Uitgebreide omschrijving methode en data-analyse bij “Kenmerken van groepen ouders die hun kinderen deels of niet laten vaccineren volgens het Rijksvaccinatieprogramma”

September 2025

Auteurs:

Lars Hanen, datascientist GGD West-Brabant

Leonard Vanbrabant, datascientist GGD West-Brabant

Dymphie Mioch, Onderzoeker GGD West-Brabant

Lieke Raaijmakers, Onderzoeker GGD West-Brabant

Marjolein Stoopendaal, Onderzoeker GGD West-Brabant

Dit project is financieel mede mogelijk gemaakt door het RIVM en uitgevoerd in samenwerking met Marijke Langenberg (Arts Infectieziektebestrijding GGD WB), Sasha Pauli (arts Jeugd en Gezondheid en aandacht functionaris RVP GGD WB), Jeannine Hautvast (AMPHI), Esther Lodder (arts M&G GGD WB), Hans Augustijn (REC regio Zuid), Susan Hahne (onderzoeker RIVM).

Inhoud

[1 Inleiding 3](#_Toc188968144)

[2 Dataset beschrijving 3](#_Toc188968145)

[2.1 Features 3](#_Toc188968146)

[2.1.1 Kidos 4](#_Toc188968147)

[2.1.2 CBS 5](#_Toc188968148)

[2.1.3 CoronIT 5](#_Toc188968149)

[2.2 Feature selection 5](#_Toc188968150)

[3 Data bewerkingen 6](#_Toc188968151)

[3.1 Feature engineering 6](#_Toc188968152)

[3.2 Missing value imputation 6](#_Toc188968153)

[3.3 Scaling 7](#_Toc188968154)

[3.4 Uniform Manifold Approximation and Projection (UMAP) dimensionaliteitsreductie 7](#_Toc188968155)

[4 Clusteranalyse 8](#_Toc188968156)

[4.1 Interpretatie clusterresultaten 9](#_Toc188968157)

[5 Bijlage 1: uitgebreide beschrijving selectie Kidos data 10](#_Toc188968158)

[5.1 Zuigelingen 10](#_Toc188968159)

[5.2 Stroomschema zuigelingen 11](#_Toc188968160)

[5.3 Schoolgaanden 12](#_Toc188968161)

[5.4 Stroomschema schoolgaanden 12](#_Toc188968162)

[6 Bijlage 2: vaccinatieschema 2020-2024 14](#_Toc188968163)

# Inleiding

Dit is een uitgebreide beschrijving van de methode bijbehorende bij het onderzoek “Kenmerken van groepen ouders die hun kinderen deels of niet laten vaccineren volgens het Rijksvaccinatieprogramma”. Dit is een project uitgevoerd door GGD West-Brabant, in samenwerking met AMPHI en het RIVM, gefinancierd door het RIVM (Projecten regionale ondersteuning).

Deze beschrijving is ondersteunend aan de Jupyter Source Files die de volledige code bevatten en voornamelijk geschreven zijn voor data-scientisten. Naast de analyses en rationalisatie daarachter, is in de appendix beschreven op welke manier data uit Kidos ontsloten zijn en met welke definities is gewerkt. Kidos is het data registratiesysteem wat de Jeugdgezondheidszorg (JGZ) in GGD West-Brabant gebruikt. Doel van deze uitgebreide beschrijving is om het werken met registratiedata van de JGZ en met clusteranalyses inzichtelijk te maken, zodat andere GGD’en deze methode ook kunnen uitvoeren.

Het volledige artikel met daarin de aanleiding, resultaten en discussie is ingediend bij TSG en is momenteel onder review.

# Dataset beschrijving

De analyse is gebaseerd op een real-time SQL (Structured Query Language) dataset van GGD West-Brabant waarin bepaalde Kidos data kunnen worden ontsloten. In Kidos worden onder andere alle vaccinaties geregistreerd die de GGD geeft aan kinderen binnen de regio van uitvoering. Er zijn twee afzonderlijke datasets ontsloten:

1. Alle vaccinatieregistraties van kinderen geboren in 2020 en 2021. Aan deze dataset wordt vanaf nu gerefereerd als ‘zuigelingen’, in overeenstemming met de definitie van het RIVM en bevat data over de vaccinaties die deze groep ontvangen zou moeten hebben volgens het Rijksvaccinatieprogramma;
2. Alle vaccinatieregistraties van kinderen geboren in 2012 en 2013. Dit betreft de dataset “schoolgaanden”, in overeenstemming met de definitie van het RIVM, gericht op de vaccinaties die zij tussen hun achtste en tiende levensjaar ontvangen zouden moeten hebben volgens het Rijksvaccinatieprogramma.

