

# Overzicht zorgvraagtyperingsanalyses



# Inhoudsopgave

<b>1</b>	<b>Vooraf</b>	<b>6</b>
1.1	Wat mist er nog?	7
1.1.1	Aandachtspunten NZa	7
<b>2</b>	<b>Zorgvraagtyperingsgegevens</b>	<b>9</b>
2.1	Verkenning ontvangen gegevens	9
2.1.1	Aantal ontvangen scorelijsten per methode	9
2.1.2	Aantal per type aanbieder	10
2.1.3	Aantal per gekozen zorgvraagtype	11
2.2	Vergelijken keuze en voorspelling	12
2.2.1	Aandeel voorspeld en gekozen zorgvraagtype	12
2.2.2	Aandeel gekozen per zorgvraagtype en type zorgaanbieder	13
2.2.3	Relatieve keuze tov voorspelling	14
2.2.4	Overeenkomst voorspelling en keuze	15
2.2.5	Keuze tov voorspelling - welk zorgvraagtype	16
2.2.6	Relatieve keuze tov voorspelling	17
2.3	Volledig en dynamisch vergelijken	21
2.3.1	Keuze en voorspeld zorgvraagtype per methode	21
2.3.2	Leereffect	23
2.3.3	Steekproeven dynamische methode	24
2.3.4	Aantal items	26
2.3.5	Aantal items per hoofdgroep	28
2.3.6	Specifieke items dynamische methode	29
2.4	Zorgvraagtypering en behandelinzet	31
2.4.1	Maanden in data	32
2.4.2	Verdeling kosten per zorgvraagtype	34
2.4.3	Kosten per maand	36
2.5	Zorgvraagtype en type prestatie	37
2.5.1	Aandeel zorgvraagtypering met prestatietype	38
2.5.2	Kosten per prestatietype	39
2.6	Verbeteren algoritmes	41

2.6.1	Afwijking voorspelling van keuze	42
2.6.2	Mate van afwijking	43
2.6.3	Rode regels	52
2.6.4	Behandelaar volgt algoritme	56
2.6.5	Verfijnen trainen LDA	58
2.6.6	Minder variabelen	59
2.6.7	Hoeveel structuur in honosvragen?	66
2.7	Wvggz/Wzd, verblijf of acute ggz in laatste jaar	68
2.7.1	Voorkomen van de kenmerken	68
2.7.2	Voorkomen kenmerken per zorgvraagtype	69
2.7.3	Verschil kosten bij kenmerk	71
2.7.4	Kosten kenmerken per prestatietype	76
2.7.5	Voorkomen kenmerk en relatieve kosten	80
2.7.6	Kenmerken en kwaliteitsparameters	85
2.7.7	Kenmerken en samengevoegde zorgvraagtypes	86
2.7.8	Andere alternatieve indelingen	93
2.7.9	Algoritme bij expertindeling 3	94
2.8	Individuele items	95
2.8.1	Scores per item - hoofdgroep X	96
2.8.2	Scores per item - hoofdgroep Y	97
2.8.3	Scores per item - hoofdgroep Z	98
2.9	Vragen en kosten binnen zorgvraagtypering	99
2.9.1	Correlatie tussen score en kosten per zorgvraagtype	100
2.9.2	Voorbeeld klein aantal hoge scores	101
2.9.3	Aandeel scorend op item	102
2.9.4	Kosten en aandeel bij splitsing - score > 0	105
2.10	Subschalen	110
2.10.1	Voorkomen subschaalscores bij zorgvraagtype	111
2.10.2	Subschaal en kosten	112
2.10.3	Splitsingspunt subschalen bepalen	113
<b>3</b>	<b>Declaratiegegevens</b>	<b>116</b>
3.1	Selectie declaratiegegevens en labelling	117
3.1.1	Selectie gegevens	117

3.1.2	Labelling trajecten	117
3.2	Verkennen declaraties	118
3.2.1	Aantal trajecten per zorgvraagtype	118
3.2.2	Looptijd per zorgvraagtype	119
3.2.3	Kosten per zorgvraagtype	122
3.3	Verdeling over settingen	125
3.3.1	Settingen bij zorgvraagtype	126
3.3.2	Zorgvraagtype in setting	127
3.3.3	Kosten zvt en setting	128
3.4	Zorgvraagtype en diagnosehoofdgroep	132
3.4.1	Meest voorkomende diagnosehoofdgroepen	133
3.4.2	Overige diagnosehoofdgroepen	134
3.4.3	Kosten per diagnosehoofdgroep	135
3.4.4	Gemiddelde kosten	136
3.5	Generalistische basis-ggz	137
3.5.1	Aandeel per gbg-profiel	138
3.5.2	Kosten per gbg-profiel	139
3.6	Eerdere s-ggz behandeling	140
3.6.1	Aandeel eerder s-ggz	141
3.6.2	Kosten en eerder s-ggz-gebruik	142
3.7	Hertypering	143
3.7.1	Aantal hertyperingen	144
3.7.2	Aandeel van trajecten	145
3.7.3	Hertypering - zvt voor en na	147
3.7.4	Op- of afschalen	148
3.8	Samenvoegen zorgvraagtypes	153
3.8.1	Samenvoegen tot hoofdgroepen	154
3.8.2	Samenvoegen subgroep	155
3.8.3	Samenvoegen obv clustering	156
<b>4</b>	<b>Beoordelen kwaliteit model</b>	<b>158</b>
4.1	Niveau beoordeling	158
4.2	Methoden beoordeling	159
4.3	Uitvoering kwaliteitsberekeningen	160

