

VERDIEPEND ONDERZOEK NAAR ZIEKTEVERZUIM IN HET VOORTGEZET ONDERWIJS

In opdracht van het ministerie van OCW en Voion

Eindrapportage, 23-05-2023

Arie-Jan van der Toorn

Paul de Hek



COLOFON

DIT IS EEN UITGAVE VAN

SEOR BV

© SEOR BV / ROTTERDAM, MEI 2023

Niets van deze uitgave mag op welke wijze dan ook worden verveelvoudigd zonder de voorafgaande toestemming van de uitgever en andere auteursrechthebbenden. SEOR is niet aansprakelijk voor gegevens die door derden werden verstrekt.

VERDIEPEND ONDERZOEK NAAR ZIEKTEVERZUIM IN HET VOORTGEZET ONDERWIJS

EINDRAPPORTAGE

CONTACTPERSOON

Naam	Arie-Jan van der Toorn
Adres	Marconistraat 16, 11 ^e etage 3029 AK Rotterdam
Telefoon	010 302 0500
Mobiel	06-409 881 21
Email	vandertoorn@seor.eur.nl

INHOUDSOPGAVE

SAMENVATTING, CONCLUSIES EN AANBEVELINGEN	1
1 INLEIDING	3
1.1 Achtergrond	3
1.2 Doel en onderzoeksvragen	3
1.3 Aanpak van het onderzoek	4
2 BESCHRIJVING INHOUD EN KWALITEIT VERZUIMDATA	5
2.1 Inhoud van de data	5
2.2 Kwaliteit van de data	6
3 PATRONEN IN ZIEKTEVERZUIM	9
3.1 Sequentieanalyse	9
3.2 Clusters op basis van vergelijkbare patronen	10
4 VERKLARENDE ANALYSE ZIEKTEVERZUIM	17
4.1 Analyses op schoolniveau	17
4.2 Aanbevelingen	23
I BESCHRIJVENDE STATISTIEKEN	25

SAMENVATTING, CONCLUSIES EN AANBEVELINGEN

Achtergrond, doel en aanpak onderzoek

Dit rapport doet verslag van het verdiepende onderzoek naar verklaringen van het ziekteverzuim van personeel op het voortgezet onderwijs (2016-2021) met als doel aanknopingspunten te vinden voor beleid en het eventueel verbeteren van de Verzuimbenchmark VO¹. Het ziekteverzuim binnen het voortgezet onderwijs (vo) is, in vergelijking met het landelijk gemiddelde, hoog te noemen. Omdat een hoog ziekteverzuim een negatief effect heeft op de kwaliteit van het onderwijs is het nodig om meer inzicht te krijgen in welke factoren meespelen in dit hoge verzuim. Scholen kunnen dit uiteindelijk gebruiken voor betere beleidsmaatregelen om verzuim terug te dringen. Allereerst is in dit onderzoek geïnventariseerd welke gegevens allemaal beschikbaar zijn en wat de kwaliteit van deze gegevens is. Op basis van beschrijvende analyses, sequentie- en clusteranalyses en regressieanalyses onderzoeken we hoe het ziekteverzuim in het voortgezet onderwijs zich heeft ontwikkeld en wat de belangrijkste factoren in deze ontwikkeling zijn.

Kwaliteit van de gegevens ligt over het algemeen vrij hoog

Over het algemeen ligt de kwaliteit van de verzuimdata die DUO heeft behoorlijk hoog. De meeste variabelen zijn zo goed als volledig gevuld, en bij variabelen waar wel missende waardes zijn, gaat dit om een klein percentage. Wel zien we dat niet alle scholen terug te vinden zijn in de verzuimdata en dat dit aantal in 2019 en 2020 hoger ligt. Ook zien we dat er een aantal 'oneffenheden' in de data zitten die betrekking hebben op de registratie van het verzuim, en met name wanneer personen meerdere dienstverbanden op één of meerdere scholen hebben. In dit onderzoek worden daarin een aantal keuzes gemaakt die als basis zouden kunnen dienen voor verder gebruik van deze bestanden.

Sequentie- en clusteranalyse levert inzicht in verschillen tussen groepen

De sequentie- en clusteranalyse groepeerde de werknemers in het voortgezet onderwijs in verschillende groepen op basis van het ziekteverzuim. Vervolgens kijken we naar verschillen in persoons- en schoolkenmerken. Over het algemeen zien we dat persoonskenmerken vaker bepalend zijn voor de groep waar iemand in valt dan schoolkenmerken.

1. Nooit verzuim (6.698 personen, 10% van het totaal):
 - Meer mannen, minder onderwijzend personeel, meer fulltimers.
2. Maximaal 2x per jaar kort of middellang verzuim (32.267 personen, 46% van het totaal):
 - Meer mannen, meer jongere personen, minder onderwijzend personeel, meer fulltimers, minder op scholen in het westen van het land en met een onvoldoende van de Onderwijsinspectie.
3. Meer dan 2x per jaar kort of middellang verzuim (13.330 personen, 19% van het totaal):
 - Meer vrouwen, meer 30-40'ers, meer onderwijzend personeel, meer parttimers, meer op scholen in het westen van het land, meer op scholen met hogere leerling-leraar ratio.
4. Minder dan een jaar lang verzuim (13.100 personen, 19% van het totaal):
 - Meer vrouwen, meer ouderen, meer op scholen met onvoldoende van de Onderwijsinspectie.
5. Minimaal een jaar lang verzuim (4.218 personen, 6% van het totaal):

¹ Zie: <https://www.voion.nl/instrumenten/verzuimbenchmark-vo/>

- Meer vrouwen, meer ouderen, meer onderwijzend personeel, meer parttimers, meer op scholen met een onvoldoende van de Onderwijsinspectie.

Geen duidelijke patronen in korte, middellange, lange verzuimperiodes

Ondanks het feit dat we bovenstaande groepen hebben kunnen identificeren, zien we binnen deze groepen geen duidelijk identificerende patronen. In de groepen die ooit lang verzuim hebben gehad, zien we bijvoorbeeld geen duidelijke patronen van kort naar middellang naar lang verzuim terug. Wel is duidelijk dat het aantal dagen verzuim gemiddeld genomen toeneemt naarmate de lange verzuimperiode dichterbij komt.

Aantal effecten van schoolkenmerken en persoonskenmerken op verzuim maar verklaringskracht is laag

Bovenstaande bevindingen gaan over individuele patronen van personen. Voor scholen is het echter lastig om te sturen op individuele gevallen, en moet vooral gekeken worden of er op schoolniveau aan knoppen te draaien is. Daarom tillen we de analyse op persoonsniveau naar de scholen. Over het algemeen geldt dat de verklaringskracht laag ligt: het is lastig om op basis van de beschikbare gegevens de ontwikkeling in het verzuimpercentage te verklaren. We zien wel dat een aantal schoolkenmerken (schooltype, denominatie, mate van verstedelijking, schoolgrootte, leerling-leraar ratio) en persoonskenmerken (leeftijdverdeling, werktijdfactor) van invloed zijn op het verzuimpercentage. Ook zien we dat kenmerken verschillende effecten hebben op kort, middellang en lang verzuim en dat het korte verzuim op basis van de beschikbare kenmerken het beste te verklaren is.

Conclusies en aanbevelingen

Uit dit onderzoek komen de onderstaande conclusies, waarbij we ook direct een aantal aanbevelingen doen:

- 1) ***Kwaliteit van de data is goed, maar er ontbreken scholen die hun verzuim niet registreren.***
Scholen moeten een duidelijke prikkel hebben om het verzuim goed te registreren. Uiteindelijk hebben scholen er zelf baat bij dat ze goed weten hoe hoog het verzuim binnen hun school is, zeker gezien het lerarentekort. Deze boodschap moet aankomen bij de scholen.
- 2) ***Ondanks de hoge kwaliteit van de data is het lastig om het ziekteverzuim goed te verklaren.***
Een mogelijkheid om de data te verrijken is door een koppeling te maken met CBS-microdata. Hierdoor kunnen meer factoren onderzocht worden, bijvoorbeeld over de levensfase van personen (huishoudsituatie, wel/geen kinderen, etc.). Ook hier blijven uiteraard beperkingen aan zitten, omdat er zeer waarschijnlijk factoren van invloed zijn op het ziekteverzuim die niet kwantitatief te meten zijn.
- 3) ***Weinig aanknopingspunten om verzuimbeleid te verbeteren***
Bestaande initiatieven zoals het richten op 55+'ers lijken terecht op basis van de uitkomsten uit dit onderzoek. Daarnaast is er een groep 30- tot 40-jarige leraren op scholen met een hoge leerling-leraar ratio die wellicht meer aandacht nodig hebben, omdat we zien dat deze groep relatief vaak kort of middellang ziekteverzuim heeft.
- 4) ***Kleine aanpassingen in Verzuimbenchmark Voion mogelijk***
Het is terecht dat de huidige verzuimbenchmark van Voion voor het samenstellen van de referentiegroepen onderscheid maakt naar schooltype. Het kan goed zijn om nog meer gedetailleerde uitsplitsingen te maken naar bijvoorbeeld verstedelijking en aandelen 55+'ers. We zien dat er gemiddeld genomen verschillen zijn in het verzuimpercentage tussen scholen op basis van deze factoren.
- 5) ***Data biedt een goede basis om meer onderzoek te doen, bijvoorbeeld in de vorm van pilots***
Om meer aanknopingspunten voor beleid te vinden, zou er aanvullend onderzoek gedaan moeten worden. De beschikbare data biedt wel een goede basis voor bijvoorbeeld het uitvoeren van een pilot. Zo zou er geëxperimenteerd kunnen worden met een bepaald verzuimbeleid op een (gerandomiseerde) selectie van de scholen. Vervolgens kan de verzuimdata gebruikt worden om de effecten van het beleid te analyseren. Als alternatief kan ook geïnventariseerd worden wat de huidige verschillen zijn in het verzuimbeleid tussen scholen waarna dit gekoppeld kan worden aan de verzuimdata.

1 INLEIDING

1.1 ACHTERGROND

Het ziekteverzuim binnen het voortgezet onderwijs (vo) is in vergelijking met het landelijk gemiddelde, hoog te noemen. Zo was het verzuimpercentage onder onderwijzend personeel in 2019 gemiddeld 5,6%², terwijl dit landelijk gemiddeld 4,4% was³. Een hoog ziekteverzuim heeft meerdere negatieve consequenties. Het maakt het lerarentekort nog nijpender, heeft een negatief effect op de onderwijskwaliteit, raakt uiteraard de leraren die zich vaak of langdurig ziek moeten melden zelf, het maakt de bedrijfsvoering van de scholen niet makkelijker en heeft ook financiële gevolgen (doorbetaling en vervanging). In het kader van de toenemende lerarentekorten wordt het lerarenverzuim ook een steeds belangrijker thema. In veel regionale aanpakken lerarentekorten zijn 'Terugdringen verzuim' en 'Verzuimpreventie' en 'Duurzame inzetbaarheid' veel voorkomende actielijnen⁴. Daarbij worden veel verschillende initiatieven aangedragen, van het inzetten van coaches tot het opleiden van vitalogen onder leraren. Ook internationaal wordt veel aandacht geschonken aan het topic van lerarenverzuim, bijvoorbeeld in de policy toolbox van UNESCO⁵.

In het kader van de negatieve effecten van ziekteverzuim is het dus van belang om tot een effectieve(re) verzuimaanpak te komen. Om de verzuimaanpak effectiever te kunnen maken, is beter inzicht nodig in welke factoren een (belangrijke) rol spelen in (de stijging van) het ziekteverzuim. In het algemeen is er al veel onderzoek gedaan naar de factoren die een rol spelen bij ziekteverzuim onder personeel. Zo is er jarenlang onderzoek gedaan bij de PTT (1991-2000) naar beïnvloedbare en niet-beïnvloedbare factoren in verzuim, waarbij leeftijd, geslacht, functie en bedrijfsgrootte als belangrijkste niet-beïnvloedbare factoren naar boven kwamen. Daarnaast komt uit onderzoek ook naar voren dat de fase in het leven van belang is bij verzuim (het hebben van kinderen), en dat de cultuur op een school ook bijdraagt aan de mate van verzuim onder onderwijspersoneel.

1.2 DOEL EN ONDERZOEKSVRAGEN

Uit bovenstaande paragraaf blijkt al dat de redenen achter verzuim divers zijn, en vaak gerelateerd aan zowel persoons- als schoolkenmerken. In de Verzuimbenchmark-VO zijn met name op basis van schoolsoort de referentiegroepen (waarmee scholen hun verzuimpercentage kunnen vergelijken) samengesteld. De verzuimbenchmark-VO (door Voion) geeft individuele scholen handvatten om het verzuimbeleid verder te verbeteren. De gegevens ter ondersteuning hiervan worden aangeleverd door DUO. Voion en OCW willen nu weten of er meer en/of betere stuurinformatie gehaald kan worden uit de beschikbare databronnen van DUO, waarmee onderwijswerkgevers en/of Voion gericht kunnen sturen op minder verzuim.

Concreet moet het onderzoek het volgende opleveren:

² [DUO \(2020\). Verzuimonderzoek PO en VO 2019.](#)

³ [CBS \(2020\) Ziekteverzuim onder werknemers in 2019 verder toegenomen.](#)

⁴ <https://www.aanpaklerarentekort.nl/regios/voortgezet-onderwijs>

⁵ [Teacher Absenteeism | IIEP Education Policy Toolbox \(unesco.org\)](#)

- 1) Informatie uit de gegevensverzameling van DUO die:
 - a) **inzicht** geeft in de verzuimontwikkeling (patronen) en
 - b) mogelijk **handelingsperspectieven** (voor Voion en/of vo-scholen) biedt die kunnen helpen bij het bepalen van het verzuimbeleid.
- 2) **Aanbevelingen** hoe de huidige gegevensverzameling, de Verzuimrapportage-VO en Verzuimbenchmark-VO verbeterd kunnen worden en wat dit betekent voor de gegevenslevering van DUO (aanvullende gegevens en/of andere opzet).
- 3) Voor zover mogelijk, een **duiding** van de sectorale verzuimontwikkeling in de afgelopen jaren (welke waarneembare factoren hebben deze ontwikkeling beïnvloed en welke (andere) oorzaken zijn hiervoor aan te wijzen?).

