj in laatste 4 wkn?	Zie LBSDK308
LBSDA308	Hoeveel dgn wiet of hasj in hele leven? (opgeschoond)	1 = Nooit 2 = 1 of 2 dagen 3 = 3 t/m 5 dagen 4 = 6 t/m 9 dagen 5 = 10 dagen of meer
LBSDA309	Hoeveel dgn wiet of hasj in laatste 4 wkn? (opgeschoond)	Zie LBSDA308
LBSDK3S1*	Ooit wiet (marihuana) of hasj gebruikt	0 = Nee 1 = Ja
LBSDK3S2*	Laatste 4 wkn wiet (marihuana) of hasj gebruikt	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

‘LBSDK301’ ‘LBSDK308’ ‘LBSDK309’ zijn de ruwe variabelen. ‘LBSDA308’ en ‘LBSDA309’ zijn de opgeschoonde vraag. Overige zijn de indicatoren.

4.6.13 Pesten op school

Overzicht van de variabelen m.b.t. pesten op school in het totaalbestand

Variabele naam	Label/Toelichting	Codering
SBRLK301	Hoe vaak laatste 3 mndn gepest op school?	1 = Nooit 2 = Minder dan 2 keer per week 3 = 2 of 3 keer per maand 4 = Ongeveer 1 keer per week 5 = Meerdere keren per week
SBRLK302	Hoe vaak laatste 3 mndn zelf ander gepest op school?	Zie SBRLK301
SBRLK3S11*	Laatste 3 mndn gepest op school	0 = Nee 1 = Ja
SBRLK3S12*	Laatste 3 mndn meerdere keren per mnd gepest op school	0 = Nee 1 = Ja
SBRLK3S13*	Laatste 3 mndn \geq wekelijks gepest op school	0 = Nee 1 = Ja
SBRLK3S14*	Laatste 3 mndn zelf ander gepest op school	0 = Nee 1 = Ja
SBRLK3S15*	Laatste 3 mndn meerdere keren per mnd zelf ander gepest op school	0 = Nee 1 = Ja
SBRLK3S16*	Laatste 3 mndn \geq wekelijks zelf ander gepest op school	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

‘SBRLK301’ en ‘SBRLK302’ zijn de ruwe variabelen. ‘SBRLK3S11’ t/m ‘SBRLK3S16’ zijn de indicatoren.

4.6.14 Cyberpesten

Overzicht van de variabelen m.b.t. cyberpesten in het totaalbestand

Variabele naam	Label/Toelichting	Codering
sbrlk311	Hoe vaak laatste 3 mndn gepest via internet?	1 = Nooit 2 = Minder dan 2 keer per week 3 = 2 of 3 keer per maand 4 = Ongeveer 1 keer per week 5 = Meerdere keren per week
sbrlk312	Hoe vaak laatste 3 mndn zelf ander gepest via internet?	Zie sbrlk311
SBRLK3S21*	Laatste 3 mndn gepest via internet	0 = Nee 1 = Ja
SBRLK3S22*	Laatste 3 mndn meerdere keren per mnd gepest via internet	0 = Nee 1 = Ja
SBRLK3S23*	Laatste 3 mndn \geq wekelijks gepest via internet	0 = Nee 1 = Ja
SBRLK3S24*	Laatste 3 mndn zelf ander gepest via internet	0 = Nee 1 = Ja
SBRLK3S25*	Laatste 3 mndn meerdere keren per mnd zelf ander gepest via internet	0 = Nee 1 = Ja
SBRLK3S26*	Laatste 3 mndn \geq wekelijks zelf ander gepest via internet	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

‘sbrlk311’ en ‘sbrlk312’ zijn de ruwe variabelen. ‘SBRLK3S21’ t/m ‘SBRLK3S26’ zijn de opgeschoonde variabelen.

4.6.15 Sexting

Overzicht van de variabelen m.b.t. sexting in het totaalbestand

Variabele naam	Label/Toelichting	Codering
sbsxk301	Hoe vaak laatste 6 mndn naaktfoto of seksfilmpje van jezelf gestuurd?	1 = Nooit 2 = 1 keer 3 = 2 keer of vaker
SBSXK3S1*	Laatste 6 mndn naaktfoto of seksfilmpje van zichzelf gestuurd	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

'sbsxk301' is de ruwe variabele. 'SBSXK3S1' is de aangemaakte indicator.

4.6.16 Sociale mediagebruik

Overzicht van de variabelen m.b.t. sociale mediagebruik in het totaalbestand

Variabele naam	Label/Toelichting	Codering
sbsgk302	Hoe vaak moeilijk om met sociale media te stoppen?	1 = Nooit 2 = Bijna nooit 3 = Soms 4 = Vaak 5 = Zeer vaak
sbsgk303	Hoe vaak zeggen anderen minder tijd aan sociale media?	Zie sbsgk302
sbsgk304	Hoe vaak liever sociale media dan dat tijd met anderen?	Zie sbsgk302
sbsgk305	Hoe vaak onrustig etc wanneer geen sociale media?	Zie sbsgk302
sbsgk306	Hoe vaak raffel je je huiswerk af om sociale media?	Zie sbsgk302
sbsgk307	Hoe vaak sociale media omdat je je rot voelt?	Zie sbsgk302
sbsgk308	Hoe vaak slaap te kort door sociale media?	Zie sbsgk302
SBSGA302	Hoe vaak moeilijk om met sociale media te stoppen?	0 = Nooit 1 = Zelden 2 = Soms 3 = Vaak 4 = Zeer vaak
SBSGA303	Hoe vaak zeggen anderen minder tijd aan sociale media?	Zie SBSGA302
SBSGA304	Hoe vaak liever sociale media dan tijd met anderen?	Zie SBSGA302
SBSGA305	Hoe vaak onrustig etc wanneer geen sociale media?	Zie SBSGA302
SBSGA306	Hoe vaak raffel je je huiswerk af om sociale media?	Zie SBSGA302
SBSGA307	Hoe vaak sociale media omdat je je rot voelt?	Zie SBSGA302
SBSGA308	Hoe vaak slaap te kort door sociale media?	Zie SBSGA302
SBSGK3S1*	Ervaart zelf risico op problematisch gebruik social media (obv 7 items), totaalscore	
SBSGK3S2*	Risico op problematisch gebruik sociale media (obv 7 items, verzaard) in 2 klassen	0 = Geen of weinig risico 1 = Wel risico (vaker dan soms)
SBSGK3S3*	Risico op problematisch gebruik sociale media (obv 6 items; zoals IVO), totaalscore	
SBSGK3S4*	Risico op problematisch gebruik sociale media (obv 6 items; zoals IVO) in 2 klassen	0 = Geen of weinig risico 1 = Wel risico (vaker dan soms)

* = aangemaakte variabele, indicator

Variabelen met een sterretje (*) zijn indicatoren, aangemaakt in de bijbehorende syntax. Om sociale media te meten is gebruik gemaakt van de verkorte versie van de Compulsive Internet Use Scale van het IVO (7 items: 'SBSGK302 t/m SBSGK308'; hercodering: 'SBSGA302' t/m 'SBSGA308').

De indicatoren geven risico op problematisch gebruik sociale media weer:

'SBSGK3S1' is de totale score van risico op problematisch gebruik sociale media gebaseerd op 7 items. Officieel is de verkorte lijst gebaseerd op 6 items die 6 componenten meet (respectievelijk controleverlies, sociaal conflict, preoccupatie, terugval, conflict en coping). Hier is slaap item aan toegevoegd. De verkorte lijst is niet gevalideerd, maar al wel in verschillende onderzoeken meegenomen.