4.3.1	Gemiddelde absolute fout (MAE)	163
4.3.2	Verklaren basisfout	165
4.3.3	MAE per type aanbieder	169
4.3.4	Verklaarde basisfout per type aanbieder	171
<b>5</b>	<b>Vergelijking met evaluatie van het zorgclustermodel</b>	<b>172</b>

# 1 Vooraf

De Tweede Kamer heeft een motie over de zorgvraagtypering aangenomen. Met deze motie wil de Kamer vernietiging van de zorgvraagtyperingsgegevens in gang zetten en wil ze beperkingen stellen aan het type gegevens dat de NZa opvraagt of verwerkt. Dit is een duidelijk signaal dat de politieke werkelijkheid is veranderd. De afweging tussen de winst die de ggz kan behalen aan de ene kant en de privacy aan de andere kant wordt in deze nieuwe werkelijkheid anders gemaakt. De NZa kan niet anders dan consequenties aan dit signaal verbinden en zal de gegevens eerder dan gepland vernietigen. Dit betekent dat we, als adviescommissie zorgvraagtypering ggz, met elkaar moeten versnellen op het gebied van de verbeteringen die we nog aan het model willen doorvoeren. Hiervoor is het belangrijk dat de partijen uit de commissie aangeven welke analyses zij nog nodig achten en welke verbeteringen dit op zal leveren. De adviescommissie heeft gevraagd om een overzicht met uitgevoerde analyses voordat zij in kaart kunnen brengen wat er nog nodig is. Dit document bevat het overzicht van analyses.

De opdracht van de adviescommissie is het adviseren over hoe de zorgvraagtypering nog beter een verband kan leggen tussen het zorgvraagtype en de zorginzet. Hiervoor zijn twee verbanden in het systeem belangrijk, namelijk 1) het verband tussen het zorgvraagtype en de zorginzet en 2) het verband tussen de zorgvraag, zoals die is gescoord op het zorgvraagtyperingsinstrument en het gekozen zorgvraagtype. De adviescommissie kan twee verschillende datasets gebruiken om deze verbanden te onderzoeken. Ten eerste zijn er de declaratiegegevens, deze bestaan uit alle zorgdeclaraties in de ggz onder het Zorgprestatie-model en bevatten gedetailleerde gegevens over de geleverde zorgprestaties en hebben aan de kant van de zorgvraag alleen informatie over de zorgvraagtypering, diagnosehoofdgroep en basis-ggz-profiel. Ten tweede kan gebruik worden gemaakt van de zorgvraagtyperingsgegevens. Deze gegevens bevatten meer informatie over de zorgvraag en minder detail over de ingezette zorg. Specifiek bevatten de zorgvraagtyperingsgegevens de zorgvraagtyperingen die tussen 1 juli 2022 en 1 juli 2023 zijn uitgevoerd en die in de tweede helft van 2023 zijn aangeleverd bij de NZa. De gegevens bevatten scorelijsten, de methode waarop de typering is uitgevoerd, diagnosehoofdgroep, basis-ggz-profiel, de behandelinzet geaggregeerd in brede categorieën per maand en drie aanvullende kenmerken van de patiënt, namelijk of er in het afgelopen jaar sprake is geweest van acute ggz, verblijf of Wvvgg/Wzd. In dit document worden de analyses vanuit de twee verschillende datasets achtereenvolgens benoemd, te beginnen met de zorgvraagtyperingsgegevens.

- Behandelinzet: De geleverde zorg bij de zorgvraagtypering geaggregeerd naar relatief brede categorieën per maand.
- Zorgvraagtype: Het zorgvraagtype dat de behandelaar bij de scorelijst heeft gekozen
- Zorgvraagtypering: De gescoorde scorelijst met gekozen zorgvraagtype en behandelinzet. Dit is niet noodzakelijkerwijs een volledig zorgtraject. Behandelinzet bij zorgvraagtypering bevat in veel gevallen slechts een deel van het zorgtraject, omdat soms zorg is geleverd voor het uitvoeren van de zorgvraagtypering, omdat soms een nieuwe zorgvraagtypering is uitgevoerd waar het vervolg van het zorgtraject onder is geschaard of omdat de behandeling nog niet was voltooid op het moment dat de gegevens zijn aangeleverd.