1.3 AANPAK VAN HET ONDERZOEK

De aanpak die we volgen is dat we allereerst inventariseren welke gegevens allemaal beschikbaar zijn, wat de kwaliteit van deze gegevens is, en welke informatie hier op dit moment al voor wordt benut. Vervolgens voeren we aanvullende analyses uit om te ontdekken hoe het ziekteverzuim in het voortgezet onderwijs zich heeft ontwikkeld en om te achterhalen wat de belangrijkste factoren in deze ontwikkeling zijn. Dit doen we door een combinatie van beschrijvende analyses, sequentie- en clusteranalyses en regressieanalyses. Op basis van deze analyses krijgen we een beter beeld van de achtergronden van (de stijging van) het ziekteverzuim in het voortgezet onderwijs. Dit kan vervolgens handvatten bieden om het verzuimbeleid te optimaliseren, waarbij we ook aandacht besteden aan de gegevensleveringen vanuit DUO.

2 BESCHRIJVING INHOUD EN KWALITEIT VERZUIMDATA

2.1 INHOUD VAN DE DATA

De data voor dit onderzoek is afkomstig van DUO, die de volgende bestanden heeft overgedragen aan het onderzoeksbureau:

- Meldingsfrequentie 2016-2021;
- Verzuimduur 2016-2021;
- Verzuimpercentage 2016-2021;
- Nulverzuim 2016-2021.

De verschillende bestanden zijn in principe allemaal aparte bewerkingen van dezelfde brongegevens. In het bestand 'Meldingsfrequentie' is een aantal selecties gemaakt (bijvoorbeeld alleen verzuimgevallen die in dat betreffende jaar beginnen worden meegenomen, lopende van het jaar ervoor niet) om zodoende tot een juiste variabele te komen. Voor de analyses in dit onderzoek maken we met name gebruik van het bestand 'Nulverzuim'. Hierin staan alle docenten die in een bepaald jaar actief zijn geweest in het voortgezet onderwijs, dus ook de docenten die in een bepaald jaar geen verzuimgeval hebben gehad. Het bestand bevat gegevens voor elke persoon per maand. De verzuimgevallen in dit bestand zijn echter wel afgekapt op 31 december van een bepaald jaar: ziekteverzuimperiodes die de jaargrens overgaan, zijn verwijderd uit dit bestand. Uiteindelijk willen we voor de patronen voor elke persoon op elk moment weten of diegene wel of geen ziekteverzuim heeft, en dus moeten we ook de gevallen weten die in jaar X starten en in jaar Y verdergaan. Deze gevallen staan wel in het bestand 'Verzuimpercentage' en zijn dus toegevoegd aan het uiteindelijke analysebestand.

In de bestanden zitten de volgende variabelen:

- Bestuur: het schoolbestuur waaronder één of meerdere scholen (BRIN's) vallen;
- BRIN: de Basisregistratie Instellingen, aparte code voor elke unieke school;
- ID: unieke code per werknemer (versleutelde BSN);
- Peilmaand: de maand waarover wordt gerapporteerd;
- Leeftijd;
- Geslacht;
- Functiecategorie: gedetailleerde categorie van functie (18 verschillende categorieën);
- Functiegroep: brede categorie van functie (4 verschillende categorieën);
- Werktijdfactor: dienstverband van de medewerker in termen van fte;
- Salarisschaal en trede;
- Aantal werknemers op de betreffende school;
- Schooltype (havo/vwo/vmbo/pro/combi);
- Denominatie;
- Regio (noord/oost/zuid/west);
- Stedelijkheid locatie school;
- Begin- en einddatum periode van ziekteverlof;
- Omvang van ziekteverlof.

Aan deze dataset hebben wij nog de volgende drie variabelen toegevoegd op BRIN-niveau:

- Percentage leerlingen uit APCG-gebied (Armoedeprobleem-cumulatie-gebieden)⁶;
- Gemiddelde inspectiescore van een school (Geen oordeel/onvoldoende/onvoldoende, tenzij/voldoende);
- Leerling leraar ratio (LLR).

2.2 KWALITEIT VAN DE DATA

Weinig missende waardes variabelen in verzuimdata, wel andere aantallen dan in personeelsbestanden

Over het algemeen ligt de kwaliteit van de data behoorlijk hoog. De meeste variabelen zijn zo goed als volledig gevuld. Het aandeel missende waardes voor de meeste variabelen ligt onder de 1% per jaar. Alleen bij de salarisschalen zien we wat hogere missende waardes: bij ongeveer 1,5% van de personen is deze variabele onbekend.⁷ Als we de aantallen van de ruwe data vergelijken met de aantallen die DUO zelf rapporteert, dan zien we wel een aantal verschillen (zie tabel 2.1). Wat opvalt is dat de aantallen directie hoger liggen in de verzuimbestanden dan in de personeelsbestanden, en de aantallen ondersteunend personeel juist lager. Ook zien we dat de totalen in de verzuimbestanden behoorlijk dalen in 2019 en 2020, terwijl dit in de personeelsbestanden niet het geval is.⁸

De verschillen worden deels verklaard doordat de gegevenslevering vanuit scholen voor personeelsaantallen en verzuimgegevens gescheiden zijn. Hierdoor kan het zijn dat scholen voor bepaalde medewerkers een andere registratie hebben in de verzuimgegevens dan in de personeelsgegevens. Verder ontstaat het verschil in totalen vooral op schoolniveau. Wanneer we het aantal personen per bestuur en brinnummer per jaar op een rij zetten, en dit leggen naast de bestanden van DUO, dan zien we allereerst dat er negen brinnummers wel in de personeelsdata zitten, maar niet in de verzuimdata. Daarnaast zien we dat met name in 2019 en 2020 een aantal brinnummers geen observaties hebben in de verzuimdata, maar wel in de personeelsbestanden.

Tabel 2.1 Aantallen personeelsleden over de tijd naar functiegroep, peilmaand oktober

Jaar	Directie	Onderwijsgevend personeel	Onderwijs ondersteunend personeel	Onbekend	Totaal
<i>Vanuit verzuimbestanden</i>					
2016	3.779	75.043	24.108	875	103.805
2017	3.853	76.639	24.513	102	105.107
2018	3.683	75.182	24.338	141	103.344
2019	3.573	72.365	23.750	105	99.793
2020	3.579	70.947	23.688	72	98.286
2021	3.678	73.909	25.445	44	103.076
<i>Vanuit personeelsbestanden⁹</i>					
2016	3.036	75.958	26.619	964	106.578

⁶ Deze indicator is afkomstig van [de Onderwijsinspectie](#).

⁷ Om een idee te krijgen van de hoogte van het salaris zijn de salarisschalen en tredes gekoppeld aan de salaristabellen VO.

⁸ Dit kan te maken hebben met de coronacrisis, waardoor meer scholen moeite hadden om de verzuimgegevens aan te leveren. (De gegevens over 2019 worden pas in 2020 geleverd).

⁹ https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/personeel/in-aantal-personen.jsp

Jaar	Directie	Onderwijsgevend personeel	Onderwijs ondersteunend personeel	Onbekend	Totaal
2017	2.905	76.406	26.892	107	106.310
2018	2.916	76.191	27.017	147	106.271
2019	2.886	75.809	27.380	129	106.204
2020	2.807	75.117	27.791	121	105.836
2021	2.830	76.511	29.340	106	108.787

Bron: Bewerking SEOR op basis van DUO-data

Aantal oneffenheden in data gevonden

De data is vervolgens ook gecontroleerd op onrealistische waardes. Daarbij kwam een aantal onlogische zaken naar voren. Zo zien we dat de werktijdfactor een minimum heeft van -2,4 en een maximum van 1,8. Dit kan bijvoorbeeld komen doordat er correcties worden gemaakt op eerdere maanden waarbij er met een te hoge of te lage werktijdfactor is gewerkt. Omdat niet volledig is te achterhalen hoe deze waardes tot stand zijn gekomen, selecteren we alle waardes tussen 0 en 1,2. Dit betreft een verlies van 0,4% van het aantal observaties. Daarnaast loopt de leeftijdsverdeling van 13 tot en met 95, wat wel zeer extreme waardes zijn. Het gaat hier zeer waarschijnlijk om invoerfouten vanuit de scholen. Omdat niet geheel duidelijk is welke waardes er wel zouden moeten staan, en omdat het maar om een zeer klein aantal gaat waarbij dit het geval is, kiezen we ervoor om een selectie te maken van personen tussen 18 en 75 jaar. We verliezen hierbij 0,04% van de observaties. Als laatste zien we nog dat de omvang van het verlof in sommige gevallen hoger ligt dan de werktijdfactor. Omdat de omvang van het verlof in principe aangeeft welk deel van de werktijd iemand verlof heeft, zou dit nooit hoger kunnen zijn dan de werktijdfactor. Daarom passen we bij deze gevallen de omvang van het verlof aan naar het maximum (zijnde de werktijdfactor). Dit betreft een aanpassing voor 0.8% van het totaal aantal observaties.

Naast bovengenoemde extreme waarden in de dataset, zijn we ook nog een aantal fouten in de data tegengekomen. Zo zagen we dat op een bepaalde school de volgnummers van alle medewerkers in oktober 2017 ineens allemaal veranderden. Dit leidde ertoe dat er in oktober 2017 dubbele waarnemingen waren voor alle medewerkers van deze school. In de analyse van de patronen zijn deze dubbele waarnemingen allemaal uit de dataset gehaald. Ook komt het vaak voor dat personen meerdere dienstverbanden hebben binnen eenzelfde school of op verschillende scholen. Dit is niet direct een fout, maar personen hebben dan wel meerdere waarnemingen binnen dezelfde maand. Wanneer personen op meerdere scholen werken, hebben we de school gekozen waar deze persoon het grootste deel van de tijd werkt. Bij meerdere dienstverbanden (volgnummers) binnen een school worden de werktijdfactoren en omvang van het verlof bij elkaar opgeteld. Echter zorgde dit in enkele gevallen voor zeer extreme waardes van de werktijdfactor, wat ook zeer waarschijnlijk komt door administratieve fouten. Deze werktijdfactoren zijn aangepast naar de meest voorkomende werktijdfactor in een bepaalde maand.

In onderstaande tabel is een overzicht gegeven van alle oneffenheden in de data die wij in dit onderzoek tegen zijn gekomen.

Tabel 2.2 Oneffenheden verzuimdata

Type	Toelichting	Toegepaste correctie	Deel van observaties
Onrealistische waarde werktijdfactor	Werktijdfactor loopt van -2,4 tot 1,8	Werktijdfactor van 0 tot 1,2	0,4%
Onrealistische waardes leeftijden	Leeftijd loopt van 13 tot 95	Leeftijd van 18 tot 75 jaar	0,04%
Omvang verlof groter dan omvang dienstverband	Omvang verlof zou in theorie altijd lager moeten liggen dan de werktijdfactor van een persoon (een persoon kan niet langer ziek zijn dan hoeveel diegene werkt)	Omvang verlof maximaal gelijk aan de volledige werktijdfactor	0,8%

Type	Toelichting	Toegepaste correctie	Deel van observaties
Volgnummer veranderd zonder dat andere waardes veranderen	In sommige gevallen krijgen werknemers een ander volgnummer zonder dat andere variabelen zoals functiecategorie en salaris veranderen	Waarnemingen ontdebeld	0,8%
Dienstverbanden op meerdere scholen	Dit is niet direct een fout, maar hier moet wel rekening mee worden gehouden bij de berekening van verlofstatistieken: personen kunnen meerdere dienstverbanden hebben op meerdere scholen, waardoor verlof dubbel geteld zou kunnen worden.	Gekozen voor de school waar een persoon het grootste deel van de tijd werkt	0,6%
Meerdere dienstverbanden binnen een school	Ook dit is niet direct een fout, maar ook hiermee moet rekening worden gehouden. Wel zie je dat de optelling van de werktijdfactoren bij personen met meerdere dienstverbanden binnen een school in sommige gevallen leidt tpt extreme waardes.	Aangepast naar meest voorkomende werktijdfactor in een bepaalde maand	2%
Ziekmelding bij ene dienstverband maar niet bij andere dienstverband	Het komt ook voor dat personen binnen eenzelfde maand wel een ziekmelding hebben bij het ene, maar niet bij het andere dienstverband. De vraag is of dit altijd juist is, of dat de ziekmelding voor beide gevallen zou moeten gelden.	Geen aanpassing toegepast.	2,2%

Toelichting: de kolom 'deel van observaties' geeft het deel van alle waarnemingen aan dat van toepassing is op het criterium. Het gaat hier niet om het aantal personen, maar om het aantal observaties.

In bijlage I is een tabel opgenomen met beschrijvende informatie over een groot deel van de variabelen in de dataset per jaar.