'SBSGK3S2' is gebaseerd op 7 items: SBSGA302' t/m 'SBSGA307'. Het slaap-item is verzwaard.

'Verzwaren' gebeurt door te stellen dat er positief (vaker dan 'soms') gescoord moet worden op minimaal één van de volgende items: SBSGK306 (huiswerk-item) en SBSGK308 (slaap-item).

'SBSGK3S2' heeft de volgende afkappunten voor wel of geen risico op problematisch sociale media gebruik: 0 t/m 2 is geen of weinig risico / > 2 t/m 4 is wel risico (vaker dan 'soms') (Bron: afkomstig van het IVO).

'SBSGK3S' is gebaseerd op de verkorte Compulsive Internet Use Scale van het IVO (6 items: SBSGA302' t/m 'SBSGA307').

'SBSGK3S4' is risico op problematisch gebruik sociale media net als SBSGK3S3 maar ingedeeld in 2 klassen met de volgende afkappunten: 0 t/m 2 is geen of weinig risico / > 2 t/m 4 is wel risico (vaker dan 'soms') (Bron: afkomstig van het IVO).

4.6.17 Gamen

Overzicht van de variabelen m.b.t. gamen in het totaalbestand

Variabele naam	Label/Toelichting	Codering
SBGMK301	Hoe vaak speel je games?	1 = (Bijna) nooit 2 = Minder dan 1 dag per week 3 = 1 dag per week 4 = 2 of 3 dagen per week 5 = 4 of 5 dagen per week 6 = (Bijna) elke dag
SBGMK302	Hoe vaak moeilijk om met gamen te stoppen?	1 = Nooit 2 = Bijna nooit 3 = Soms 4 = Vaak 5 = Zeer vaak
SBGMK303	Hoe vaak zeggen anderen minder tijd aan gamen?	Zie SBGMK302
SBGMK304	Hoe vaak liever gamen dan tijd met anderen?	Zie SBGMK302
SBGMK305	Hoe vaak onrustig etc wanneer niet gamen?	Zie SBGMK302
SBGMK306	Hoe vaak raffel je je huiswerk af om te gamen?	Zie SBGMK302
SBGMK307	Hoe vaak gamen omdat je je rot voelt?	Zie SBGMK302
SBGMK308	Hoe vaak slaap te kort door gamen?	Zie SBGMK302
SBGMA302	Hoe vaak moeilijk om met gamen te stoppen? (opgeschoond)	0 = Nooit 1 = Bijna nooit 2 = Soms 3 = Vaak 4 = Zeer vaak
SBGMA303	Hoe vaak zeggen anderen minder tijd aan gamen? (opgeschoond)	Zie SBGMA302
SBGMA304	Hoe vaak liever gamen dan tijd met anderen? (opgeschoond)	Zie SBGMA302
SBGMA305	Hoe vaak onrustig etc wanneer niet gamen? (opgeschoond)	Zie SBGMA302
SBGMA306	Hoe vaak raffel je je huiswerk af om te gamen? (opgeschoond)	Zie SBGMA302
SBGMA307	Hoe vaak gamen omdat je je rot voelt? (opgeschoond)	Zie SBGMA302
SBGMA308	Hoe vaak slaap te kort door gamen? (opgeschoond)	Zie SBGMA302
SBGMK3S1*	Ervaart zelf risico op problematisch gamen (obv 7 items), totaalscore	
SBGMK3S2*	Ervaart zelf risico op problematisch gamen (obv 7 items, verzwaard) in 2 klassen	0 = geen of weinig risico 1 = wel risico (vaker dan soms)
SBGMK3S3*	Ervaart zelf risico op problematisch gamen (obv 6 items; zoals IVO), totaalscore	
SBGMK3S4*	Ervaart zelf risico op problematisch gamen (obv 6 items; zoals IVO) in 2 klassen	0 = geen of weinig risico

* = aangemaakte variabele, indicator

Variabelen met een sterretje (*) zijn indicatoren, aangemaakt in de bijbehorende syntax.

Om problematisch gamen te meten is gebruik gemaakt van de Video game Addiction Test (VAT) van het IVO (7 items: 'SBGMK302' t/m 'SBGMK308'; hercodering: 'SBGMA302' t/m 'SBGMK308'). Dit is dezelfde lijst als de CIUS, maar dan toegepast op gamen in plaats van internetten. De verkorte lijst bestaat uit 6 items. De keuze voor de 6 items is gemaakt op basis van de hoogste factorlading binnen 6 onderscheiden componenten (respectievelijk controleverlies, sociaal conflict, preoccupatie, terugval, conflict en coping) op basis van historische data van het IVO. De subwerkgroep, bestaande uit vier GGD-epidemiologen, Trimbos Instituut en het IVO, heeft een 7^e item uit de oorspronkelijke lange lijst toe gevoegd aan de verkorte lijst: een item over slaaptkort als gevolg van gamen.

Indicator 'SBGMK3S1' geeft de totaal score van risico op problematisch gamen op basis van 7 items van de 'IVO schaal' ('SBGMA302' t/m 'SBGMA308').

Indicator 'SBGMK3S2' is ook op basis van 7 items ('SBGMA302' t/m 'SBGMA308') maar hierbij is het slaap item verzwaaard. Deze indicator wordt 'verzwaaard' door te stellen dat er positief (vaker dan 'soms') gescoord moet worden op minimaal één van de volgende items: SBGMK306 (huiswerk-item) en SBGMK308 (slaap-item). Afkappunten van 'SBGMK3S2': 0 t/m 2 is geen of weinig risico / > 2 t/m 4 is wel risico (vaker dan 'soms') (afkomstig van het IVO).

Indicator 'SBGMK3S3' en 'SBGMK3S4' zijn op basis van 6 items, de verkorte IVO-schaal ('SBGMA302' t/m 'SBGMA307'). De volgende afkappunten zijn gebruikt voor 'SBGMK3S4': 0 t/m 2 is geen of weinig risico / > 2 t/m 4 is wel risico (vaker dan 'soms') (afkomstig van het IVO).

'SBSGK3S1' is de totale score van risico op problematisch gebruik sociale media gebaseerd op 7 items.

Bron originele vragenlijst: Van Rooij, A. J., Schoenmakers, T. M., Van den Eijnden, R. J. J. M., Vermulst, A. a, & Van de Mheen, D. (2012). Video game addiction test: validity and psychometric characteristics. *Cyberpsychology, behavior and social networking*, 15(9), 507–11.
doi:10.1089/cyber.2012.0007

4.6.18 Seksualiteit

Overzicht van de variabelen m.b.t seksualiteit in het totaalbestand

Variabele naam	Label/Toelichting	Codering
lbskk311	Heb je ooit met iemand gezoend (tongzoenen)?	1 = Ja 2 = Nee, nooit
lbskk315	Heb je ooit seksuele gemeenschap gehad?	1 = Ja 2 = Nee, nooit
lbskk313	Heb je altijd een condoom gebruikt?	1 = Ja, altijd 2 = Nee, niet altijd 3 = Nee, nooit 4 = Ik weet het niet meer
LBSKA315	Heb je ooit seksuele gemeenschap gehad? (opgeschoond)	1 = Ja 2 = Nee, nooit
LBSKA313	Heb je altijd een condoom gebruikt? (opgeschoond)	1 = Ja, altijd 2 = Nee, niet altijd 3 = Nee, nooit 4 = Ik weet het niet meer
LBSKK3S11*	Ooit seksuele gemeenschap gehad	0 = Nee 1 = Ja
LBSKK3S12*	Altijd condooms gebruikt	0 = Nee 1 = Ja

* = aangemaakte variabele, indicator

‘lbskk311’, ‘lbskk315’, ‘lbskk313’ zijn de ruwe variabelen. ‘LBSKA315’, ‘LBSKA313’ zijn de opgeschoonde variabelen. ‘LBSKK3S11’ ‘LBSKK3S12’ zijn de indicatoren.