Een meer gedetailleerde beschrijving van de ontsluiting van de vaccinatieregistraties en het toekennen van de definities gevaccineerd, ongevaccineerd of deels gevaccineerd is in bijlage 1 opgenomen.

Aan de kenmerken van zowel de zuigelingen als de schoolgaanden zijn de volgende gegevens gekoppeld:

* **CBS**: Sociodemografische variabelen op [buurtniveau](https://www.cbs.nl/nl-nl/cijfers/detail/85318NED) uit 2022;
* **CBS**: Sociodemografische variabelen op [PC4](https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode) niveau uit 2022;
* **CoronIT**: COVID-19 vaccinatiestatus van beide ouders (peildatum 1 juli 2024).

## Features

Een volledig overzicht van alle features en de operationalisatie daarvan is op te vragen via [onderzoek@ggdwb.nl](mailto:onderzoek@ggdwb.nl)

### Kidos

Binnen de onderzoeksdatabase (ODB) van GGD West-Brabant zijn verschillende tabellen opgenomen met Kidos data. De Kidos\_observatie tabel bevat losse registraties van verschillende contactmomenten. Eerst is een selectie gemaakt om alleen contactmomenten rondom vaccinaties te includeren. Uit de tabel Kidos\_observaties zijn vervolgens de volgende variabelen gehaald:

* Geboortejaarmaand
* Tweeling
* Buurt\_BRP (basisregistratie personen)
* CBS\_Buurtcode\_BRP
* GeboortejaarmaandBiologischeMoeder
* LeeftijdBiologischeMoederBijGeboorte
* Geslacht
* Geboorteland
* Postcode6\_BRP (voor schoolgaanden uit oproeplijst)
* Huisnummer\_BRP (indien nodig en voor schoolgaanden indien nodig uit oproeplijst). Huisnummer is alleen gebruikt wanneer er de combinatie geboortejaarmaand van ouder + PC6 + geslacht geen unieke combinatie opleverde, welke nodig is voor de koppeling met CoronIT (zie 2.1.3)

Voor zuigelingen is voor postcode, CBS buurtcode en huisnummer een random waarde tussen de 0-24 maanden geselecteerd. De reden hiervoor is dat de relevante vaccinaties tussen de leeftijd van 0-2 jaar zouden moeten worden gegeven. Indien zuigelingen binnen die twee jaar op verschillende adressen gewoond hebben, wordt er random één geselecteerd.

In het geval van schoolgaanden worden postcode en huisnummer (indien nodig), uit de oproeplijsten gehaald. Andere variabelen, zoals geboortejaarmaand, geboorteland, geslacht en tweeling zijn onveranderlijk.

Uit een andere ODB tabel, Kidos\_relaties, zijn vervolgens voor zowel zuigelingen als schoolgaanden de volgende variabelen gehaald:

* Siblings
* Geboortejaarmaand\_relatie\_ouder1
* Geboorteland\_relatie\_ouder1
* Geslacht\_relatie\_ouder1
* Geboortejaarmaand\_relatie\_ouder2
* Geboorteland\_relatie\_ouder2
* Geslacht\_relatie\_ouder2

De geboortelanden van ouders in combinatie met het geboorteland van de schoolgaande of zuigeling zelf, zijn gebruikt om de migratieachtergrond te bepalen. Deze variabele bestaat op basis van de CBS indeling uit drie categorieën: ‘Nederlands’, ‘Kind van migrant’, ‘Migrant’.

Uit de ODB tabel Kidos\_toestemmingen wordt de variabele Toestemming gegevensuitwisseling RVP gehaald.