3 PATRONEN IN ZIEKTEVERZUIM

3.1 SEQUENTIEANALYSE

Met behulp van een sequentieanalyse worden de patronen in het ziekteverzuim van werknemers in het voortgezet onderwijs in kaart gebracht. Naast de genoemde correcties uit tabel 2.2 in het vorige hoofdstuk, selecteren we hier verder een groep die sinds 1 januari 2016 in de data te vinden is en ook de hele periode gevolgd kan worden. Deze selectie wordt gemaakt om ervoor te zorgen dat we goede vergelijkingen tussen de personen kunnen maken. In totaal gaat het om een groep van 69.613 personen (ongeveer 45% van het totaal aantal unieke personen in de dataset). Van deze personen zetten we per week het ziekteverzuim op een rij over de periode 2016-2019¹⁰. We hebben dus in totaal $69.613 \cdot 208 =$ ruim 14 miljoen observaties.

Het ziekteverzuim delen we op in vier categorieën¹¹:

- Geen verzuim;
- Kort verzuim (aansluitende verzuimperiode korter dan 7 dagen);
- Middellang verzuim (aansluitende verzuimperiode van 7 dagen tot en met 41 dagen);
- Lang verzuim (aansluitende verzuimperiode van 42 dagen of langer).

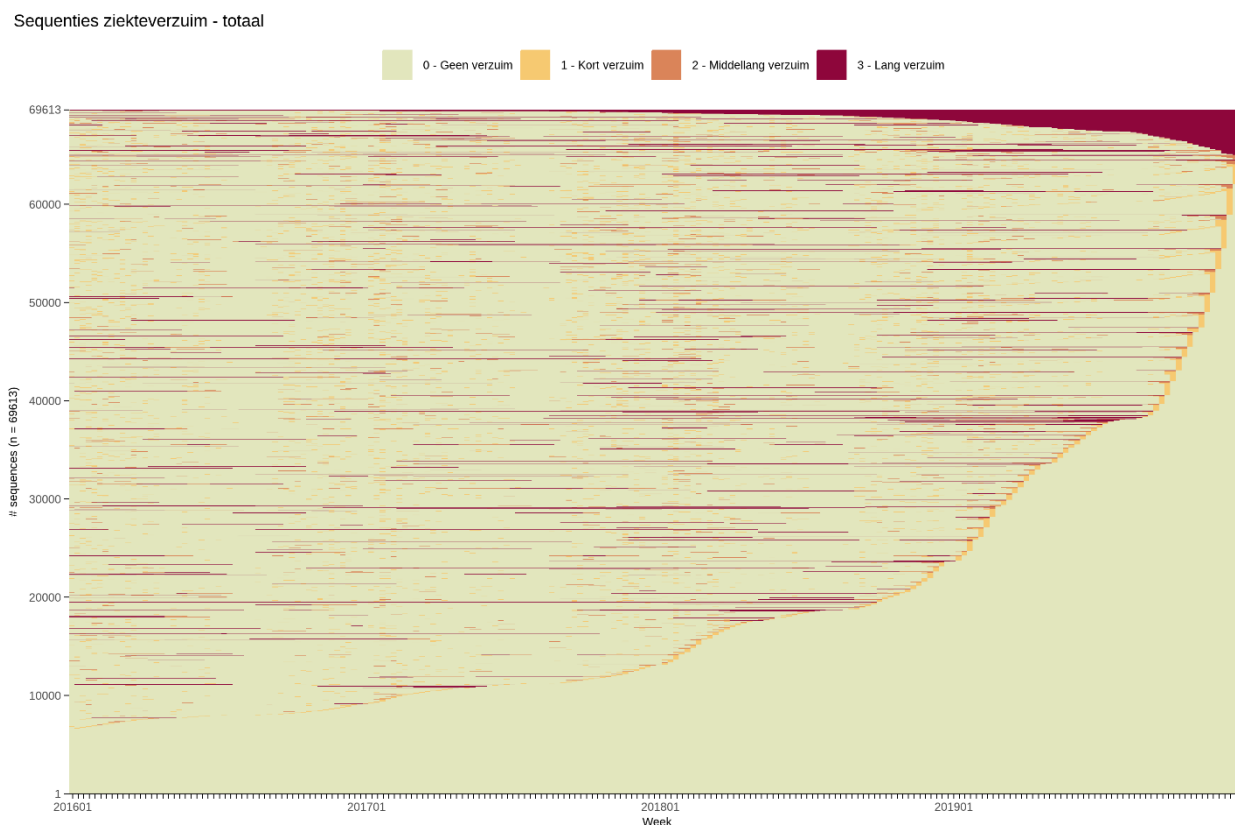
Er is ook gekeken naar de verdeling van de verzuimdagen binnen deze categorieën, en of een andere indeling andere resultaten oplevert in de sequentieanalyse. Dit laatste blijkt niet het geval te zijn. Zie hiervoor bijlage I.

Per week hebben we dus een indicatie van de 'status' van een persoon. Deze 'statussen' worden in de sequentieanalyse voor elke persoon allemaal achter elkaar gezet en geordend. Visueel is dit weergegeven in figuur 3.1. Deze figuur laat in feite de hele dataset in één plaatje zien. Elke lijn van links naar rechts representeert 1 uniek persoon. Onderaan de grafiek zie je personen die helemaal nooit verzuim hebben gehad van 2016 tot 2019, bovenaan de grafiek zie je personen die eind 2019 een lange verzuimperiode hadden. Uiteraard is het nog erg lastig om hier iets over de patronen te concluderen. Daarom maken we op basis van deze sequenties groepen personen die vergelijkbare patronen doorlopen.

¹⁰ Vanwege de coronacrisis laten we in de analyse van patronen in het ziekteverzuim de jaren 2020 en 2021 buiten beschouwing.

¹¹ Categorieën gebaseerd op Vroonhof (2020).

Figuur 3.1 Grafiek sequentieanalyse ziekteverzuim



Bron: Bewerking SEOR op basis van DUO-data

3.2 CLUSTERS OP BASIS VAN VERGELIJKBARE PATRONEN

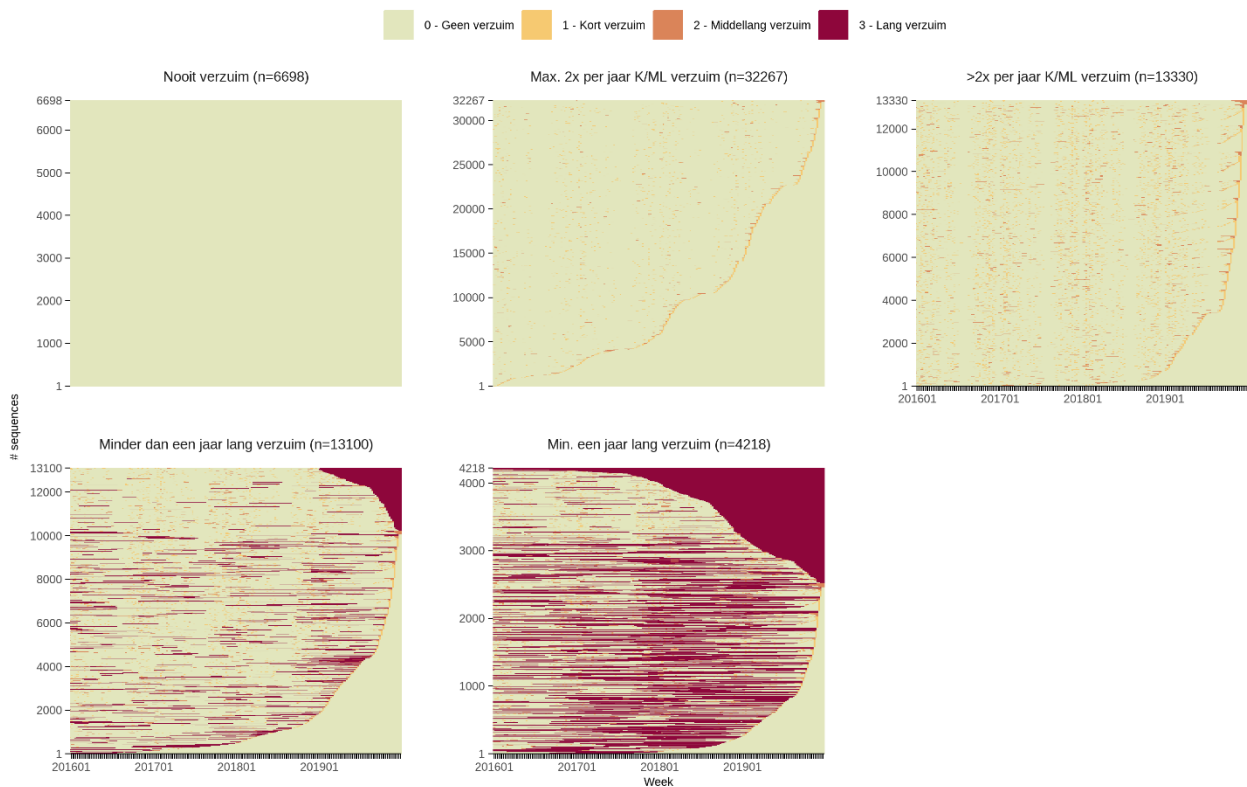
In eerste instantie laten we een model de vergelijkbare clusters bepalen. Het model zoekt voor elke sequentie (elke persoon) de meest vergelijkbare andere sequentie. Vervolgens maakt het clusters waarbij de verschillen tussen de sequenties *binnen* een cluster zo gering mogelijk zijn, terwijl de verschillen tussen de sequenties *tussen* de clusters zo groot mogelijk zijn. Uit de berekeningen van het model maken wij vervolgens zelf beslisregels om logische clusters te vormen. Dit doen we door enerzijds visueel te kijken welke groepen er uit het model gevormd worden, en anderzijds door naar statistieken te kijken als gemiddeldes, minima en maxima over het aantal verzuimperioes per groep. Uiteindelijk maken we de volgende vijf groepen:

1. Nooit verzuim: deze groep heeft tussen 2016 en 2019 geen enkele keer ziekteverzuim gehad. (6.698 personen, 10% van het totaal).
2. Maximaal 2x per jaar kort of middellang verzuim: deze groep heeft maximaal 2 verzuimperioes (ofwel een kort ofwel een middellange verzuimperioe) per jaar. Deze groep heeft nooit lang verzuim gehad. (32.267 personen, 46% van het totaal).
3. Meer dan 2x per jaar kort of middellang verzuim: deze groep heeft meer dan 2 verzuimperioes (ofwel kort ofwel middellange verzuimperioes) per jaar. Deze groep heeft nooit lang verzuim gehad. (13.330 personen, 19% van het totaal)

4. Minder dan een jaar lang verzuim: deze groep heeft ooit een lange verzuimperiode gehad in de periode 2016-2019, maar in totaal niet langer dan een jaar. (13.100 personen, 19% van het totaal)
5. Minimaal een jaar lang verzuim: deze groep heeft ooit een lange verzuimperiode (meer dan 41 dagen aaneengesloten) gehad in de periode 2016-2019, en in totaal een jaar of langer lang verzuim gehad. (4.218 personen, 6% van het totaal).

Figuur 3.2 Clustering ziekteverzuim sequenties

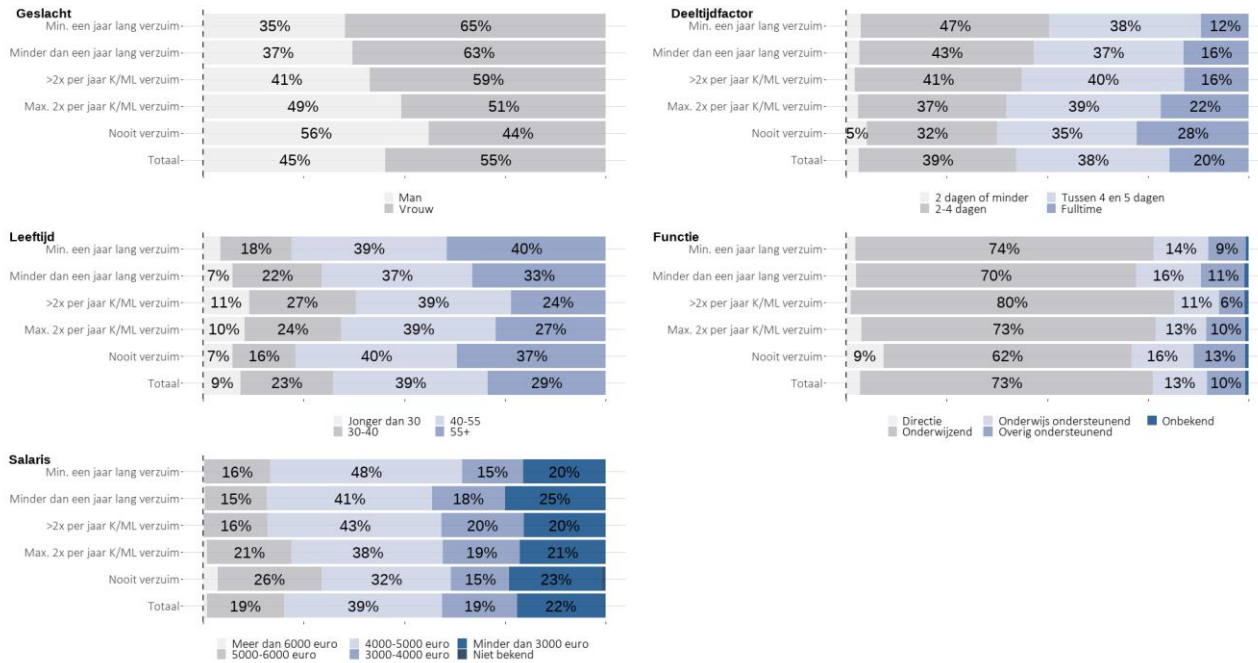
Sequenties ziekteverzuim - per cluster



Bron: *Bewerking SEOR op basis van DUO-data*

Bovenstaande gegevens laten zien dat voor de grootste groep (bijna 50%) geldt dat ze minder dan twee keer per jaar kort of middellang verzuim hebben gehad. Toch heeft in totaal 25% van de populatie ooit in de periode 2016 tot en met 2019 een keer lang verzuim (meer dan 41 dagen aaneengesloten) gehad (groepen 4 en 5 bij elkaar opgeteld). Het is vervolgens interessant om te onderzoeken welke personen in welke groepen zitten. We doen dit allereerst beschrijvend. Figuur 3.3 en 3.4 geven respectievelijk de verdeling van persoonskenmerken en schoolkenmerken over de clusters weer. In bijlage I zijn deze figuren ook omgekeerd weergegeven: de verdeling van de clusters over de persoons- en schoolkenmerken.