5. Complex Samples in SPSS

Voor het berekenen van gewogen prevalentiecijfers kan de procedure 'weight cases' in SPSS worden gebruikt.

Voor het toetsen van (statistisch significante) verschillen moeten gewogen cijfers met elkaar worden vergeleken.

Een manier hiervoor is het vergelijken van de betrouwbaarheidsintervallen rondom de gewogen schattingen. Bij het bepalen van betrouwbaarheidsintervallen rondom gewogen prevalentiecijfers moet echter rekening worden gehouden met een grotere onbetrouwbaarheid als gevolg van het wegen. Hiervoor kan de module 'Complex Samples' van SPSS worden gebruikt. In deze module wordt rekening gehouden met het zogenaamde designeffect.

In de SPSS procedure 'Complex Samples' wordt rekening gehouden met het steekproefdesign. Hiervoor wordt de variabele 'stratum' in het totaalbestand gebruikt die de steekproef stratificatie weergeeft. Er is voor de GGD'en die mee doen op regionaal niveau per GGD een steekproef van scholen getrokken. GGD-regio is de eerste variabele waarop gestratificeerd wordt. Vervolgens zijn er binnen die scholen klassen (=leerlingen) getrokken/aangewezen. Van elk schoolniveau en leerjaar. Voor de regio's waar alle scholen zijn benaderd is er maar 1 stratum, namelijk de GGD-regio (zie paragraaf 4.3 weegfactoren). Dit betekent de volgende indeling voor de strata:

* GGD Groningen.

IF (GGDregio_school=1) stratum=1.

* GGD Fryslan.

IF (GGDregio_school=2 AND KLAS=2 AND MBOKA3S33=1) stratum=9.

IF (GGDregio_school=2 AND KLAS=4 AND MBOKA3S33=1) stratum=10.

IF (GGDregio_school=2 AND KLAS=2 AND MBOKA3S33=2) stratum=11.

IF (GGDregio_school=2 AND KLAS=4 AND MBOKA3S33=2) stratum=12.

IF (GGDregio_school=2 AND KLAS=2 AND MBOKA3S33=3) stratum=13.

IF (GGDregio_school=2 AND KLAS=4 AND MBOKA3S33=3) stratum=14.

IF (GGDregio_school=2 AND KLAS=2 AND MBOKA3S33=4) stratum=15.

IF (GGDregio_school=2 AND KLAS=4 AND MBOKA3S33=4) stratum=16.

* GGD Drenthe.

IF (GGDregio_school=3) stratum=17.

* GGD IJsselland.

IF (GGDregio_school=4) stratum=25.

* GGD Twente.

IF (GGDregio_school=5) stratum = 33.

* GGD Noord- en Oost-Gelderland.

IF (GGDregio_school=6) stratum=41.

* VG Gelderland-Midden.

IF (GGDregio_school=7 AND KLAS=2 AND MBOKA3S33=1) stratum=49.

IF (GGDregio_school=7 AND KLAS=4 AND MBOKA3S33=1) stratum=50.

IF (GGDregio_school=7 AND KLAS=2 AND MBOKA3S33=2) stratum=51.

IF (GGDregio_school=7 AND KLAS=4 AND MBOKA3S33=2) stratum=52.

IF (GGDregio_school=7 AND KLAS=2 AND MBOKA3S33=3) stratum=53.
IF (GGDregio_school=7 AND KLAS=4 AND MBOKA3S33=3) stratum=54.
IF (GGDregio_school=7 AND KLAS=2 AND MBOKA3S33=4) stratum=55.
IF (GGDregio_school=7 AND KLAS=4 AND MBOKA3S33=4) stratum=56.

*GGD Gelderland zuid.
IF (GGDregio_school=8) stratum=57.

* GGD regio Utrecht.
IF (GGDregio_school=9) stratum=65.

* GGD Hollands Noorden.
IF (GGDregio_school=10) stratum=73.

* GGD Zaanstreek-Waterland.
IF (GGDregio_school=11) stratum=81.

* GGD Kennemerland.
IF (GGDregio_school=12) stratum=89.

* GGD Amsterdam.
IF (GGDregio_school=13 AND KLAS=2 AND MBOKA3S33=1) stratum=97.
IF (GGDregio_school=13 AND KLAS=4 AND MBOKA3S33=1) stratum=98.
IF (GGDregio_school=13 AND KLAS=2 AND MBOKA3S33=2) stratum=99.
IF (GGDregio_school=13 AND KLAS=4 AND MBOKA3S33=2) stratum=100.
IF (GGDregio_school=13 AND KLAS=2 AND MBOKA3S33=3) stratum=101.
IF (GGDregio_school=13 AND KLAS=4 AND MBOKA3S33=3) stratum=102.
IF (GGDregio_school=13 AND KLAS=2 AND MBOKA3S33=4) stratum=103.
IF (GGDregio_school=13 AND KLAS=4 AND MBOKA3S33=4) stratum=104.

* GGD Gooi en Vechtstreek.
IF (GGDregio_school=14) stratum=105.

* GGD Haaglanden.
IF (GGDregio_school=15 AND KLAS=2 AND MBOKA3S33=1) stratum=113.
IF (GGDregio_school=15 AND KLAS=4 AND MBOKA3S33=1) stratum=114.
IF (GGDregio_school=15 AND KLAS=2 AND MBOKA3S33=2) stratum=115.
IF (GGDregio_school=15 AND KLAS=4 AND MBOKA3S33=2) stratum=116.
IF (GGDregio_school=15 AND KLAS=2 AND MBOKA3S33=3) stratum=117.
IF (GGDregio_school=15 AND KLAS=4 AND MBOKA3S33=3) stratum=118.
IF (GGDregio_school=15 AND KLAS=2 AND MBOKA3S33=4) stratum=119.
IF (GGDregio_school=15 AND KLAS=4 AND MBOKA3S33=4) stratum=120.

* GGD Hollands Midden.
IF (GGDregio_school=16) stratum=121.

* GGD Rotterdam-Rijnmond.
IF (GGDregio_school=17) stratum=129.

* DGJ Zuid-Holland Zuid.
IF (GGDregio_school=18) stratum=137.

* GGD Zeeland.

IF (GGDregio_school=19) stratum=145.

* GGD West-Brabant.

IF (GGDregio_school=20) stratum=153.

* GGD Hart voor Brabant.

IF (GGDregio_school=21 AND KLAS=2 AND MBOKA3S33=1) stratum=161.

IF (GGDregio_school=21 AND KLAS=4 AND MBOKA3S33=1) stratum=162.