### CBS

De Kidos data zijn verrijkt met CBS data op postcode-4 niveau (PC4 omdat een koppeling met PC6 tot veel onbekenden leidde) en buurtniveau. Dit betreft sociodemografische variabelen waaronder:

* **Bevolking**: huishoudens (eenouder- of tweeouder gezinnen), leeftijden, geslacht, burgerlijke staat, herkomst en migratieachtergronden;
* **Bouwen en wonen**: verhouding koop- en huurwoningen, leegstand, WOZ-waarden, stedelijkheid, meergezinswoningen;
* **Arbeid**: participatie;
* **Inkomen**: huishoudensinkomens, vermogen, aandeel onder sociaal minimum;
* **Sociale zekerheid**: uitkeringen;
* **Opleidingsniveau**: aandeel laag/hoog;
* **Zorg**: jongeren met jeugdzorg;
* **Motorvoertuigen**: personenauto's;

De CBS variabelen die een postcode- of buurtgebied in een absoluut cijfers uitdrukken, zoals het aantal hoog- en laagopgeleiden, eenoudergezinnen of woningen, zijn door het corresponderende totale aantal gedeeld om tot proporties te komen.

Voor sommige variabelen geldt dat ze door het CBS afgerond zijn en daardoor een ordinaal karakter hebben in plaats van continu. Ordinale variabelen met >5 categorieën zijn behandeld als continue variabelen. Ordinale variabelen met 5 categorieën of minder one-hot-encoden we zodat deze als binaire variabelen worden behandeld.

### CoronIT

CoronIT is het registratiesysteem van coronavaccinaties en coronatesten van de GGD’en. De Kidos data van elke geregistreerde ouder is verrijkt met een COVID-19 vaccinatiestatus vanuit CoronIT. Koppeling vond plaats op basis van zijn of haar PC6, huisnummer (indien nodig), geboortejaar en -maand en geslacht (indien nodig).

## Feature selection

Voorafgaand de verdere databewerkingen is een selectie gemaakt in de variabelen op verschillende manieren om op basis van logica een eerste stap in dimensionaliteitsreductie uit te voeren;

1. Er is gekeken de mate van correlatie tussen continue features om te bekijken of bepaalde continue features in hoge mate met elkaar correleren volgens de Pearson correlatie coëfficiënt. Echter is besloten deze features niet bij voorbaat uit te sluiten. Na de dimensionaliteitsreductie die later plaatsvindt, spelen sterk correlerende features namelijk automatisch een minder grote rol.
2. Er is een selectie van de CBS variabelen op buurt- en PC4 niveau gemaakt op basis van relevantie (e.g. bouwjaar en aantal niet bewoonde woningen zijn niet geïncludeerd; er zijn voldoende andere variabelen die meer informatie geven over het soort buurt), wederzijdse exclusiviteit (e.g. voor percentage huur- en percentage koopwoningen is alleen percentage huurwoningen geïncludeerd),interpreteerbaarheid (e.g. stedelijkheid geselecteerd in plaats van adressendichtheid).
3. Variabelen die meer dan 50% missings hadden in een subgroep waarbinnen de clusteranalyse zal plaatsvinden (de groep ongevaccineerde zuigelingen, deels gevaccineerde zuigelingen of niet/deels gevaccineerde schoolgaanden) zijn voor die betreffende groep uitgesloten van de analyses (zie bijlage 2 voor het volledige overzicht).

# Data bewerkingen

## Feature engineering

Na de verrijking van de data uit Kidos met CBS data en data uit CoronIT, zijn features geconstrueerd voor alle schoolgaanden en zuigelingen. Dit zijn voornamelijk continue kenmerken zoals het aandeel huurwoningen, maar soms ook binaire kenmerken die uit one-hot coding van categorische variabelen voorkomen (zoals ‘CBS indeling migratieachtergrond’ en ‘Tweeling’).

One-hot coding representeert iedere categorische variabele als een vector met een lengte gelijk aan het aantal categorieën, waarin één element 1 is (dit indiceert de categorie) en de rest 0.

## Missing value imputation

De ontbrekende waarden zijn geïmputeerd door middel van een enkelvoudige imputatiemethode genaamd ‘IterativeImputer’. Dit is een implementatie van de python package Scikit-learn en is geïnspireerd door de meervoudige imputatiemethode ‘MICE’.