Figuur 3.3 Verdeling persoonskenmerken over de clusters



Bron: *Bewerking SEOR op basis van DUO-data Toelichting: de categorieën van continue variabelen zijn elkaar uitsluitende categorieën.*

Vrouwen, parttimers en 55+’ers relatief vaak in lang verzuim clusters

Qua persoonskenmerken zien we een aantal vrij duidelijke verschillen tussen de groepen. In de groep ‘nooit verzuim’ zitten bijvoorbeeld relatief veel mannen, terwijl in de groep ‘langer dan een jaar lang verzuim’ juist relatief veel vrouwen zitten. Ook zien we dat personen die fulltime werken minder snel in de lang verzuim clusters vallen, en vaker in het cluster ‘nooit verzuim’. Personen die 2-4 dagen werken vallen juist relatief vaak in het cluster dat langer dan een jaar lang verzuim heeft.

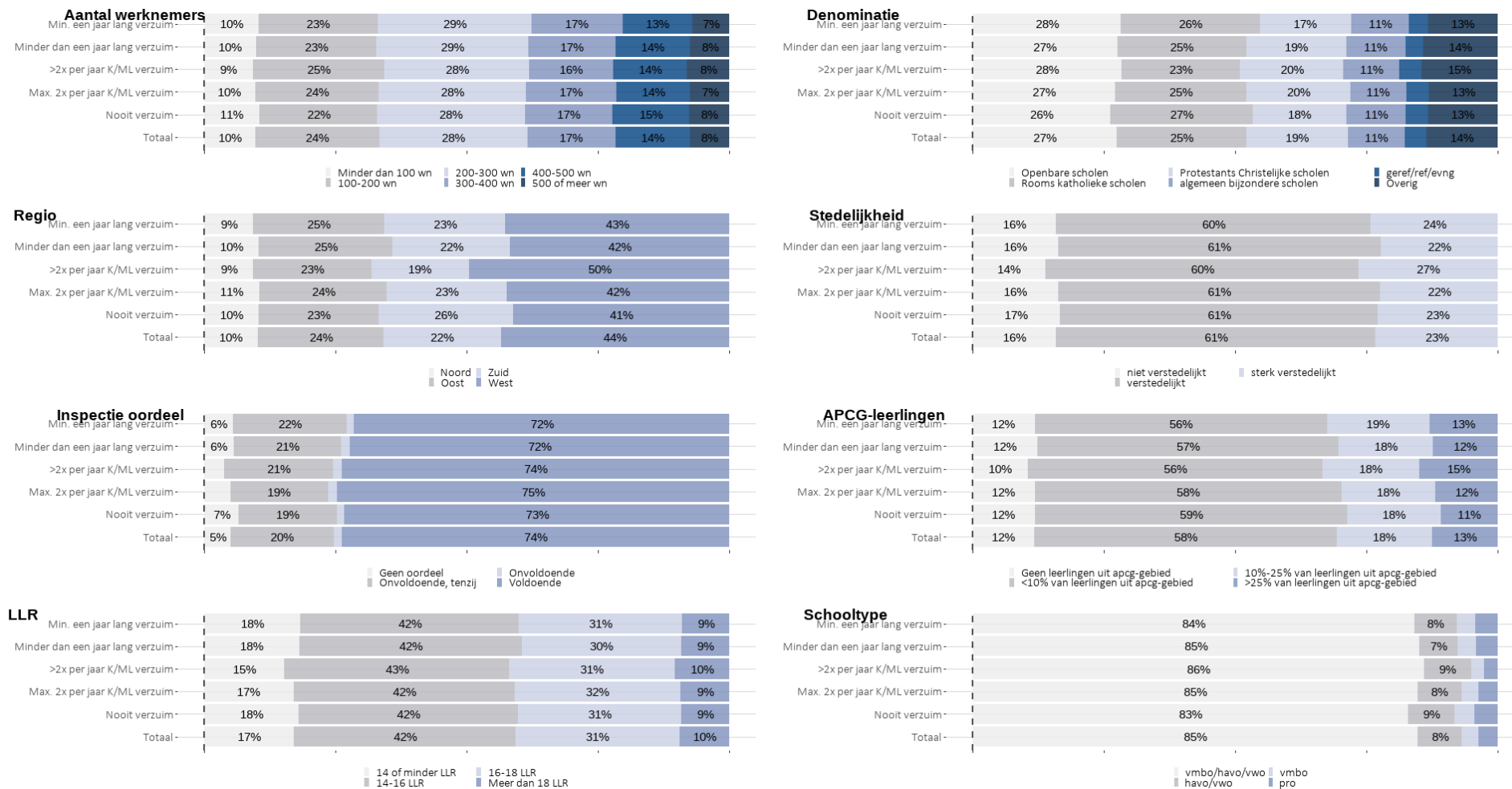
Ook bij de leeftijdsverdeling zien we een aantal verschillen. 55+’ers vallen zowel vaker in het cluster ‘nooit verzuim’ als het cluster ‘langer dan een jaar lang verzuim’. Personen tot 40 jaar lijken wat vaker in het cluster ‘meer dan 2 keer per jaar kort/middellang verzuim’ te vallen. Bij de functiegroepen zien we met name dat de directie vaker in het cluster ‘nooit verzuim’ valt. Ook valt op dat het onderwijzende personeel relatief wat vaker in het cluster ‘meer dan 2 keer per jaar kort/middellang verzuim’ zit.

Als laatste zien we dat personen met een gemiddeld salaris boven de 5000 euro vaker in het cluster ‘nooit verzuim’ vallen, en dat personen met een salaris tussen 4000 en 5000 euro vaker in het cluster langer dan een jaar lang verzuim zitten. De eerste observatie is duidelijk gecorreleerd met de functie: directie zit vaker in het ‘nooit verzuim’ cluster en valt ook in de hogere salarisschalen.

Weinig verschillen in schoolkenmerken, cluster >2 keer per jaar kort/middellang verzuim wel afwijkend

Over het algemeen zien we weinig verschillen naar schoolkenmerken tussen de clusters. De meeste kenmerken hebben een gelijke verdeling in elk van de clusters. Alleen bij het cluster ‘meer dan 2 keer per jaar kort of middellang verzuim’ zien we een aantal afwijkingen. Met name onderwijspersoneel op scholen in het westen van het land en in sterk verstedelijkte gebieden vallen relatief vaak in dit cluster.

Figuur 3.4 Verdeling schoolkenmerken over de clusters



Bron: Bewerking SEOR op basis van DUO-data. Toelichting: de categorieën van continue variabelen zijn elkaar uitsluitende categorieën.

Bovenstaande figuren laten puur een beschrijving zien van de kenmerken verdeeld over de clusters. Door middel van een verklarende analyse kijken we verder welke type personen een grotere kans hebben om in een bepaald cluster te komen. We gebruiken hiervoor een multinomiaal logit model. We zien de volgende uitkomsten:

Kans op cluster 'Nooit verzuim'

- Vrouwen kleinere kans dan mannen;
- Onderwijzend personeel kleinere kans dan overig personeel;
- Parttimers kleinere kans dan fulltimers, behalve personen die minder dan 2 dagen werken;
- Personen met hoger salaris grotere kans.

Kans op cluster 'Maximaal 2 keer per jaar kort/middellang verzuim'

- Vrouwen kleinere kans dan mannen;
- Hogere leeftijden kleinere kans;
- Onderwijzend personeel kleinere kans dan overig personeel;
- Parttimers kleinere kans dan fulltimers, behalve personen die minder dan 2 dagen werken;
- Personen met hoger salaris grotere kans;
- Personen op scholen in het westen van het land kleinere kans dan in het noorden van het land;
- Personen op scholen met een onvoldoende van de inspectie kleinere kans dan scholen met een voldoende.

Kans op cluster 'Meer dan 2 keer per jaar kort/middellang verzuim'

- Vrouwen grotere kans dan mannen;
- 30-40'ers grotere kans, 55+'ers kleinere kans dan personen jonger dan 30;
- Onderwijzend personeel grotere kans dan overig personeel;
- Parttimers grotere kans dan fulltimers, behalve personen die minder dan 2 dagen werken;
- Personen met hoger salaris kleinere kans;
- Personen op scholen in het westen van het land grotere kans dan in het noorden;
- Personen op scholen met een hogere LLR een grotere kans.

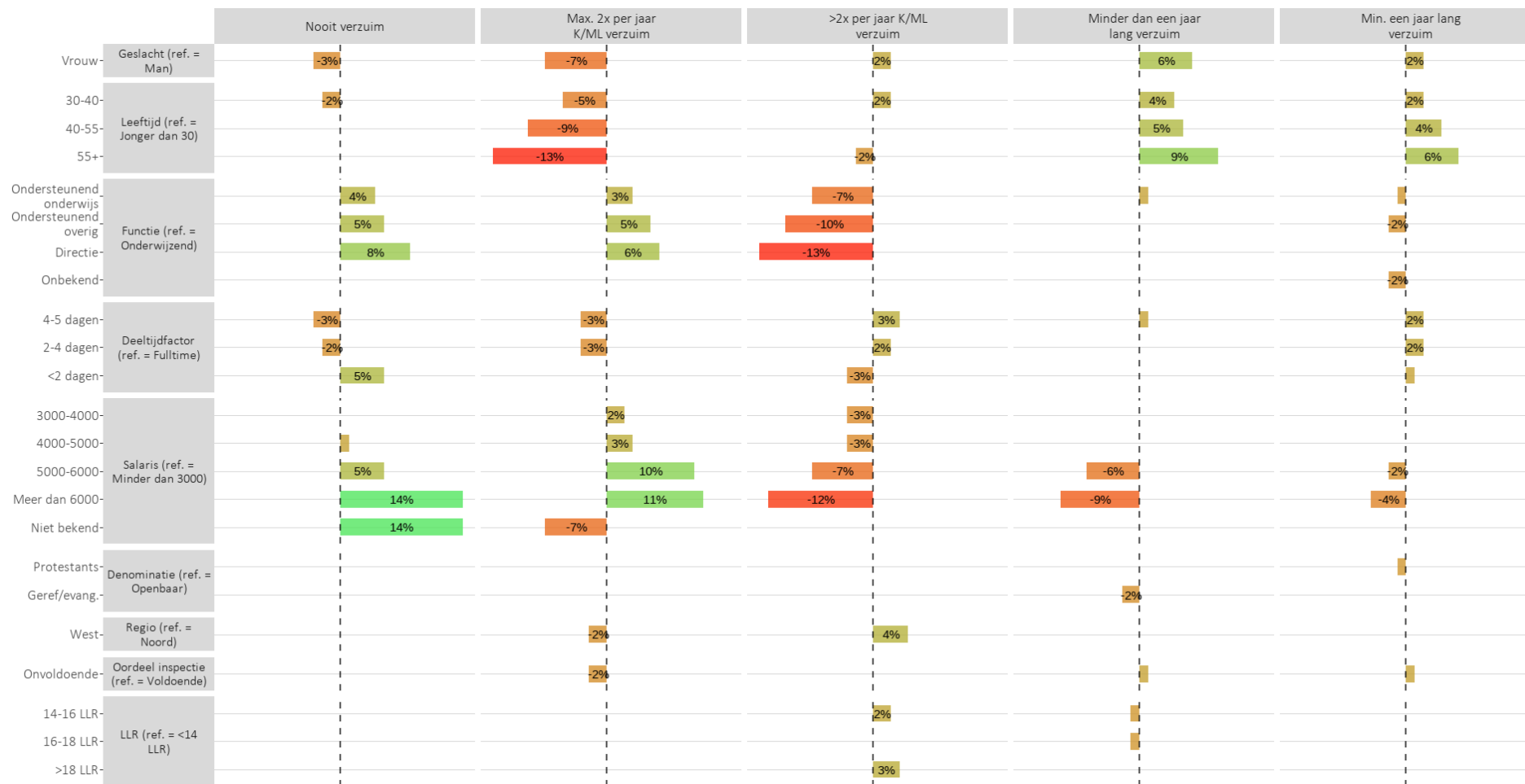
Kans op cluster 'Minder dan een jaar lang verzuim'

- Vrouwen grotere kans dan mannen;
- Hogere leeftijden grotere kans;
- Personen met hoger salaris kleinere kans;
- Personen op scholen met gereformeerd of evangelische denominatie kleinere kans dan openbare scholen;
- Personen op scholen met een onvoldoende van de inspectie een hogere kans dan scholen met een voldoende;
- Personen op scholen met een LLR tussen 14 en 18 een kleinere kans dan scholen met LLR onder de 14.

Kans op cluster 'Minimaal een jaar lang verzuim'

- Vrouwen grotere kans dan mannen;
- Hogere leeftijden grotere kans;
- Onderwijzend personeel grotere kans dan overig personeel;
- Parttimers grotere kans dan fulltimers;
- Personen met hoger salaris kleinere kans;
- Personen op scholen met een onvoldoende van de inspectie een hogere kans dan scholen met een voldoende.