IF (GGDregio_school=21 AND KLAS=2 AND MBOKA3S33=2) stratum=163.

IF (GGDregio_school=21 AND KLAS=4 AND MBOKA3S33=2) stratum=164.

IF (GGDregio_school=21 AND KLAS=2 AND MBOKA3S33=3) stratum=165.

IF (GGDregio_school=21 AND KLAS=4 AND MBOKA3S33=3) stratum=166.

IF (GGDregio_school=21 AND KLAS=2 AND MBOKA3S33=4) stratum=167.

IF (GGDregio_school=21 AND KLAS=4 AND MBOKA3S33=4) stratum=168.

* GGD Brabant-Zuidoost.

IF (GGDregio_school=22 AND KLAS=2 AND MBOKA3S33=1) stratum=169.

IF (GGDregio_school=22 AND KLAS=4 AND MBOKA3S33=1) stratum=170.

IF (GGDregio_school=22 AND KLAS=2 AND MBOKA3S33=2) stratum=171.

IF (GGDregio_school=22 AND KLAS=4 AND MBOKA3S33=2) stratum=172.

IF (GGDregio_school=22 AND KLAS=2 AND MBOKA3S33=3) stratum=173.

IF (GGDregio_school=22 AND KLAS=4 AND MBOKA3S33=3) stratum=174.

IF (GGDregio_school=22 AND KLAS=2 AND MBOKA3S33=4) stratum=175.

IF (GGDregio_school=22 AND KLAS=4 AND MBOKA3S33=4) stratum=176.

* GGD Limburg-Noord.

IF (GGDregio_school=23) stratum=177.

* GGD Zuid-Limburg.

IF (GGDregio_school=24) stratum=185.

* GGD Flevoland.

IF (GGDregio_school=25) stratum=193.

Planfile en frequentietabel

Hieronder staan de syntax waarmee je in SPSS een planfile kunt aanmaken. Deze planfile is nodig om de module Complex Samples te gebruiken. Hiermee kun je onder meer betrouwbaarheidsintervallen berekenen. In de syntax staat een voorbeeld van een frequentietabel met betrouwbaarheidsintervallen in SPSS Complex Samples.

Voor elke weegfactor moet een aparte planfile worden gemaakt. Dus een planfile voor landelijke en regionale cijfers. En een aparte voor gemeentelijke weegfactoren.

*aanmaken planfile:

```
CSPLAN ANALYSIS  
/PLAN FILE='C:\Monitor\cs_plan_monitor_weging.csplan'  
/PLANVARS ANALYSISWEIGHT=variabele weegfactor  
/SRSESTIMATOR TYPE=WOR  
/PRINT PLAN  
/DESIGN STRATA=stratum
```

```
/ESTIMATOR TYPE=WR.  
CSTABULATE
```

* frequentietabel in complex samples.

```
/PLAN FILE='C:\ Monitor\ cs_plan_monitor_weging.csaplan'  
/TABLES VARIABLES= XXXXXX  
/CELLS TABLEPCT  
/STATISTICS CIN(95)  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

* kruistabel in complex samples.

```
CSTABULATE  
/PLAN FILE='C:\ Monitor\ cs_plan_monitor_weging.csaplan'  
/TABLES VARIABLES= VARXXX1 VARXXX2 etc  
/CELLS TABLEPCT  
/SUBPOP TABLE=VARXXX3 DISPLAY=LAYERED  
/STATISTICS CIN(95)  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Bijlage 1. Overzicht steekproefniveaus GGD-vragenlijsten per GGD-regio

Gezondheidsmonitor Jeugd 2019 per GGD-regio

Wijze van deelname

- Regionale cijfers
- Gemeentecijfers



Steekproefniveau kan op verschillende niveaus: 1) GGD vraagt alle scholen om mee te doen. Wat zorgt voor gemeentelijk cijfers (N=19). 2) Regionale steekproef levert alleen informatie voor regionale cijfers op GGD regio niveau. Alle GGD'en leveren landelijke referentie cijfers (N=25)

Bijlage 2. Responsaantallen per GGD-regio

GGD-regio	Leerlingen*	Schoollocaties**
GGD Groningen	5161	36
GGD Fryslân	1401	11
GGD Drenthe	5347	21
GGD IJsselland	7282	28
GGD Twente	10748	42
GGD Noord- en Oost- Gelderland	13951	55
VG Gelderland-Midden	2395	13
GGD Gelderland-Zuid	9257	38
GGD regio Utrecht	16426	62
GGD Hollands Noorden	10074	39
GGD Zaanstreek- Waterland	5410	19
GGD Kennemerland	7430	23
GGD Amsterdam	1129	9
GGD Gooi en Vechtstreek	4351	19
GGD Haaglanden	1469	13
GGD Hollands Midden	11663	46
GGD Rotterdam- Rijnmond	11912	61
DGJ Zuid-Holland Zuid	5154	18
GGD Zeeland	6763	25
GGD West-Brabant	10770	41
GGD Hart voor Brabant	1282	8
GGD Brabant-Zuidoost	703	9
GGD Limburg-Noord	8000	22 [#]
GGD Zuid-Limburg	7282	25
GGD Flevoland	5832	24
TOTAAL	171.192	707

*Gebaseerd op variabele regio waar leerling woont (variabele naam: GGDregio_leerling)

**gebaseerd op variabele met unieke schoollocaties (variabele naam: Schoolcode_uniek).

In het databestand zie je 23 school locaties voor GGD Limburg-Noord. Echter 2 schoollocaties zijn dezelfde school. Het zijn dan in totaal 22 schoollocaties

Totale respons:

Het totaal aantal schoollocaties in Nederland was 1238. Totaal aantal benaderde scholen was 910. Hiervan hebben 707 scholen deelgenomen aan de Gezondheidsmonitor Jeugd. De respons bedroeg hiermee 78% en de dekkingsgraad 57%. Van de benaderde scholen deed 22% niet mee om verschillende redenen (te druk/overbelasting vanwege JGZ-activiteiten, te druk/overbelasting vanwege andere vragenlijsten/onderzoeken, te druk vanwege andere zaken, principiële bezwaren, andere redenen, nl... en reden onbekend).

Bijlage 3. Notitie Weging Gezondheidsmonitor Jeugd 2019

December 2020 versie 2.0 met extremen correctie
Rudolf Hoogenveen, Marieke Hiemstra, Bianca Snijders

1. Inleiding en achtergrond

De Gezondheidsmonitor Jeugd (GMJ) is voor de tweede keer uitgevoerd in 2019 (schooljaar 2019-2020) in de periode september 2019- december 2019. In alle 25 GGD'en (dwz 25 GGD regio's + Utrecht Stad) van Nederland is in die periode een basisvragenlijst over leefstijl, gezondheid en gedrag afgenomen bij leerlingen van klas 2 en klas 4 van het reguliere Voortgezet Onderwijs. Alle GGD'en namen een identieke set van basisvragen af, eventueel aangevuld met 'eigen', lokale vragen.

De opbrengst van de Gezondheidsmonitor Jeugd 2019:

- Landelijke prevalentiecijfers over leefstijl, gezondheid en gedrag van leerlingen in 2^e en 4^e klassen van het regulier voortgezet onderwijs (gebaseerd op alle 25 GGD'en).
- Regionale prevalentiecijfers voor alle 25 GGD'en.
- Gemeentelijke prevalentiecijfers in 19 van de 25 GGD'en.
- Cijfers ten behoeve van schoolprofielen voor alle scholen die hebben meegedaan.