De gebruikte methode imputeert ontbrekende waarden door over kolommen te itereren en per stap één kolom als output te behandelen en alle andere kolommen als input. Een regressiemodel, in ons geval een Random Forest, wordt vervolgens gefit op de bekende waarden in de outputkolom. Dit model voorspelt vervolgens de ontbrekende waarden, alvorens de volgende kolom te gaan behandelen. ‘max\_iter’ is een hyperparameter die vaststelt hoe vaak dit proces zich herhaald en is voor dit onderzoek ingesteld op 5 iteraties.

De datasets bevatten zowel continue als binaire variabelen. Aangezien deze missing value imputation strategie voor de imputation van de ene variabele informatie uit alle andere variabelen benut, wordt de dataset in zijn geheel als input aan IterativeImputer gegeven. De keerzijde hiervan is dat ook de voorspellingen voor binaire variabelen continue zijn, terwijl voor deze variabelen de binaire structuur behouden moet worden (van belang voor UMAP, zie hoofdstuk 3). Om dit op te lossen zijn continue voorspellingen tussen 0 en 1 afgerond tot 0 of 1. Dit gold voor de variabelen ‘Geslacht=mannelijk’, ‘Gevaccineerd\_ouder1’ en ‘Gevaccineerd\_ouder2’.

Over het algemeen is een goede imputatiewaarde een waarde die geobserveerd had kunnen worden, als deze niet ontbrekend zou zijn geweest. Om te inspecteren of imputaties passend zijn, is de eerste stap gevolgd zoals voorgesteld door (Van Buuren, 2011): het plotten van de distributies van de geobserveerde en geïmputeerde variabelen. Verschillen daarin wijzen op mogelijke problemen die nadere inspectie behoeven.

## Scaling

Continue features zijn vervolgens gestandaardiseerd om de invloed van verschillende eenheden en distributies te mitigeren. Dit is van belang voor de meeste machine learning modellen omdat ze gevoelig zijn voor de schaal van inputvariabelen. Binaire features zijn niet gescaled omdat ze voor de machine learning modellen in de volgende bewerkingsstappen al een geschikte vorm hebben.

## Uniform Manifold Approximation and Projection (UMAP) dimensionaliteitsreductie

Dimensionaliteitsreductie houdt in dat we proberen zo veel mogelijk van de variantie in de oorspronkelijke data vast te leggen in een kleiner aantal dimensies (i.e., variabelen). Voor elke rij in de dataset (een zuigeling of schoolgaand kind) resulteert dit in een nieuwe representatie in minder dimensies. Hoewel deze nieuwe representaties niet direct interpreteerbaar zijn, bevatten ze zo veel mogelijk informatie uit de oorspronkelijke data.

Dimensionaliteitsreductie is uitgevoerd, net als Cooper (2021) en Maleki (2024), om twee redenen. Ten eerste maakt een reductie naar twee dimensies visuele analyse van de ruimtelijke ordening van datapunten mogelijk. Ten tweede zijn afstandsberekeningen tussen datapunten, een fundamenteel onderdeel van clustering algoritmen, beter mogelijk naarmate de dimensionaliteit lager is.

UMAP UMAP is een non-lineair algoritme voor dimensionaliteitsreductie van datasets (McInnes, 2020). Deze behoudt ook complexere structuren, die lineaire dimensionaliteitsreductietechnieken als Principal Component Analyse vaak niet behouden. Tevens faciliteert UMAP een afweging tussen globaal en lokaal structuurbehoud (‘n\_neighbours’ instelling) en stimuleert het een duidelijke scheiding tussen dicht opeengepakte clusters.

UMAP biedt de mogelijkheid om verschillende UMAP modellen met elkaar te [combineren](https://umap-learn.readthedocs.io/en/latest/composing_models.html) als de onderliggende data hetzelfde zijn, bijvoorbeeld om twee data typen afzonderlijk te reduceren en vervolgens bij elkaar te brengen. In het huidige onderzoek is dit toegepast op continue en binaire variabelen, zodat voor beide modellen afzonderlijk een afstandsmaat ingesteld kan worden die past bij het type data. Namelijk respectievelijk de Euclidische afstand voor continue variabelen en de ‘Dice distance’ voor binaire variabelen. De modellen zijn gecombineerd door ze met elkaar te vermenigvuldigen, volgens de door UMAP voorgestelde en meest voor de hand liggende methode (zie figuur 1).