Figuur 3.5 Kans om in een bepaald cluster te komen op basis van achtergrondkenmerken



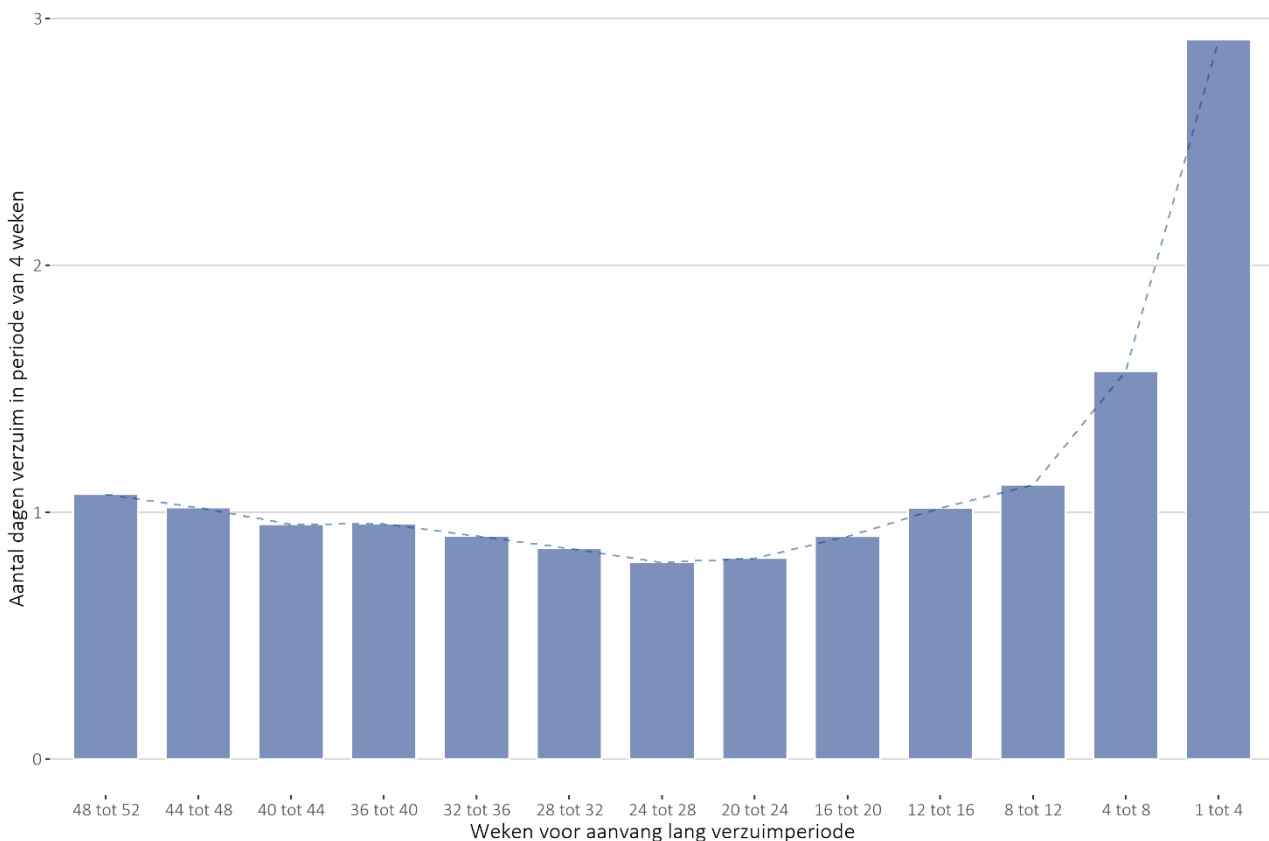
Bron: Bewerking SEOR op basis van DUO data. Toelichting: de figuur toont de effecten die significant zijn op het 5%-niveau.

Verzuim loopt gemiddeld genomen op richting lange verzuimperiode

We zoomen nog iets verder in op de sequenties van de groepen die ooit lang verzuim (meer dan 41 dagen aaneengesloten) hebben gehad. Hebben personen voorafgaand aan een lange verzuimperiode bijvoorbeeld eerst meer kort of middellang verzuim? In de sequentieanalyses kunnen we onder andere op zoek naar ‘representatieve’ en veel voorkomende sequenties binnen de clusters. Deze statistieken geven in dit geval helaas weinig waardevolle informatie: er zijn geen hele duidelijke patronen in kort of middellang verzuim voorafgaand aan een lang verzuimperiode te ontdekken.

Wel zien we dat voorafgaand aan de eerste lange verzuimperiode van een persoon, het aantal dagen verzuim gemiddeld genomen toeneemt naarmate de lange verzuimperiode dichterbij komt. Figuur 3.5 geeft deze ontwikkeling weer. Voor elke persoon kijken we steeds een jaar voor de aanvang van de eerste lange verzuimperiode naar het aantal dagen ziekteverlof in een periode van 4 weken. Aan de linkerkant van de grafiek is het gemiddeld aantal ziekteverlofdagen 52-48 weken voor aanvang van de lange verzuimperiode te zien, en dit loopt steeds af tot en met 1-4 weken voor aanvang van de lange verzuimperiode. Het gemiddeld aantal dagen verzuim in een periode van 4 weken stijgt van ongeveer 1 naar bijna 3 dagen. Vanaf 20 weken voor aanvang van de lange verzuim periode zien we het aantal verlofdagen gestaag oplopen, waarbij de piek duidelijk in de laatste vier weken voor aanvang van de lange verzuimperiode is te zien.

Figuur 3.6 Ontwikkeling gemiddeld aantal dagen verzuim in periode van vier weken voor start lang verzuim



Bron: Bewerking SEOR op basis van DUO data

4 VERKLARENDE ANALYSE ZIEKTEVERZUIM

4.1 ANALYSES OP SCHOOLNIVEAU

In het voorgaande hoofdstuk zijn individuele patronen van personen bekeken. In dit hoofdstuk kijken we naar het schoolniveau. Met de regressieanalyse op schoolniveau dienen we twee doelen:

- a) verklaren van de ontwikkeling van het verzuimpercentage (welke factoren hebben het meest invloed op deze ontwikkeling);
- b) verbeteren van de verzuimbenchmark per school (hoe goed scoort een school ten opzichte van het te verwachten verzuimpercentage, gegeven een aantal – niet te beïnvloeden – factoren).

De belangrijkste te verklaren variabele in deze analyses is het verzuimpercentage per school per jaar. Er zijn twee categorieën verklarende variabelen: variabelen over personen en variabelen over de school. De variabelen over personen zijn als aandelen of gemiddelden per school in het model opgenomen.

Naast het verklaren van de (totale) verzuimpercentages, voeren we ook aparte analyses uit naar kort, middellang en langdurig verzuim, om te onderzoeken in hoeverre de verschillende verzuimduren andere oorzaken hebben. Dit doen we door steeds het aandeel korte/middellange/langdurige verzuimgevallen in het totaal aan verzuimgevallen binnen een school te berekenen, en dat als te verklaren variabele op te nemen in het model.

We hebben in deze analyse te maken met zogeheten ‘panel data’, d.w.z. meerdere observaties per school over de tijd. Dit zorgt ervoor dat we geen ‘simpele’ Ordinary Least Squares (OLS) regressie uit kunnen voeren, omdat de observaties van een school over de tijd niet onafhankelijk zijn van elkaar (wat wel een voorwaarde is binnen de OLS-regressie). Om hiervoor te corrigeren maken we gebruik van een random-effects model. Formeel ziet het model dat we zullen schatten er als volgt uit:

$$\%Verzuim_{it} = \alpha_0 + \beta_1 Persoonskenmerken_{it} + \beta_2 Schoolkenmerken_{it} + \mu_i + \varepsilon_{it}$$

Waarbij i staat voor de school en t voor de tijdsperiode (de μ_i zijn de random-effects).

4.1.1 Verklaring verzuimpercentage

Tabel 4.1 vat samen welke kenmerken (significant) van invloed zijn op het totale verzuimpercentage en de verzuimpercentages gebaseerd op kort, middellang of lang verzuim. Dit is gebaseerd op de uitkomsten van de regressieanalyses, welke te zien zijn in tabel 4.2. Het totale verzuimpercentage wordt dus beïnvloed door de volgende schoolkenmerken:

- Schooltype: scholen van het type pro, vmbo, vmbo-t/havo/vwo en vmbo/havo/vwo hebben gemiddeld een hoger verzuim dan havo/vwo-scholen;
- Denominatie: Protestants Christelijke scholen hebben gemiddeld een lager verzuim dan openbare scholen;
- Verstedelijking: niet-verstedelijkte gebieden hebben gemiddeld een lager verzuimpercentage dan sterk verstedelijkte gebieden;
- Aantal werknemers: scholen met meer werknemers hebben gemiddeld lager verzuim;
- Leerling-leraar ratio: scholen met een hogere leerling-leraar ratio hebben gemiddeld een lager verzuim.

Daarnaast hebben de volgende persoonskenmerken een significante invloed op het verzuimpercentage:

- Leeftijd: scholen met een hoger percentage werknemers boven de 55 jaar hebben gemiddeld een hoger verzuim;
- Werktijdfactor: scholen met een gemiddelde werktijdfactor die hoger ligt dan het eigen langjarige gemiddelde hebben gemiddeld een lager verzuim.

Een aantal dingen valt hierbij op:

- a) Grotere scholen met meer werknemers hebben gemiddeld een lager verzuim. Het kan zijn dat grotere scholen vaker een beter beleid hebben op het gebied van verzuimpreventie. Dit is echter niet uit deze data te halen.
- b) Het effect van de leerling-leraar ratio (LLR) zou je eerder andersom verwachten, namelijk dat grotere klassen meer werkdruk met zich meebrengen en dus tot meer verzuim leiden. Uit een nadere inspectie van de relatie tussen LLR en verzuim blijkt dat het negatieve effect veroorzaakt wordt door het feit dat scholen met een lage LLR (onder de 10) gemiddeld een hoger verzuim kennen dan scholen met een hogere LLR (boven de 10). Het merendeel van de scholen met een LLR onder de 10 betreft pro-scholen. Echter verklaart het schooltype niet het hele effect: ook binnen pro-scholen zien we nog een negatief verband tussen LLR en verzuim. Het is aannemelijk dat het hebben van (zeer) kleinere klassen, met name iets zegt over de problematiek van de leerling populatie, wat de functie voor een docent zwaarder kan maken en daardoor leidt tot meer verzuim. Het verband dat grotere klassen tot meer werkdruk en dus tot meer verzuim leiden wordt niet gevonden in de data, ook niet als we alleen kijken naar scholen met een LLR boven de 10, of binnen verschillende schooltypes.
- c) Bij het gevonden effect van de werktijdfactor (in afwijking van het gemiddelde) ligt de causaliteit mogelijk andersom. Een toename van verzuim, bijvoorbeeld door verslechtering van gezondheid, leidt dan (uiteindelijk) tot een daling in de werktijdfactor. Het langjarige gemiddelde van de werktijdfactor van een school heeft wel een positief effect op verzuim, waarbij een hogere werktijdfactor dus leidt tot meer verzuim.

Tabel 4.1 Invloed van school- en persoonskenmerken op verzuimpercentages*

	Verzuim	Kort verzuim	Middellang verzuim	Lang verzuim
Schoolkenmerken				
<i>Schooltype</i>	x	x	x	x
<i>Denominatie</i>	x	x		x
<i>Regio</i>		x	x	
<i>Verstedelijking</i>	x	x		
<i>Aantal werknemers</i>	x	x		x
<i>Leerling-leraar ratio</i>	x			x
<i>APC-gebied**</i>				x
Persoonskenmerken				
<i>Leeftijd</i>	x		x	x
<i>Functie</i>			x	x
<i>Salarisschaal</i>		x	x	
<i>Werktijdfactor***</i>	x	x	x	

Bron: *Bewerking SEOR op basis van DUO-data.*

* Een 'x' geeft aan dat het betreffende kenmerk een significant (op 5%-niveau) effect heeft op het betreffende verzuimpercentage. ** Armoedeprobleemcumulatiegebied.