In deze notitie wordt de methode en verantwoording van de wegingsprocedure beschreven. De weegfactoren zijn berekend op het RIVM door de afdeling Statistiek Informatica en Modelling (SIM) met het statistisch pakket R (versie 3.6.0) en het 'survey' package. De betreffende code en de documentatie over de weging zijn opvraagbaar.

Waarom is weging nodig?

De waarde die we kunnen hechten aan deze prevalentiecijfers staat of valt met de mate waarin ze beschouwd kunnen worden als een representatieve weergave van de werkelijkheid. Weging van de cijfers is een middel om de prevalenties zo representatief mogelijk te laten zijn voor *alle* tweede en vierdeklassers van het voortgezet onderwijs (VO) in de Nederlandse populatie. In het algemeen geldt dat de steekproef niet exact de verhoudingen in de populatie zal weerspiegelen. Dit is op te lossen door gebruik te maken van weegfactoren, die gebaseerd zijn op de onderzochte kenmerken.

Door het gekozen steekproefdesign van de GMJ 2019 zijn er drie mogelijke weegfactoren:

- 1) Weegfactor = 1: dit betekent dat de personen met deze kenmerken in dezelfde verhouding in de steekproef voorkomen als in de populatie.
- 2) Weegfactor > 1, bijv. 1.5. Dan telt deze jongere dus voor 1.5 jongeren.
- 3) Weegfactor < 1, dit betekent dat de personen met deze kenmerken verhoudingsgewijs 'te veel' voorkomen in de steekproef. Dit laatste is een gevolg van de verschillen tussen de trekkingskansen van leerlingen tussen GGD regio's (de rol van de zogenaamde grensoverschrijders) en wordt later uitgelegd.

De representativiteit van prevalentiecijfers kan op twee manieren geschaad worden:

- 1) Ongelijke trekkingskansen voor leerlingen door het design van de dataverzameling. Een leerling die woont in een GGD regio die op regionaal niveau meedoet heeft minder kans om geïnccludeerd te worden in de steekproef omdat dit afhangt van de kans dat de school geïnccludeerd wordt. Deze vertekening door verschillende trekkingskansen is vooraf te bepalen (o.a. aan de hand van het design van de steekproeftrekking).

- 2) Selectieve sampling waardoor de steekproef afwijkt van de totale populatie. Bijvoorbeeld, stel je trekt alleen leerlingen met een hoge urbanisatiegraad dan mis je dus leerlingen met een lage urbanisatiegraad. Dan zijn de cijfers dus niet representatief in termen van urbanisatiegraad. Of anders gezegd, de trekkingskansen voor leerlingen met een hoge urbanisatiegraad zijn veel groter dan die met een lage urbanisatiegraad. In hoeverre de steekproef representatief is kunnen we achteraf in kaart brengen door te kijken hoe de verdeling van specifieke subgroepen (bijv. jongens en meisjes of vwo- en vmbo-leerlingen) in onze dataset is, in vergelijking met de verdeling van die groepen in de totale populatie in de betreffende GGD regio op basis van CBS data.

Trekkingskansen

Tijdens de uitvoer van de GMJ 2015 werd destijds duidelijk dat door het design van de GMJ de trekkingskansen per regio sterk verschillen. Dit kwam door de 3 verschillende deelname niveaus destijds (landelijk, regionaal, gemeentelijk. In 2019 is de situatie anders. Er zijn minder verschillen tussen de deelnemende leerlingen van de verschillende regio's. Dit komt omdat in 2019 er 2 deelname niveaus zijn en het merendeel van de GGD'en (n=19) alle leerlingen includeren voor gemeentelijke cijfers.

Gezondheidsmonitor Jeugd 2019

per GGD-regio

Wijze van deelname
 Regionale cijfers
 Gemeentecijfers



Bij 6 GGD'en die op regionaal niveau meedoen (de zgn steekproef regio's) is een steekproef van scholen getrokken door Trimbos, rekening houdend met het Peilstation Onderzoek van Trimbos. In elke steekproef regio waren 386 leerlingen nodig voor VO2 en VO4 (afkomstig van ongeveer 10 scholen om clusteringeffecten te voorkomen. De verhouding tussen het totaal aantal scholen in een regio en het aantal scholen dat gevraagd wordt mee te doen, verschilt per regio. Hierdoor

hebben sommige leerlingen minder kans en andere leerlingen juist meer kans om uitgenodigd te worden om mee te doen.

In de 'steekproef-regio's' wordt de trekkingskans van een leerling ook beïnvloed door het aantal leerlingen op zijn/haar school. Door het trekken van een random steekproef hebben "grote scholen" evenveel kans als "kleine scholen" om in de steekproef te komen. Echter, het aantal klassen dat een school moest leveren was niet afhankelijk van het aantal leerlingen op die school. Daardoor is de trekkingskans voor leerlingen op een grote school kleiner dan voor leerlingen op kleine scholen. Deelnemende leerlingen van grote scholen tellen dus zwaarder mee dan leerlingen van kleine scholen en krijgen dus een hoger gewicht.

Representativiteit van de gerealiseerde steekproef

Van de random getrokken steekproef scholen moesten klassen worden uitgenodigd naar rato van de verdeling van onderwijsniveaus in de betreffende GGD regio. Door de selectie van klassen en door non-respons van scholen en/of leerlingen kan het zijn dat de werkelijke steekproef afweek van de beoogde steekproef en daarmee niet representatief was voor de leerlingen in de regio. De representativiteit is verschillend gedefinieerd tussen de landelijke/regionale en de gemeentelijke gewichten. Stedelijkheid is bijvoorbeeld wel van toepassing voor de landelijke gewichten maar niet voor de gemeentelijke gewichten.

2. Aanpak van de weging

Voor de procedure van de weging zijn een aantal bronnen met referentiecijfers gebruikt.

Bronnen voor het bepalen van verschillen in trekkingskansen:

- 1) De lijst met de aantallen VO scholen per regio die ook is gebruikt voor de steekproeftrekking (zie bijlage 3 van het handboek) is gebruikt om te wegen voor het aantal scholen dat heeft meegedaan in verhouding tot het totaal aantal scholen in de regio.
- 2) Het DUO bestand "leerlingen-vo-per-vestiging-naar-onderwijstype-2017-2018.xls" waarin voor elke school de leerlingenaantallen per leerjaar te vinden zijn, is gebruikt om in de 'steekproef-regio's' te wegen voor de verschillen in trekkingskansen tussen leerlingen van grote en kleine scholen.