Voor de UMAP-embeddings moet de parameter n\_neighbours worden ingesteld, die bepaalt hoeveel nabije datapunten worden meegenomen bij het berekenen van de structuur. Voor schoolgaande kinderen is deze ingesteld op 100 en voor zuigelingen op 20 (default is 15). Deze verschillende waarden zijn gekozen vanwege de verschillen in datasetomvang tussen de groepen; de dataset voor schoolgaande kinderen was in deze studie groter dan die van de zuigelingen. Een hogere waarde voor n\_neighbours helpt om meer globale patronen in de data te behouden in plaats van sterk gefocust te zijn op lokale details. De parameter min\_dist, die bepaalt hoe dicht punten in een cluster op elkaar worden geplaatst, is op de standaardwaarde van 0.1 gelaten om compactere clusters te stimuleren. Deze instellingen zijn gekozen om de clustering te optimaliseren voor de volgende stappen in onze analyse.

Afbeelding met tekst, Lettertype, lijn, wit

Automatisch gegenereerde beschrijving

Figuur 1 Conceptueel ontwerp van bewerkingsstappen

# Clusteranalyse

Hierarchical Density-Based Clustering (HDBSCAN) is een clusteringmethode die datapunten in groepen deelt op basis van dichtheid en automatisch het aantal clusters bepaalt en uitschieters detecteert. Met de hyperparameter ‘min\_cluster\_size’ wordt een minimale clustergrootte ingesteld om te voorkomen dat kleine groepen datapunten door ruis als clusters worden geïnterpreteerd. Hoe hoger de ingestelde waarde, hoe minder clusters er zullen worden gevonden (Campello, 2013). ‘min\_samples’ is een andere hyperparamater en is standaard zo hoog als ‘min\_cluster\_size’. De min\_samples is verlaagd om te voorkomen dat te veel datapunten als ruis worden geclassificeerd.

Voor niet-gevaccineerde schoolgaanden is uiteindelijk een ‘min\_cluster\_size’=200 en ‘min\_samples’=150 toegepast. Voor de zuigelingen was dit respectievelijk 25 en 10. Deze instellingen resulteerden in de meest duidelijke scheiding van clusters die goed interpreteerbaar zijn.

Eerder werk heeft de bruikbaarheid van HDBSCAN in combinatie met UMAP aangetoond (Allaoui, 2020). Hierin wordt gebruik gemaakt van externe evaluatiemethoden waarvoor ground truth clusterlabels nodig zijn. Door het ontbreken daarvan hebben wij ons gebaseerd op intrinsieke eigenschappen van de clusterresultaten, met de volgende methoden:

* **De compactheid en scheiding van clusters**

De compactheid en scheiding van clusters wordt uitgedrukt in een silhouette score [-1, 1] zonder inachtneming van ruis datapunten. Hoewel deze een indicatie geeft van de mate van clustering van datapunten, biedt het gebruik in het geval van HDBSCAN beperkte evaluatiemogelijkheden door (a) de aanwezigheid van ruispunten die niet aan een cluster worden toegewezen, (b) het vermogen van HDBSCAN om clusters van onregelmatige vormen en verschillende dichtheden te identificeren, (c) en de gevoeligheid van de silhouette score voor de wijze waarop outliers en clusters worden verwerkt. De silhouette score is desondanks wel gebruikt als indicatie, maar gecombineerd met de volgende twee methoden.

* **Consistentie van clusterkenmerken met bestaande kennis in de epidemiologie**

Indien één of meerdere clusterkenmerken uit de resultaten bekend zijn in de regionale epidemiologie (ingeschat door experts), verhoogt dit de betrouwbaarheid van alle gevonden clusterkenmerken in de desbetreffende data.