*** In afwijking van het gemiddelde over de periode 2016-2021.

Kenmerken hebben verschillend effect op kort, middellang en lang verzuim

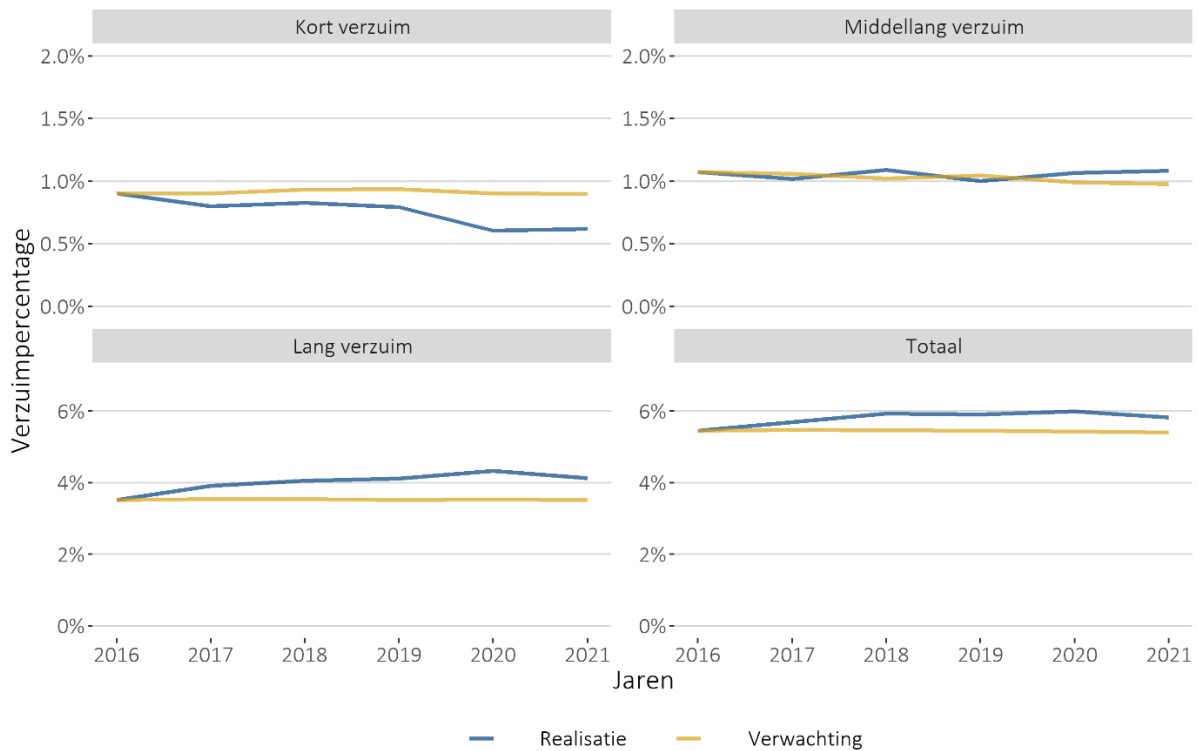
De invloed van de verschillende school- en persoonskenmerken op de afzonderlijke verzuimpercentages (kort, middellang en lang; op basis van de duur van de verzuimperiodes) is verschillend. Sommige kenmerken, die geen invloed hebben op het totale verzuimpercentage, hebben wel invloed op één of twee van de afzonderlijke verzuimpercentages. Regio heeft bijvoorbeeld alleen invloed op kort en middellang verzuim en het percentage leerlingen uit een APC-gebied heeft juist alleen invloed op lang verzuim. De persoonskenmerken functie en salarisschaal blijken vooral invloed te hebben op het middellange verzuim. Alleen het schoolkenmerk schooltype is op alle verzuimcategorieën van invloed.

Verklaringskracht ligt over het algemeen vrij laag

Figuur 4.1 laat de ontwikkelingen zien in de verzuimpercentages. Naast de realisaties (blauwe lijn) zijn hierin ook de ontwikkelingen in de verzuimpercentages te zien die verklaard worden door de ontwikkelingen in de schoolkenmerken en de persoonskenmerken (dus afgezien van de autonome ontwikkelingen in verzuim, in het model weergegeven door de zogenoemde jaardummies) (gele lijn). We zien dat het totale verzuimpercentage in 2017 en 2018 stijgt en vanaf 2018 redelijk constant blijft. De stijging wordt met name veroorzaakt door de stijging van het lange verzuim. Sowieso levert het lange verzuim veruit de grootste bijdrage aan het totale verzuimpercentage op scholen. We zien dat het korte verzuim in de afgelopen jaren juist gedaald is. Het middellange verzuimpercentage schommelt per jaar iets.

Op basis van de ontwikkelingen in de school- en persoonskenmerken zouden de verzuimpercentages weinig veranderen over de tijd. De veranderingen over de tijd in de mate van verzuim komen dus vooral door andere ontwikkelingen die niet door de beschikbare kenmerken verklaard kunnen worden.

Figuur 4.1 Ontwikkeling verzuimpercentages realisatie versus verwachting op basis van kenmerken



Bron: Bewerking SEOR op basis van DUO-data. Toelichting: de assen van de bovenste twee figuren zijn anders dan de onderste twee figuren voor het bevorderen van de leesbaarheid van de figuur.

Ook op basis van de R-kwadraat maatstaven (onderaan tabel 4.2) kunnen we concluderen dat de verklaaringskracht van het model aan de lage kant is. We zien wel dat het korte verzuimpercentage beter verklaard wordt door de beschikbare kenmerken dan de andere verzuimpercentages.

Tabel 4.2 Uitkomsten regressies (random effects)

Variabelen	Verzuimpercentage			
	Totaal	Kort verzuim	Middellang verzuim	Lang verzuim
Schooltype (ref = havo/vwo)				
pro	0.955*** (0.338)	-0.121*** (0.0357)	0.254*** (0.0919)	0.639** (0.295)
vmbo	0.648** (0.321)	-0.0385 (0.0381)	0.0886 (0.0862)	0.448 (0.316)
vmbo-t	1.104 (0.771)	-0.193*** (0.0593)	-0.0335 (0.121)	1.665*** (0.625)
vmbo-t/havo/vwo	0.505** (0.211)	0.0589** (0.0264)	0.0959* (0.0554)	0.275 (0.193)
vmbo/havo/vwo	1.011*** (0.210)	0.0456* (0.0244)	0.182*** (0.0555)	0.685*** (0.186)
Denominatie (ref = openbare scholen)				
Rooms katholieke scholen	-0.256 (0.173)	-0.0521** (0.0241)	-0.00204 (0.0523)	-0.166 (0.151)
Protestants Christelijke scholen	-0.376** (0.184)	-0.0294 (0.0230)	-0.0210 (0.0492)	-0.391** (0.162)
algemeen bijzondere scholen	0.520* (0.302)	-0.0131 (0.0301)	0.0501 (0.0626)	0.439* (0.264)
geref/ref/evng	-0.558 (0.452)	-0.0899* (0.0514)	-0.0854 (0.0920)	-0.402 (0.386)
islamitisch	-0.101 (2.037)	0.128 (0.596)	0.379 (0.438)	-0.651 (1.015)
vrije scholen	0.277 (0.270)	0.0727 (0.107)	0.323* (0.170)	-0.171 (0.346)
Overig	-0.362* (0.209)	-0.00917 (0.0269)	0.0665 (0.0614)	-0.339* (0.181)
Regio (ref = noord)				
oost		0.0575** (0.0259)	0.191*** (0.0619)	
west		0.117*** (0.0280)	0.190*** (0.0623)	
zuid		-0.0349 (0.0288)	0.0731 (0.0696)	
Verstedelijking (ref = sterk verstedelijkt)				
Verstedelijkt	-0.290 (0.183)	-0.0429** (0.0194)	-0.0856* (0.0458)	
niet verstedelijkt	-0.465** (0.230)	-0.0914*** (0.0287)	-0.117* (0.0619)	
onbekend	-0.470 (0.631)	-0.165** (0.0691)	-0.183* (0.102)	

Variabelen	Verzuimpercentage			
	Totaal	Kort verzuim	Middellang verzuim	Lang verzuim
aantal werknemers (honderdtallen)	-0.133** (0.0611)	-0.0182** (0.00789)		-0.128** (0.0532)
leerling-leraar ratio	-0.0598** (0.0281)			-0.0650** (0.0265)
armoedeprobleemcumulatiegebied	0.635 (0.449)		0.139 (0.116)	0.726** (0.349)
Leeftijd (aandelen)				
> 25 en <= 35			-0.443 (0.473)	
> 35 en <=45			0.569 (0.430)	
> 45 en <= 55			0.829** (0.387)	
> 55 en <= 60	5.598*** (1.486)	0.211 (0.150)	0.766* (0.421)	4.632*** (1.409)
> 60	4.539*** (1.578)		0.625 (0.460)	3.668** (1.429)
Functie (aandelen)				
OP - groepsleerkracht			1.947*** (0.299)	
OP - Vakleraar			2.014*** (0.298)	
OP - overig			1.458*** (0.316)	
OOP			2.683*** (0.210)	
OBP				3.554*** (0.790)
onbekend			1.628*** (0.278)	
salarisschaal (aandelen)				
> 3.000 en <= 4.000			-0.438 (0.316)	
> 4.000 en <= 5.000		0.204*** (0.0635)	-0.635* (0.326)	
> 5.000 en <= 6.000			-0.754** (0.343)	
> 6.000			2.751*** (1.004)	
onbekend		-0.315*** (0.0730)	-0.710** (0.285)	
werktijdfactor (afwijking gemiddelde)				
	-3.153** (1.284)	-0.348** (0.152)	-1.264*** (0.302)	
Jaar (ref = 2016)				
2017	0.236*** (0.0915)	-0.102*** (0.0106)	-0.0285 (0.0254)	0.363*** (0.0878)
2018	0.495*** (0.109)	-0.101*** (0.0149)	0.0778* (0.0400)	0.531*** (0.105)

Variabelen	Verzuimpercentage			
	Totaal	Kort verzuim	Middellang verzuim	Lang verzuim
2019	0.491*** (0.113)	-0.140*** (0.0153)	-0.0273 (0.0379)	0.645*** (0.108)
2020	0.603*** (0.116)	-0.294*** (0.0124)	0.0921 (0.0563)	0.844*** (0.112)
2021	0.447*** (0.124)	-0.278*** (0.0123)	0.120** (0.0577)	0.647*** (0.118)
Constante	4.685*** (0.619)	0.857*** (0.0443)	-1.036** (0.435)	2.707*** (0.586)
R-kwadraat:				
<i>Within</i>	0.03	0.33	0.05	0.05
<i>Between</i>	0.17	0.25	0.14	0.15
<i>Overall</i>	0.10	0.28	0.11	0.10
Waarnemingen	3,395	3,415	3,413	3,390
Aantal scholen	613	620	620	612

*Bron: Bewerking SEOR op basis van DUO-data. Toelichting: het aantal sterren geeft de mate van significantie aan: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. De linker kolom bevat het originele model, de rechterkolom een aangepast model waar alleen de significante variabelen zijn meegenomen.*

4.1.2 Verzuimbenchmark

De huidige Voion-verzuimbenchmark kan verbeterd worden door hierin naast schooltype ook andere factoren mee te nemen. Hierbij kan gedacht worden aan de schoolkenmerken denominatie en verstedelijking. Daarnaast zouden ook persoonskenmerken meegenomen kunnen worden, zoals leeftijd (aandelen 55-60 en 60+). Verder kan het nuttig zijn om de benchmark op te knippen in kort, middellang en lang verzuim omdat we zien dat verschillende kenmerken hier een ander effect op hebben.

Tabel 4.3 geeft een overzicht van de sterkte van de effecten die de verschillende factoren hebben op het verzuimpercentage. De effecten van de schoolkenmerken geven steeds het relatieve effect van een categorie ten opzichte van de referentiecategorie. Bij schooltype is havo/vwo de referentiecategorie. Vmbo-t-scholen hebben bijvoorbeeld gemiddeld een verzuimpercentage dat ruim 1 procentpunt hoger ligt dan havo/vwo-scholen. Maar ook in geval van denominatie kan het verschil tussen scholen ruim 1 procentpunt zijn (verschil tussen algemeen bijzondere scholen en gereformeerde, reformatorische en evangelische scholen).

De persoonskenmerken betreffen geen categorieën, maar zijn continue variabelen. Om de effecten van de persoonskenmerken met elkaar te kunnen vergelijken en ook enigszins te kunnen vergelijken met de effecten van de schoolkenmerken, zijn de effecten berekend van een verandering met de standaarddeviatie (een maat voor de spreiding van een variabele). Hieruit blijkt dat de leeftijdscategorie 55-60 het sterkste effect heeft op het verzuimpercentage.

Tabel 4.3 Sterkte van de effecten van schoolkenmerken en persoonskenmerken op verzuimpercentage

Schoolkenmerken	Effect*	Persoonskenmerken	Effect**
Schooltype (ref. = havo/vwo)		Aantal werknemers	-0,17
pro	0,96	Leerling-leraar ratio	-0,19
vmbo	0,65	APC-gebied	0,11
vmbo-t	1,10	Leeftijdscategorie > 55 en <= 60	0,27
vmbo-t/havo/vwo	0,50	Leeftijdscategorie > 60	0,20
vmbo/havo/vwo	1,01	Werktijdfactor (afwijking gemiddelde)	-0,12
Denominatie (ref. = openbare scholen)			
Rooms katholieke scholen	-0,26		

Schoolkenmerken	Effect*	Persoonskenmerken	Effect**
Protestants Christelijke scholen	-0,38		
Algemeen bijzondere scholen	0,52		
Gereformeerde, reformatische, evangelische scholen	-0,56		
Islamitische scholen	-0,10		
Vrije scholen	0,28		
Overige scholen	-0,36		
Verstedelijking (ref. = sterk verstedelijkt)			
verstedelijkt	-0,29		
niet verstedelijkt	-0,47		
onbekend	-0,47		

Bron: *Bewerking SEOR op basis van DUO-data.*

* *Het relatieve effect ten opzichte van de referentiegroep.*

** *Het effect van een toename in de betreffende factor met de standaarddeviatie van die factor.*

Het voordeel van het meenemen van meer factoren in de benchmark is dat het streefpercentage voor scholen realistischer wordt. Nu hebben bijvoorbeeld scholen met een laag aandeel 55+'ers het relatief makkelijker om het streefpercentage te halen dan scholen met een hoog aandeel 55+'ers. Wanneer dit in de benchmark wordt meegenomen worden scholen met een laag aandeel 55+'ers meer uitgedaagd. Het aanpassen van de benchmark zal waarschijnlijk wel meer van DUO vragen, omdat de benchmark meer gevoed moet worden met gegevens vanuit de data.

4.2 AANBEVELINGEN

Hieronder doen we een aantal aanbevelingen op basis van dit onderzoek. De aanbevelingen hebben enerzijds betrekking op de data zelf, en anderzijds op aanknopingspunten voor verzuimbeleid (en de Verzuimbenchmark-VO).