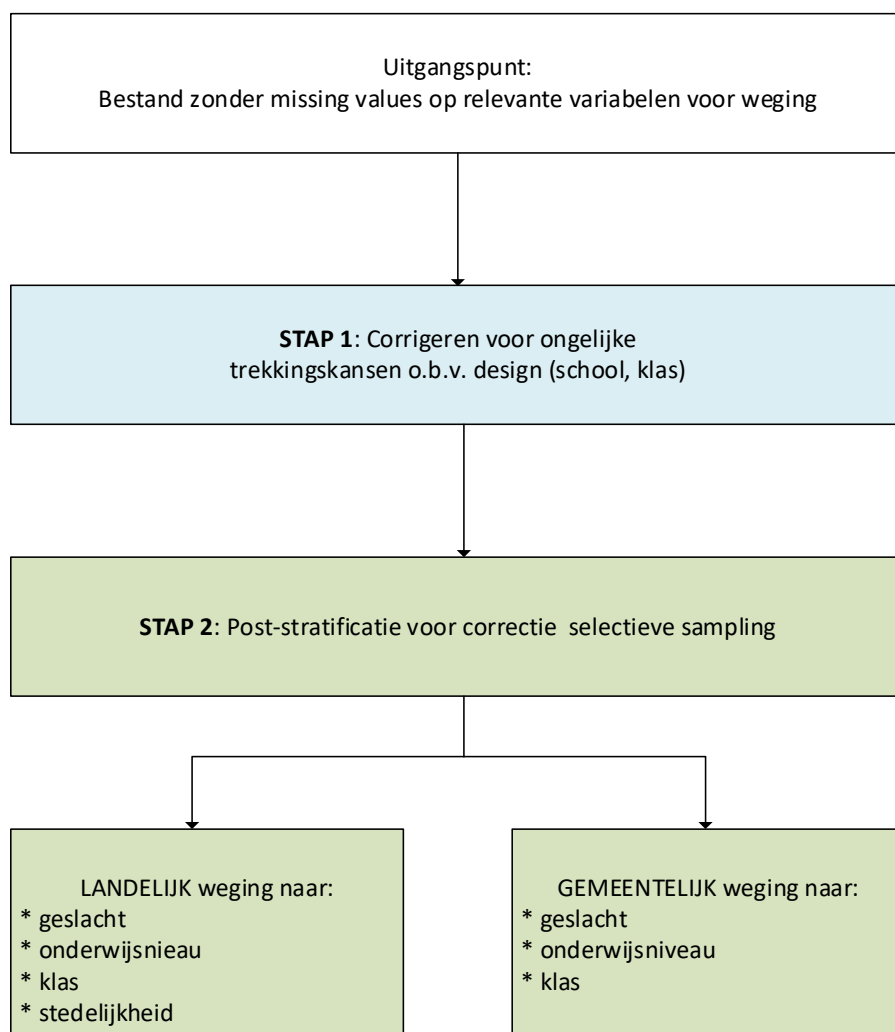
Bron voor het bepalen van niet-representativiteit:

Voor de weging voor niet-representativiteit van de landelijke en regionale gewichten is gebruik gemaakt van een bestand met voor elke GGD-regio de aantallen leerlingen per geslacht per leerjaar (tweede of vierde) per onderwijsniveau per stedelijkheidsgraad van de gemeente waar de leerling woont. Dit bestand is door het CBS speciaal voor de GMJ aangemaakt (cq maatwerktable CBS). Voor de gemeentelijke gewichten is gebruik gemaakt van een bestand met voor elke gemeente de aantallen leerlingen per geslacht, per leerjaar en per onderwijsniveau. Deze is voor je eigen GGD opvraagbaar bij het RIVM.

3. Constructie van de weegfactoren

Constructie van de weegfactoren

Voor de constructie van weegfactoren is het uitgangspunt dat er géén *missende waarden* in de relevante variabelen voor de weging mogen zitten. Dit is bewerkstelligd door deze indicatoren op te schonen of indien dit niet mogelijk bleek de persoon in kwestie uit de steekproef te verwijderen. Met dit uitgangspunt, bestaat de constructie van de weegfactor uit 2 stappen. In onderstaand stroomdiagram worden de twee stappen van de weging samengevat waarna een toelichting volgt.



- 1) De eerste stap bestaat uit het bepalen van de trekkingskansen van leerlingen op basis van het steekproefdesign. De inversen ($1 / \text{trekkingskans}$) van deze kansen zijn de zogenaamde gewichten.
- 2) In de tweede stap (de post-stratificatie) worden deze gewichten aangepast door de steekproefaantallen te vergelijken met de leerlingenaantallen volgens het CBS voor verschillende zogenaamde strata (de zogenaamde representativiteit). Bijvoorbeeld, is de verdeling over geslacht en stedelijkheid in de steekproef gelijk aan die in de populatie. Zo nee, dan moeten de gewichten aangepast worden.

De eerste stap is gelijk voor de bepaling van de landelijke/regionale en gemeentelijke gewichten. De tweede stap is verschillend, omdat de representativiteit verschillend gedefinieerd is. Bij het bepalen van gemeentelijke weegfactoren hoeven we geen rekening te houden met stedelijkheid, immers deze is voor elke inwoner hetzelfde per gemeente.

De regionale en landelijke cijfers worden gewogen naar de volgende vier kenmerken: Geslacht, leerjaar (klas 2 of klas 4), onderwijsniveau (vwo, havo, vmbo gt, of vmbo bk), en stedelijkheid van de gemeente waar de leerling woont.

De gemeentecijfers worden gewogen naar drie kenmerken: Geslacht, leerjaar (klas 2 of klas 4), onderwijsniveau (vwo+ havo, of vmbo-gt+vmbo bk).

Opbrengst

De constructie van de weegfactoren in deze twee stappen levert 3 verschillende weegfactoren op.

- 1) Alle respondenten krijgen een weegfactor die wordt gebruikt om de regionale en landelijke prevalenties zo goed mogelijk te schatten. Deze weegfactor geeft aan hoeveel leerlingen één respondent vertegenwoordigt, rekening houdend met de bovengenoemde achtergrondkenmerken.
- 2) In 19 GGD regio's krijgen deelnemers een tweede weegfactor voor de gemeentecijfers. In de 6 GGD regio's die alleen regionaal meededen is deze variabele missing voor de betreffende respondenten. Voor de gemeentelijke weegfactoren is uitgegaan van de gemeentecodes van 2019. Er is dus rekening gehouden met de gemeentelijke herindelingen van 2018 t.o.v. 2019.
- 3) Weegfactor voor gecombineerde gemeenten d.w.z. die gemeenten die (onder 2) géén gemeentelijk gewicht konden krijgen, maar na combineren met andere gemeenten wel. Zie ook verdere uitleg bij hoofdstuk over Beperkingen.

Statistisch pakket

Voor het berekenen van de weegfactoren wordt het R-package **survey** gebruikt, met als referenties Lumley T (2019). "survey: analysis of complex survey samples." R package version 3.35-1. Lumley T (2004). "Analysis of Complex Survey Samples." *Journal of Statistical Software*, 9(1), 1-19. R package version 2.2.

4. Beperkingen

Bij het berekenen van de weegfactoren is rekening gehouden met de complexiteit van het design van de studie en beperkingen van de verkregen data. De haken en ogen van het design en de beperkingen van de data worden in dit hoofdstuk besproken, alsmede de gekozen oplossingen.

Werkelijk gehaalde versus geplande deelname van scholen

Non-respons van scholen heeft invloed op het al dan niet verkrijgen van betrouwbare gewichten.

Door non-respons van scholen kan het voorkomen dat in een bepaalde gemeente te weinig leerlingen hebben meegedaan. Voor gemeenten waar dat het geval is, kunnen dan geen

gewogen gemeentecijfers worden geschat. Een mogelijke oplossing was het combineren van gemeenten.