* **Stabiliteit van clusterkenmerken over verschillende runs**

De resultaten van de cluster analyses zijn vergeleken over verschillende runs. Een daarvan is de mate van reductie van de dataset; lage dimensionaliteit (2D) is vergeleken met de reductie naar middelhoge dimensionaliteit (15D), omdat beide opties voordelen bieden. Lage dimensionaliteit is eenvoudig te visualiseren maar het geeft UMAP slechts twee dimensies waarin zoveel mogelijk van de oorspronkelijke variantie opgenomen kan worden. Middelhoge dimensionaliteit is niet te visualiseren maar biedt meer ruimte om variantie op te nemen.

Daarnaast zijn de UMAP resultaten over twee verschillende random states vergeleken, om de gevoeligheid voor het stochastische karakter van UMAP te evalueren. Dit is gedaan door de 2D embeddings visueel te inspecteren, om het aantal te maken vergelijkingen van clusterkenmerken beperkt te houden.

## Interpretatie clusterresultaten

In de vorige paragraaf is al kort gesproken over clusterkenmerken. Een belangrijke uitdaging is de gevonden clusters goed te interpreteren. Daarvoor zijn de volgende methoden gebruikt;

* **Supervised learning met SHAP-waarden**

Lundberg (2017) presenteerde in de originele paper een methode om machine learning modellen te verklaren door middel van Shapley values uit ‘cooperative game theory’. De rationale hierachter is dat de bijdrage van individuele features aan de output van een model inzichtelijk gemaakt kan worden, en daarmee zowel individuele voorspellingen als het algehele gedrag van een model interpreteerbaar worden.

In deze studie is net als Louhichi (2023) de SHAP library gebruikt om clusterkenmerken te interpreteren, door te visualiseren welke features van schoolgaanden en zuigelingen bijdragen aan de voorspelde clusters. Er is een multiclass Random Forest classifier getraind op de oorspronkelijke variabelen om de corresponderende clusterlabels van HDBSCAN te voorspellen. 30% van de afzonderlijke datasets is als test set gebruikt, om het voorspellende vermogen van de classifiers te evalueren. De invloed van variabelen op voorspellingen (= de clusterindelingen) zijn vervolgens gevisualiseerd met ‘SHAP Force plots’. Uit deze plots wordt de mate van invloed en richting van invloed van de verschillende variabelen op de clusterindeling inzichtelijk.

* **Vergelijking distributies van oorspronkelijke variabelen op clusterniveau**

HDBSCAN koppelt aan ieder datapunt (= ieder kind) een clusterlabel. Om de duiding te vergemakkelijken, is gekeken naar de distributies van verschillende variabelen per cluster. Bij continue variabelen is het gemiddelde, mediaan en uitschieters per clusters onderling bekeken en vergeleken. Voor binaire variabelen vergelijken we de verhoudingen.

* **Geografische clusterweergave**

Voor de onderzoekers en experts (niet opgenomen in uiteindelijke resultaten wegens privacy redenen) is als laatste stap tevens een geografische clusterweergave op wijkniveau inzichtelijk gemaakt. Per cluster zijn de CBS buurten, waar de kinderen in dit cluster wonen, gevisualiseerd. Doordat onderzoekers en experts bekend zijn met kenmerken van wijken in de eigen regio, helpt dit bij de duiding.

# Bijlage 1: uitgebreide beschrijving selectie Kidos data

## Zuigelingen

De selectie van zuigelingen uit de Kidos data en het bepalen van de vaccinatiestatus is verlopen volgens de volgende stappen:

1. Selectie van alle zuigelingen uit geboortecohort 2020 en 2021;
2. Uitsluiten van alle zuigelingen met bezwaar tegen deelname aan wetenschappelijk onderzoek;
3. Zuigelingen zonder observaties over vaccinaties zijn uitgesloten. Voor kinderen die in zorg zijn bij de GGD West-Brabant worden in principe altijd gegevens vastgelegd m.b.t. vaccinaties in Kidos. Ook wanneer vaccinaties al bij voorbaat worden geweigerd of wanneer vaccinaties niet kunnen worden toegediend bij een bepaald contactmoment wordt dit geregistreerd. Zuigelingen met minimaal één observatie over minimaal één van de volgende vaccinaties zijn geïncludeerd:
   * BMR
   * Pneumokokken
   * DKTP-Hib-Hepatitis B
   * DKTP
   * Men C
   * MenACWY
4. Zuigelingen met een onbekend woonadres in de volledige eerste 24 levensmaanden zijn uitgesloten;
5. Zuigelingen die van hun eerste tot en met twintigste levensmaand niet uitsluitend in het werkgebied van de betreffende JGZ organisatie hebben gewoond zijn uitgesloten
6. Bepaal de vaccinatiestatus o.b.v. het aantal gegeven vaccinaties tussen de leeftijd 0-2 jaar. De vaccinatiestatus is gedefinieerd in overeenstemming met de jaarverslagen van het RIVM (2024), zie bijlage 3 voor het schema. Zuigelingen zijn gedefinieerd als volledig gevaccineerd wanneer zij onderstaande vaccinaties hebben ontvangen op uiterlijk de leeftijd van 2 jaar:
   * 1 B.M.R. vaccinatie
   * Minimaal 3 vaccinaties tegen Pneumokokken
   * Minimaal 3 DKTP vaccinaties, minimaal 3 Hib vaccinaties, minimaal 3 Hepatitis B vaccinaties. Hierbij is de aanname dat zuigelingen die 3 DKTP-Hib-HepB vaccinaties krijgen, ook de 4e krijgen indien de moeder geen maternale vaccinatie heeft ontvangen. Deze aanname is gevalideerd door de verhouding tussen het aantal DKTP vaccinaties en de vaccinatiegraad voor de maternale kinkhoestvaccinatie met elkaar te vergelijken
   * 1 MenACWY/MenC

Zuigelingen zijn gedefinieerd als deels gevaccineerd wanneer zij minimaal één van onderstaande vaccinaties hebben ontvangen maar niet volledig gevaccineerd zijn. Zuigelingen die geen één van deze vaccinaties hebben ontvangen in de leeftijd 0-2 jaar, zijn geclassificeerd als niet gevaccineerd.

## Stroomschema zuigelingen

In schematische weergave zien deze stappen er, inclusief de andere beschreven stappen in dit document, als volgt uit:

Afbeelding met tekst, diagram, Technische tekening, Plan

Automatisch gegenereerde beschrijving

## Schoolgaanden

De selectie van schoolgaanden en het bepalen van de vaccinatiestatus is als volgt verlopen:

1. Selecteer de oproeplijsten van het geboortecohort, vaccinatiesoorten en jaar van toepassing. In dit geval zijn de oproeplijsten van geboortecohorten 2012 en 2013 gebruikt, m.b.t. de DTP en BMR vaccinatie op 9-jarige leeftijd.
2. Koppel Kidos cliëntnummers aan bestaande oproeplijsten
3. Sluit schoolgaanden met bezwaar tegen deelname aan wetenschappelijk onderzoek uit;
4. Tel de vaccinaties die tussen de 96e en 120e maanden (8e en 10e jaar) door de schoolgaanden ontvangen zijn;
5. Leid de vaccinatiestatus af. Volledig gevaccineerd zijn betekent:
   * Minimaal 1 BMR vaccinatie
   * Minimaal 1 DTP vaccinatie

Kinderen die niet voldoen aan deze definitie worden geclassificeerd als deels- of niet gevaccineerd.

## Stroomschema schoolgaanden

In schematische weergave zien deze stappen er, inclusief toevoeging van andere variabelen (uit Kidos), als volgt uit:

Afbeelding met tekst, diagram, Plan, handschrift

Automatisch gegenereerde beschrijving

In rood zijn de stappen aangegeven die het meest verschillen van de zuigelingen. Ten eerste is een oproeplijst gebruikt waar direct het totaal aantal schoolgaanden uit te ontsluiten is. Hierbij moet opgemerkt worden dat er enkele schoolgaanden in de oproeplijst staan zonder bekend cliëntnummer, waardoor deze niet meegenomen konden worden in het vervolg van de analyse. Ten tweede worden de vaccinaties (B.M.R. en DTP) geteld op andere leeftijden.

# Bijlage 2: vaccinatieschema 2022-2024

Afbeelding met tekst, schermopname, Website, Onlineadvertenties

Automatisch gegenereerde beschrijving