- 1) Over het algemeen is de kwaliteit van de verzuimdata goed, maar er ontbreken wel scholen in de verzuimregistratie. Scholen moeten een duidelijke prikkel hebben om het verzuim goed te registreren. Uiteindelijk hebben scholen er zelf baat bij dat ze goed weten hoe hoog het verzuim binnen hun school is, zeker vanwege het lerarentekort. Deze boodschap moet aankomen bij de scholen.
- 2) Dit onderzoek laat zien dat het heel lastig is om ziekteverzuim goed te verklaren met de beschikbare data. Op persoonsniveau zien we dat er geen eenduidige patronen van verzuim te vinden zijn, en op schoolniveau zien we dat de verklaringskracht van de modellen laag liggen. De data mist informatie over persoonlijke omstandigheden en omstandigheden op de school om (de ontwikkeling in) verzuim goed te kunnen duiden. Een mogelijkheid om de data te verrijken is om een koppeling te maken met CBS-microdata. Hierdoor kunnen meer factoren onderzocht worden, bijvoorbeeld over de levensfase van personen (huishoudsituatie, wel/geen kinderen, etc.). Ook hier blijven uiteraard beperkingen aan zitten, omdat er zeer waarschijnlijk relevante factoren zijn die niet kwantitatief te meten zijn.
- 3) Door bovenstaande beperkingen is het lastig om goede aanknopingspunten te vinden voor het verbeteren van het verzuimbeleid. Bestaande initiatieven zoals het richten op 55+'ers lijken terecht op basis van de uitkomsten uit dit onderzoek. Daarnaast is er een groep 30- tot 40-jarige leraren op scholen met een hoge leerling-leraar ratio die wellicht meer aandacht nodig hebben, omdat we zien dat deze groep relatief vaak kort of middellang verzuim heeft.
- 4) Het is terecht dat de huidige verzuimbenchmark onderscheid maakt naar schooltype. Het kan goed zijn om nog meer gedetailleerde uitsplitsingen te maken naar bijvoorbeeld verstedelijking en aandelen 55+'ers. We zien dat er gemiddeld genomen verschillen zijn in het verzuimpercentage tussen scholen op basis van deze factoren.

- 5) Om meer aanknopingspunten voor beleid te vinden, zou er aanvullend onderzoek gedaan moeten worden. De beschikbare data biedt wel een goede basis voor bijvoorbeeld het uitvoeren van een pilot. Zo zou er geëxperimenteerd kunnen worden met een bepaald verzuimbeleid op een (gerandomiseerde) selectie van de scholen. Vervolgens kan de verzuimdata gebruikt worden om de effecten van het beleid te analyseren. Als alternatief kan ook worden geïnventariseerd wat de huidige verschillen zijn in het verzuimbeleid tussen scholen waarna dit gekoppeld kan worden aan de verzuimdata.

I BESCHRIJVENDE STATISTIEKEN

Tabel I.1 Beschrijvende statistieken variabelen verzuimdata

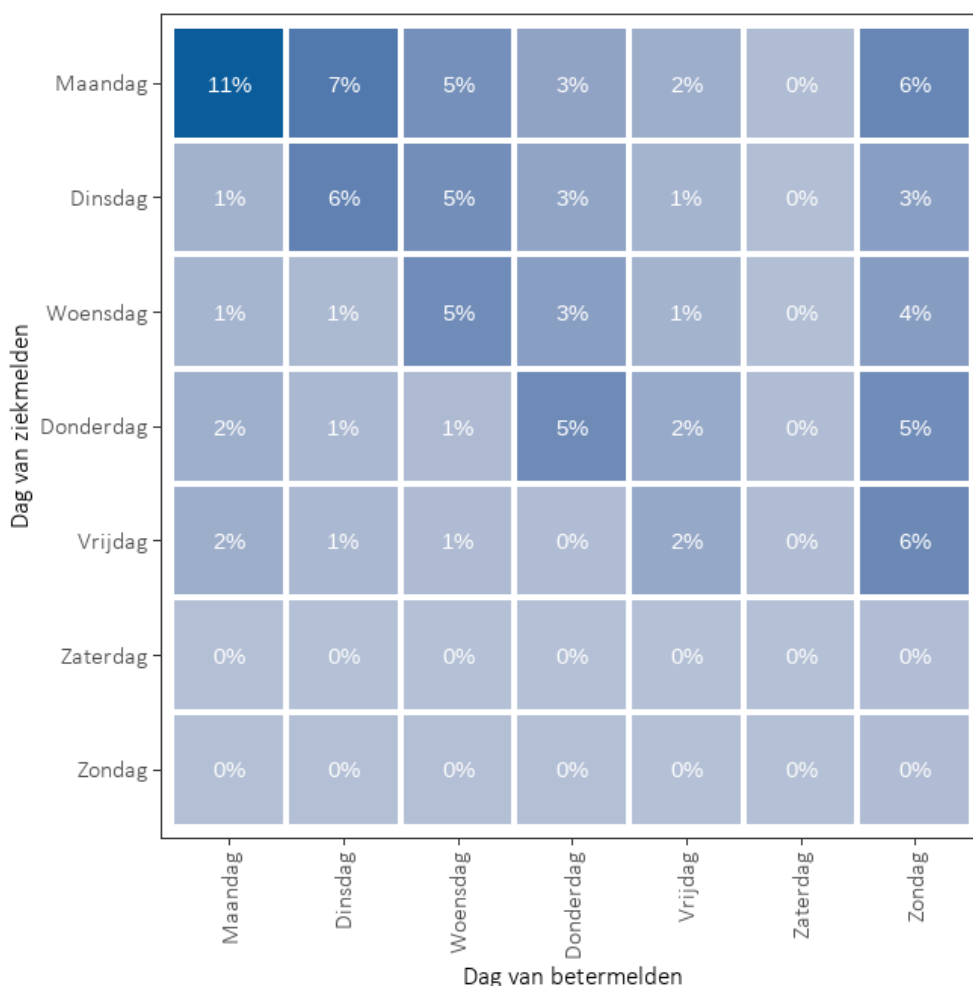
	2016	2017	2018	2019	2020	2021	Totaal
Totaal aantal personen (peilmaand oktober)	103.805	105.107	103.344	99.793	98.286	103.076	159.475
Verlof							
<i>Verdeling start verlofperiodes over dagen</i>							
Zondag	0%	1%	1%	1%	1%	1%	0%
Maandag	34%	34%	34%	34%	35%	34%	34%
Dinsdag	21%	21%	21%	21%	20%	20%	21%
Woensdag	16%	16%	15%	16%	15%	16%	16%
Donderdag	17%	16%	17%	16%	16%	16%	17%
Vrijdag	12%	12%	11%	12%	12%	12%	12%
Zaterdag	0%	0%	1%	0%	1%	0%	0%
<i>Verdeling eind verlofperiodes over dagen</i>							
Zondag	26%	25%	25%	23%	24%	22%	26%
Maandag	17%	17%	17%	17%	17%	16%	17%
Dinsdag	16%	17%	17%	17%	15%	15%	16%
Woensdag	17%	17%	17%	17%	15%	16%	17%
Donderdag	15%	15%	15%	15%	15%	15%	15%
Vrijdag	7%	8%	8%	9%	11%	13%	7%
Zaterdag	1%	1%	1%	1%	2%	3%	1%
Categorie werktijdfactor							
2 dagen of minder	19%	19%	19%	19%	18%	20%	19%
2-4 dagen	43%	44%	45%	46%	47%	46%	45%
Tussen 4 en 5 dagen	15%	14%	13%	12%	12%	11%	13%
Fulltime	24%	24%	23%	23%	23%	23%	23%
Gemiddelde werktijdfactor	0,69	0,68	0,68	0,68	0,69	0,67	0,68
Categorie brutosalaris							
Onbekend	2%	1%	1%	1%	1%	1%	1%
Minder dan 3000 euro	27%	27%	27%	24%	20%	19%	24%
3000-4000 euro	33%	33%	16%	19%	21%	24%	24%
4000-5000 euro	22%	21%	37%	37%	23%	22%	27%
5000-6000 euro	16%	17%	18%	17%	33%	32%	22%
Meer dan 6000 euro	1%	1%	1%	2%	2%	2%	1%
Gemiddeld salaris	€ 3.860	€ 3.862	€ 3.955	€ 4.067	€ 4.172	€ 4.204	€ 4.018
Leeftijdscategorie							
Jonger dan 30	16%	16%	16%	16%	16%	17%	16%
30-40	21%	21%	22%	22%	23%	23%	22%
40-55	36%	35%	35%	35%	35%	35%	35%
55+	28%	28%	28%	27%	27%	25%	27%
Gemiddelde leeftijd	45,4	45,5	45,4	45,3	45,1	44,6	45,2
Functiecategorieën							

	2016	2017	2018	2019	2020	2021	Totaal
DIR	4%	4%	4%	4%	4%	4%	4%
OP	72%	73%	73%	73%	72%	72%	72%
OOP	14%	13%	14%	14%	14%	15%	14%
OBP	10%	10%	10%	10%	10%	10%	10%
Onbekend	1%	0%	0%	0%	0%	0%	0%
Categorie aantal werknemers							
Minder dan 100 wn	10%	10%	10%	10%	11%	11%	10%
100-200 wn	25%	26%	24%	25%	26%	25%	25%
200-300 wn	28%	30%	30%	28%	26%	25%	28%
300-400 wn	17%	16%	16%	18%	17%	17%	17%
400-500 wn	13%	12%	12%	11%	11%	11%	12%
500 of meer wn	7%	6%	7%	7%	9%	11%	8%
Denominatie							
Openbare scholen	27%	28%	28%	28%	28%	28%	28%
Rooms katholieke scholen	25%	25%	24%	23%	23%	23%	24%
Protestants Christelijke scholen	18%	19%	19%	19%	19%	18%	19%
Algemeen bijzondere scholen	12%	11%	11%	11%	11%	11%	11%
Geref./ref./evangelisch	4%	4%	4%	4%	4%	4%	4%
Islamitisch	0%	0%	0%	0%	0%	0%	0%
Vrije scholen	1%	1%	1%	1%	1%	1%	1%
Overig	13%	13%	13%	14%	14%	14%	13%
Schooltype							
vmbo/havo/vwo	84%	84%	84%	84%	83%	83%	84%
havo/vwo	8%	8%	9%	9%	9%	9%	9%
vmbo	4%	4%	4%	4%	4%	4%	4%
pro	4%	4%	4%	4%	4%	4%	4%
Regio							
Noord	10%	10%	10%	10%	8%	10%	10%
Oost	23%	23%	23%	23%	24%	23%	23%
West	45%	45%	45%	46%	46%	46%	45%
Zuid	22%	22%	22%	21%	21%	21%	22%
Stedelijkheid							
sterk verstedelijkt	24%	25%	25%	27%	28%	28%	26%
verstedelijkt	60%	58%	58%	57%	56%	57%	58%
niet verstedelijkt	16%	17%	16%	16%	16%	15%	16%
Onbekend	0%	0%	1%	0%	0%	0%	0%
Oordeel inspectie							
Geen oordeel	6%	6%	6%	5%	5%	6%	6%
Onvoldoende	2%	1%	0%	0%	0%	0%	1%
Onvoldoende, tenzij	20%	16%	16%	15%	24%	8%	16%
Voldoende	73%	77%	77%	80%	71%	86%	78%
Leerlingen uit APCG-gebieden							
Geen leerlingen uit apcg-gebied	12%	12%	14%	14%	9%	8%	12%
Minder dan 10% van leerlingen uit apcg-gebied	57%	57%	55%	54%	45%	46%	52%
Tussen 10% en 25% van leerlingen uit	18%	17%	17%	18%	27%	27%	21%

	2016	2017	2018	2019	2020	2021	Totaal
apcg-gebied							
Meer dan 25% van leerlingen uit apcg-gebied	14%	14%	14%	13%	19%	19%	15%
Leerling-leraar ratio							
14 of minder LLR	18%	16%	17%	19%	23%	31%	21%
14-16 LLR	43%	45%	47%	47%	45%	46%	45%
16-18 LLR	30%	31%	29%	27%	25%	19%	27%
>18 LRR	10%	9%	7%	7%	7%	4%	7%

Bron: Bewerking SEOR op basis van DUO-data. Toelichting: voor vrijwel elke variabele is oktober als peilmaand gepakt in ieder jaar en geven de percentages de verdeling van unieke personen over de verschillende categorieën. Dit geldt niet voor de verdeling van de verlofdagen, dit is berekend over alle verlofperiodes heen, en laat dus de verdeling van verlofdagen over de dagen van de week zien.

Figuur I.1 Matrix dag van ziekmelden en betermelden



Bron: Bewerking SEOR op basis van DUO-data. Toelichting: de percentages geven het aandeel van een bepaalde combinatie (bijv. maandag ziekmelden, dinsdag betermelden) in het totaal van alle combinaties weer. De meest voorkomende combinatie is dus maandag ziekmelden, maandag betermelden (11% van het totaal).