Gemeenten die na overleg met de GGD'en zijn samengenomen zijn:

- Appingedam met Delfzijl en Loppersum
- Delfzijl met Appingedam en Loppersum
- Pekela met Veendam
- Oldambt met Westerwolde
- Laren met Blaricum
- Gennep met Bergen
- Vaals met Sittard/Geleen

Twee gevallen kunnen hier onderscheiden worden. 1) De afzonderlijke gemeenten kregen géén gewichten door te lage respons, maar kregen na combineren wél gewichten. Of 2) één van de twee (of drie) gemeenten kreeg wél gewichten door voldoende respons, maar de buurgemeente(n) niet. In dat geval behield de eerstgenoemde gemeente de oorspronkelijke gewichten, en kreeg alleen de buurgemeente(n) gewichten ná combineren.

Voor bovenstaande gemeenten is een aparte gemeentelijke weegfactor gemaakt. Niet alle gemeenten die te weinig aantallen hadden hebben zo'n gecombineerde weegfactor gekregen (deze GGD'en kiezen voor het hebben van ongewogen cijfers waarbij altijd rekening moet worden gehouden dat deze niet helemaal representatief kunnen zijn). In onderstaand kader wordt een toelichting gegeven om GGD'en te helpen bij het maken van een keuze op welke manier ze het beste de cijfers kunnen duiden bij lage aantallen: ongewogen of met een gecombineerde weegfactor.

In perspectief plaatsen van gecombineerde gewichten: Onzekerheid van uitkomsten

Bij het berekeningen van de gewichten op basis van de geïncludeerde leerlingen speelt het bekende statistische probleem van de onzekerheid van de uitkomsten. Hoe kleiner de steekproef, hoe groter de onzekerheid van de gewichten en daarmee de onzekerheid van de daarmee berekende variabelen. Vandaar dat we een minimale steekproefomvang voor iedere cel (stratum) toegepast hebben. Dit leidde ertoe dat voor sommige gemeenten vanwege de te geringe steekproefomvang geen gewichten berekend konden worden. We kunnen dit probleem concreet maken met een voorbeeld: stel dat we in een gegeven gemeente één VWO'er hebben getrokken, die rookt, in tegenstelling tot de overige niet-getrokken VWO'ers. Dan wordt het rookgedrag van 'de VWO'er' wel erg vertekend door deze leerling. Uitgaande van het statistische idee van 'borrow information from neighbours' kunnen we gewichten berekenen door de steekproefleerlingen van deze gemeente te combineren met die van een naburige, vergelijkbare gemeente. Dat leidt ertoe dat we wél gewichten kunnen berekenen voor de leerlingen van de gegeven gemeente. De toegepaste gewichten zijn gebaseerd op de gecombineerde gemeente, dus zijn enigszins gebiased. Anderzijds zijn deze gebaseerd op grotere steekproefaantallen, dus de onzekerheid is minder groot. Dit fenomeen staat in de statistiek bekend als het uitrusten van bias en variantie om de totale kwadratische afwijking van de werkelijke waarde te verkleinen. Concluderend kan gesteld worden i.g.v. de gemeenten met te kleine steekproefomvang: zonder weging zijn de resultaten niet gebiased, maar wel met een grote onzekerheid; met weging zijn de resultaten enigszins gebiased, maar met een kleinere onzekerheid. I.g.v. grote onzekerheden en vergelijkbare (buur)gemeenten zal de bias relatief klein zijn. Het is helaas in dit verband niet mogelijk om deze kwalitatieve conclusie hier kwantitatief te onderbouwen voor specifieke gemeente voor specifieke variabelen.

Missende waarden op voor de weging essentiële variabelen

Om aan een deelnemende leerling een weegfactor toe te kennen, moet van deze leerling ten minste geslacht, klas, schoollocatie (en daarmee GGD school-regio), onderwijsniveau en de woongemeente bekend zijn. Met de woongemeente zijn ook de stedelijkheid en de GGD regio waar leerling woont bekend.

Bij de GMJ 2019 is ervoor gekozen dat deze vragen niet verplicht waren om in te vullen. Dat heeft geleid tot missende waarden op deze kenmerken. Hiermee is als volgt omgegaan:

Missings geslacht:

- Deelnemers die niet hebben aangegeven of ze een jongen of een meisje zijn, worden uit het bestand verwijderd. Dit was bij 4009 deelnemers het geval. Helaas was dit onvoorzien en aanzienlijk meer dan in 2015. Er is gekeken of geslacht geïmputeerd kon worden middels een andere indicator, echter er waren geen indicatoren met een hoge correlatie met geslacht.

Missings postcode, woonplaats:

- Als er geen of een onmogelijke postcode-4 is ingevuld in de vragenlijst, werd de leerling gevraagd naar de plaats waarin hij/zij woont middels een lijst van plaatsnamen. Indien deze ook niet werd ingevuld dan is een aanname van de gemeentecode gemaakt op basis van de door de GGD'en aangereikte gemeentecode welke het meest voorkomt bij de betreffende school. Postcodes en woonplaatsen zijn omgezet naar gemeentecodes. Gemeentecodes werden als mogelijk beschouwd indien zij behoorde tot de GGD regio waar de leerling naar school gaat of tot een aangrenzende GGD regio.

Op die manier kon voor iedereen een gemeentecode aangemaakt worden en daarmee ook stedelijkheid. Stedelijkheid is niet aangemaakt voor leerlingen die in het buitenland woonachtig zijn (N=49, GGD NOG).

Missings leerjaar (klas):

Indien door de respondent niet is ingevuld in welk leerjaar (klas) hij/zij zit dan is op basis van leeftijd de klas geïmputeerd. Klas 2 voor de leeftijd 12, 13, 14 jaar en Klas 4 voor 15, 16, 17 en 18 jaar en ouder.

Missings Onderwijsniveau:

Als voor leerjaar en/of onderwijsniveau waarde is ingevuld, worden de ontbrekende waarden vervangen op basis van gegevens die GGD'en hebben aangereikt op het niveau van schoollocatie. Bijvoorbeeld in het geval een schoollocatie een gymnasium is, dan is de missing geïmputeerd als VWO. Onmogelijke waarden op onderwijsniveau zijn indien mogelijk ook vervangen op basis van de door de GGD aangereikte informatie over scholen. Het is niet altijd mogelijk om iemand toe te wijzen aan het juiste onderwijsniveau, bijvoorbeeld in het geval van brede scholen met alle niveaus. Uiteindelijk bleven er 1299 missings op onderwijsniveau (4 niveaus) over.

Als een leerling in de vragenlijst meerdere onderwijsniveaus heeft aangekruist (bijvoorbeeld als hij/zij in een gecombineerde havo-vwo klas zit), dan wordt op de volgende manier één van de vier onderwijsniveaus toegewezen:

- VWO
- HAVO bevat HAVO en de gemengde klassen VWO/HAVO en VWO/HAVO/VMBO-gt en VWO/HAVO/VMBO-alle leerwegen
- VMBO-gt bevat VMBO-gt en gemengde klassen HAVO/VMBO-gt en HAVO/VMBO-alle

leerwegen

- VMBO-bk bevat VMBO-bk en gemengde klassen VMBO-alle leerwegen

GMJ-deelnemers uit 2019 versus referentiepopulatie uit 2018

De verschillen tussen de werkelijkheid van de referentiecijfers uit 2018-2019 en de feitelijke data verzameld in schooljaar 2019-2020 resulteerden in twee soorten problemen.

- 1) 'Overvolle cellen': in theorie kunnen er niet meer leerlingen in specifieke subgroepen meedoen met de GMJ dan er leerlingen bestaan in die subgroepen. In de praktijk gebeurt dat wel. De oorzaak van dit probleem is dat de referentiecijfers het schooljaar 2018-2019 betreffen, terwijl de data zijn verzameld in het schooljaar 2019-2020 (waarvoor nog geen referentiecijfers beschikbaar waren). Verschillen tussen de twee schooljaren kunnen verschillen zijn in bijvoorbeeld het aantal vierde-klassers op een school of bijvoorbeeld een verandering in de stedelijkheid van een gemeente.
Om dit ongewenste effect te voorkomen is aangenomen dat als er meer leerlingen hadden deelgenomen aan de GMJ (in schooljaar 2019-2020) dan er volgens de referentiecijfers in de totale populatie waren (in schooljaar 2018-2019), gespecificeerd naar geslacht, klas, onderwijsniveau en stedelijkheid, dan het aantal in de totale populatie gelijk te stellen aan het aantal deelnemers.
- 2) 'Lege cellen': het kan zijn dat specifieke subgroepen (bijv. jongens in Havo-4 met stedelijkheid 2 niet voorkomen bij de GMJ-deelnemers, terwijl ze in werkelijkheid in de populatie wel bestaan. Voor de weging is het noodzakelijk dat elke cel die in de totale populatie gevuld is, ook respondenten uit de steekproef bevat. Cellen die geen of onvoldoende deelnemers bevatten (wel gevuld in de populatie) zijn daarom gecombineerd met wel c.q. voldoende gevulde cellen.

Daarvoor zijn de volgende criteria toegepast:

- voor de landelijke/regionale gewichten: cellen voldoende gevuld = minimaal 6 deelnemers per cel
- voor de gemeentelijke gewichten: cellen voldoende gevuld = minimaal 5 deelnemers of 50% van het populatie-aantal per cel.

Het combineren van cellen gebeurde m.b.v. een vaste zoekprocedure. Voor de landelijke/regionale gewichten was de zoekvolgorde als volgt:

- o eerst directe burens m.b.t. urbanisatiegraad
- o dan alle burens m.b.t. urbanisatiegraad
- o dan buur m.b.t. geslacht
- o dan directe burens m.b.t. onderwijstype
- o dan alle burens m.b.t. onderwijstype
- o dan alle burens m.b.t. urbanisatiegraad, geslacht en onderwijstype gecombineerd

De term buur verwijst naar de naburige cellen van een gegeven cel, die verschillen in de waarde voor één stratificatievariabele, maar met gelijke waarden voor de overige variabelen. Voor de gemeentelijke gewichten is altijd HAVO met VWO gecombineerd en VMBO-gt met VMBO-bk.

Grensoverschrijders

Leerlingen zijn geïnccludeerd via hun scholen en hun trekkingskansen hangen dus af van het aantal scholen en klassen dat is geïnccludeerd in de regio waar ze naar school gaan.

De regionale en gemeentelijke prevalentiecijfers zijn gebaseerd op waar leerlingen *wonen* en dat zijn dus deels andere leerlingen dan waar ze naar school gaan.

We onderscheiden hierbij 2 typen 'grensoverschrijders':

- 1) leerlingen die in een andere gemeente wonen dan waar ze naar school gaan maar wel binnen dezelfde GGD regio (gemeente grensoverschrijders).
- 2) Leerlingen die in een andere gemeente wonen dan waar ze naar school gaan waarvan de woongemeente onder een andere GGD regio valt (regio grensoverschrijders)

Beide typen grensoverschrijders vormen een uitdaging bij het berekenen van de weegfactoren. Dit is met name een probleem bij grensoverschrijders tussen GGD regio's waarvan de deelnameniveaus verschillend zijn, d.w.z. de steekproefregio's en de andere GGD regio's. Onderstaand voorbeeld ter illustratie.

Een voorbeeld

Een leerling (bijvoorbeeld een jongen in Havo-4 met stedelijkheid 2) die naar school gaat in een regio met een minimale steekproef krijgt daar op basis van zijn trekkingskansen een heel hoge weegfactor: hij telt misschien wel voor 500 vierdeklassers in z'n schoolregio. Als hij echter woont in een regio waar vrijwel alle leerlingen hebben mee gedaan en dus een weegfactor net boven de 1 hebben gekregen, kan hij met z'n hoge weegfactor zorgen dat er meer jongens in Havo-4 met stedelijkheid 2 in de steekproef lijken te zijn dan er in werkelijkheid in die regio wonen. Bij de tweede stap van de weging (de weging voor niet representativiteit) worden de 'trekkings-weegfactoren' voor alle jongens in Havo-4 met stedelijkheid 2 in die regio naar rato 'herverdeeld' zodat de weegfactoren opgeteld gelijk zijn aan het totaal aantal Havo-4 jongens met stedelijkheid 2 dat in de regio woont. Als, zoals in dit voorbeeld, één of enkele grensoverschrijders met hoge weegfactoren meetellen in een woonregio met lage weegfactoren, kan 'herverdeling' van de weegfactoren leiden tot weegfactoren < 1 voor leerlingen die in de betreffende regio naar school gaan en wonen.

In het algemeen geldt dat na post-stratificatie de som van de weegfactoren altijd gelijk is aan het aantal leerlingen in een stratum. Weegfactoren < 1 zijn dus het gevolg van het samenvoegen van leerlingen in één cel met (zeer) verschillende trekkingskansen. In 'werkelijkheid' kan een deelnemer natuurlijk niet meetellen voor minder dan 1 persoon, want hij/zij telt voor tenminste zichzelf. Om te voorkomen dat er te grote verschillen ontstaan tussen de volgens bovengenoemde procedure berekende gewichten is achteraf nog een correctieprocedure toegepast. Voor gemeenten met gewichten kleiner dan 1 zijn per stratum (geslacht, leerjaar en gecombineerd onderwijsniveau) de gewichten < 1 zo veel als mogelijk opgehoogd tot waarde 1 en de gewichten > 1 overeenkomstig verlaagd, zodanig dat de som gelijk bleef aan het aantal leerlingen in het stratum.

Extremen correctie

In 2015 is er een procedure ontwikkeld voor het corrigeren van extreme weegfactoren. Tijdens de checks van de extremen-correctie n.a.v. de rectificatie weegfactoren in juli 2020 bleek dat er onvoldoende gecorrigeerd werd voor extremen volgens de procedure uit 2015. De verdeling van de gewichten bleek nog schever te zijn dan in 2015. Daarvoor is er gezocht naar een andere vorm van correctie die met de verschillende verdelingen van de gewichten tussen de regio's rekening houdt. Er is gekozen om te corrigeren voor extreme gewichten op basis van in de literatuur beschreven methoden voor 'trimming' (Zie example 6 en 7 van Henry et al. 2012: [Methods for Adjusting Survey Weights When Estimating a Total](#)). De controles die zijn uitgevoerd zijn onder andere een vergelijking van de effecten van meer of minder stringente extremen-correctie op de verdeling van de gewichten. Daarnaast is de gebruikte methode getest op twee uitkomst variabelen (roken trekje of hele sigaret en binge drinken) om zo te komen tot een optimale mate van extremen-correctie. Deze correctie voor extremen is een *aanvulling* op de methode van 2015 welke in deze situatie beter passend is op de data. De cijfers van 2015 en 2019 zijn ondanks deze aanpassing wel vergelijkbaar.