I.1 VERDELING AANTAL DAGEN VERZUIM OVER CATEGORIEËN

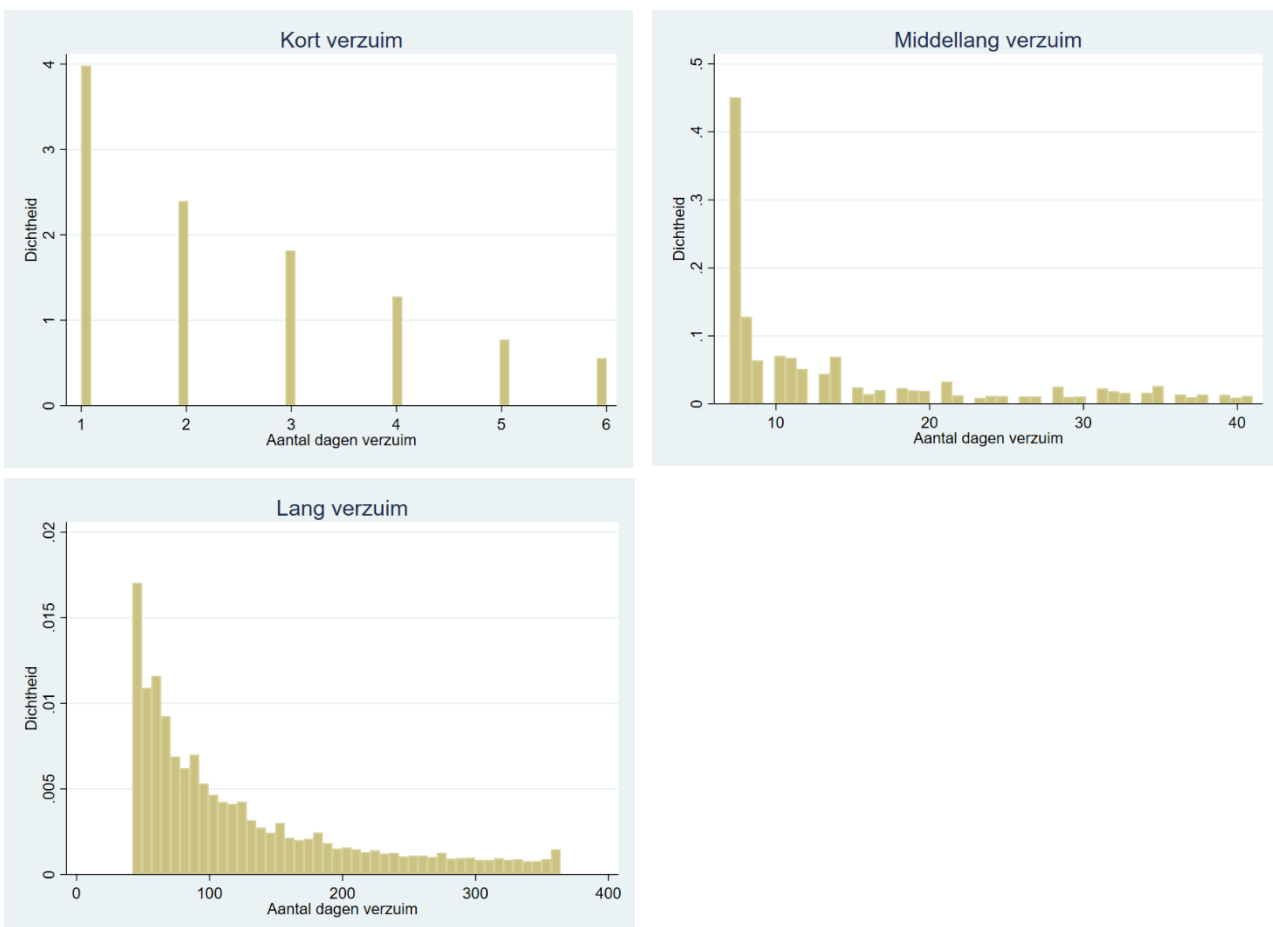
De verzuimduur is voor de sequentieanalyse opgedeeld in de volgende categorieën:

- Geen verzuim;
- Kort verzuim (aansluitende verzuimperiode korter dan 7 dagen);
- Middellang verzuim (aansluitende verzuimperiode van 7 dagen tot en met 41 dagen);
- Lang verzuim (aansluitende verzuimperiode van 42 dagen of langer).

We zien dat het aantal dagen verzuim binnen deze categorieën niet normaal verdeeld is. Bij elke categorie zien we een links-scheve verdeling van het aantal dagen verzuim. Dat wil zeggen: binnen elke categorie komt het minimale aantal dagen verzuim van die categorie het vaakst voor (zie figuur I.2).

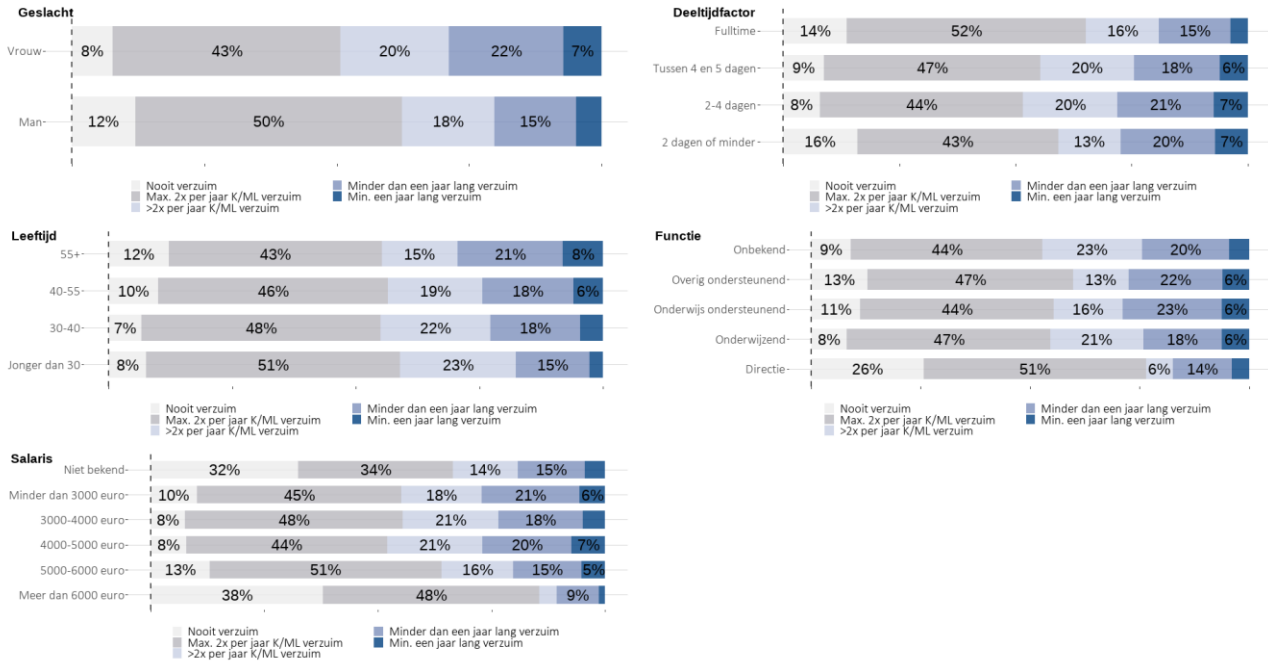
We hebben daarom voor de sequentieanalyse ook een andere indeling in verzuimcategorieën geprobeerd, die wat meer gedifferentieerd is. We hebben gekeken naar een indeling van zeer kort verzuim (max 2 dagen), kort verzuim (3-4 dagen), middel verzuim (5-7 dagen), middellang verzuim (8-35 dagen) en lang verzuim (>35 dagen). Deze indeling leverde echter geen andere resultaten op in de sequentie- en clusteranalyse. Ook voor deze indeling geldt dat het steeds links scheef verdeeld blijft binnen de categorieën, omdat de gehele verdeling afloopt van het minste naar het meeste aantal dagen.

Figuur I.2: Histogrammen verdeling aantal verzuimdagen



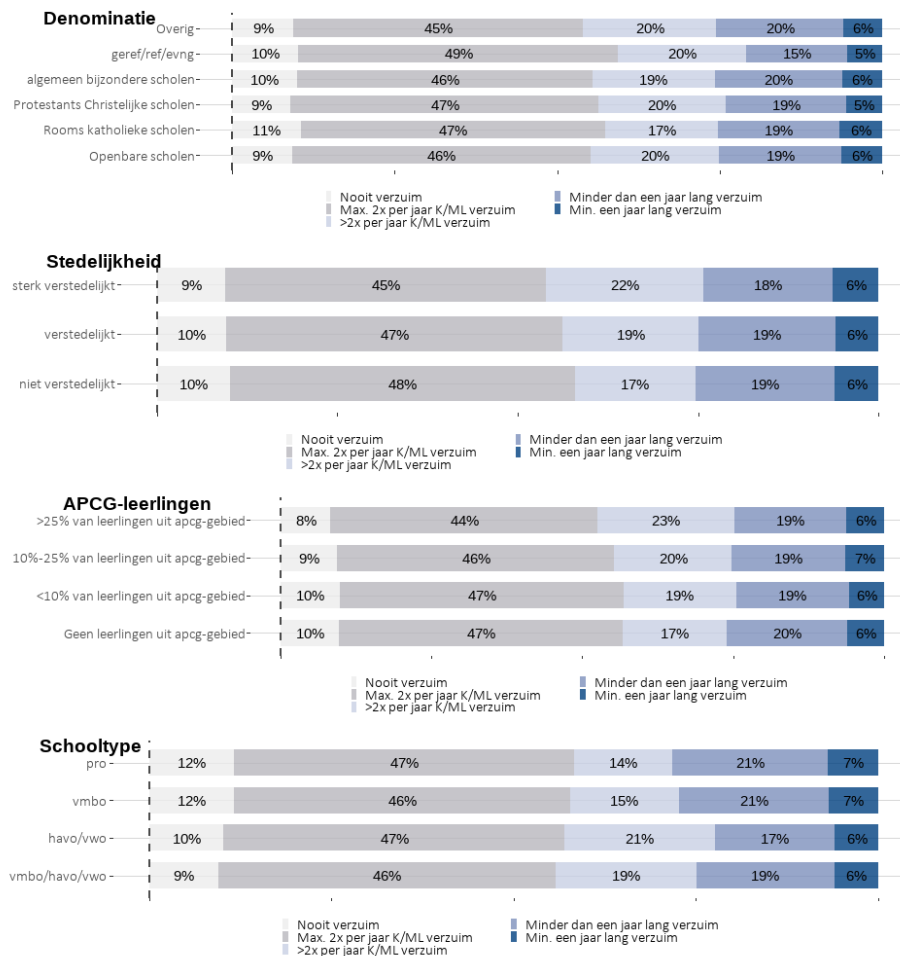
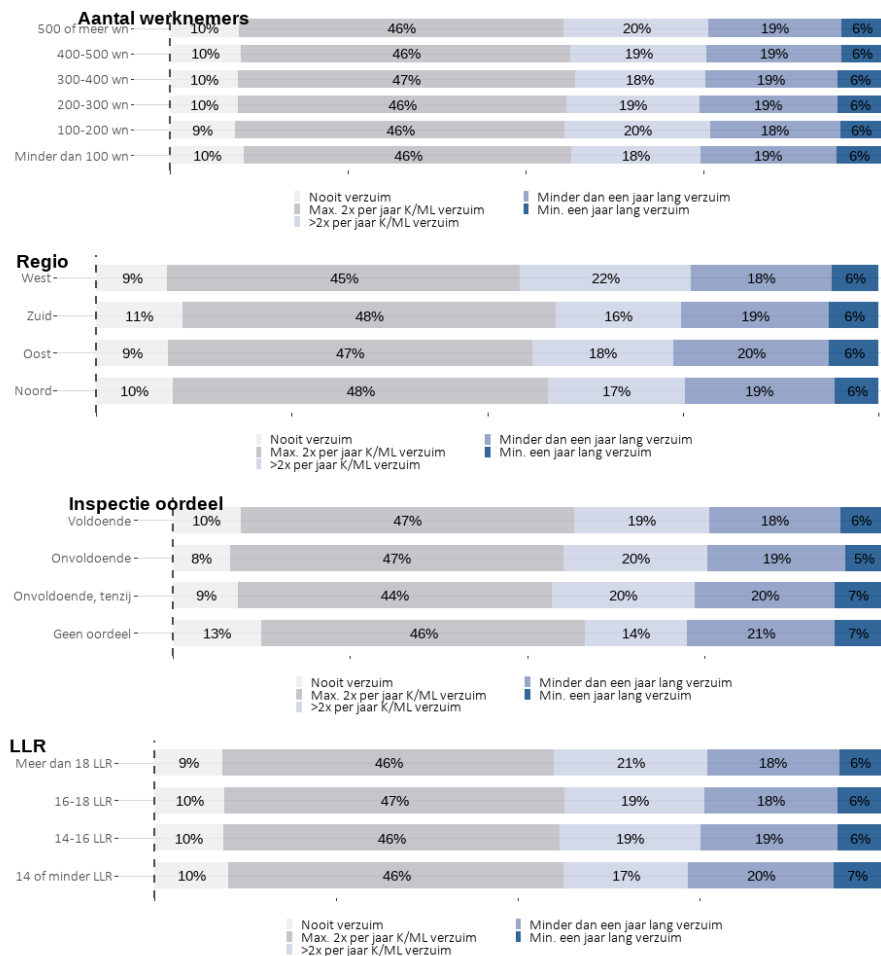
I.2 FIGUREN VERDELING KENMERKEN OVER CLUSTERS

Figuur I.3 Verdeling clusters over persoonskenmerken



Bron: Bewerking SEOR op basis van DUO-data. Toelichting: de categorieën van continue variabelen zijn elkaar uitsluitende categorieën.

Figuur I.4 Verdeling clusters over schoolkenmerken



Bron: Bewerking SEOR op basis van DUO-data. Toelichting: de categorieën van continue variabelen zijn elkaar uitsluitende categorieën.

Dit rapport doet verslag van onderzoek dat is uitgevoerd in opdracht van het ministerie van OCW en Voion.

De verantwoordelijkheid voor de inhoud berust bij SEOR BV. Het gebruik van het materiaal in dit rapport is toegestaan, mits de bron duidelijk wordt vermeld. Vermenigvuldiging en/of openbaarmaking in welke vorm dan ook is uitsluitend toegestaan na schriftelijke toestemming van SEOR BV.

Het onderzoek is uitgevoerd door SEOR BV.

Dit rapport is een uitgave van SEOR BV. Marconistraat 16. 3029 AK Rotterdam.

www.seor.nl