## 1.1 Wat mist er nog?

Zorgvraagtypering staat op de factuur en zal gebruikt worden in gesprekken tussen zorgaanbieders en zorgverzekeraars, in de regio of bilateraal. Het is hiervoor belangrijk dat het model zo goed mogelijk is, zodat de doelen van de zorgvraagtypering behaald kunnen worden. Om het model te verbeteren heeft de NZa advies gevraagd van partijen uit de ggz, verenigd in de adviescommissie ggz. De NZa ontvangt advies van de adviescommissie en zal, voor de onderdelen die onder haar invloed en verantwoordelijkheid vallen, zorgen dat er uiterlijk half 2025 een verbeterde zorgvraagtypering is. Het is daarmee niet gezegd dat de verbeterde zorgvraagtypering direct gebruikt kan worden, het is namelijk mogelijk dat implementatie, bijvoorbeeld in de ICT of voor trainen van behandelaren, tijd gaat kosten.

De werkgroep data analyse heeft zeer veel analyses uitgevoerd, besproken en verfijnd om te komen tot een verbeterde zorgvraagtypering. Deze analyses zijn in dit document gebundeld en zullen worden gedeeld met de adviescommissie en andere werkgroepen. Met de inzichten uit dit document kunnen de partijen tezamen en met dezelfde kennis de relevante adviezen tot stand brengen.

De adviescommissie en werkgroep data hebben nog enkele kleine vragen aangedragen op 7 en 8 oktober respectievelijk. Deze vragen zijn in dit document verwerkt en leidt tot een complete set aan analyses die onder de advisering van de adviescommissie zal liggen. Afronden van de analyses is ook nodig gezien de versnelde vernietiging van de zorgvraagtyperingsgegevens door de NZa, als gevolg van de verandering van toon over gegevensdeling in het maatschappelijke en politieke debat.

### 1.1.1 Aandachtspunten NZa

De NZa zal hier geen waardeoordeel geven over het systeem van zorgvraagtypering. Het is aan de adviescommissie om hierover aan de NZa te adviseren. De NZa ziet wel dat veel analyses zijn uitgevoerd. Naast de uitgebreide set met analyses die de werkgroep data-analyse heeft uitgevoerd en die hieronder wordt gedocumenteerd ziet de NZa nog enkele belangrijke punten waarover advies gewenst is.

**Herijken algoritmes** Een vernieuwing van de algoritmes onder de volledige zorgvraagtypering heeft plus- en minpunten. De nieuwe algoritmes lijken bij sommige zorgvraagtypes vaker een ander zorgvraagtype te voorspellen dan de behandelaar heeft gekozen in vergelijking met de oude algoritmes. De vernieuwde algoritmes lijken minder vaak sterk af te wijken. De afwegingen onder de keuze of de volledige algoritmes beter zijn dan de oude wijzen dus niet allemaal dezelfde kant op. De NZa heeft op dit onderwerp een advies nodig en vraagt de commissie daarom een mening te vormen over het al dan niet opnieuw schatten van de volledige algoritmes. In sectie 2.6 kunnen analyses worden gevonden waarbij kunnen helpen.

Verdere openstaande vraag is het behouden of afschaffen van (een deel van) de rode regels in de algoritmes. Deze rode regels leiden er soms toe dat geen zorgvraagtype kan worden voorspeld door de algoritmes. De keuze lijkt richting acceptatie van de rode regels te gaan, maar dit is niet expliciet gemaakt. Advies over het al dan niet afschaffen van rode regels moet gebeuren op basis van zorginhoudelijke expert-kennis.

De adviescommissie heeft op advies van de werkgroep data besloten dat de dynamische algoritmes (beslisbomen) opgebouwd zullen worden uit een basisset vragen en dat de beslisbomen tot een grotere diepte zullen worden doorgerekend.

Als laatste is het goed te weten dat het nodig is de dynamische algoritmes opnieuw op te bouwen wanneer de volledige algoritmes opnieuw worden gemaakt. De twee algoritmes bij een hoofdgroep moeten immers op elkaar aansluiten. In dit document is nog geen aandacht besteed aan het verbeteren van de dynamische zorgvraagtypering, behalve een geaccepteerd advies rond het beginnen met een vaste set items afhankelijk van de hoofdgroep.

**Aanpassen indeling zorgvraagtypes** In dit document verkennen we enkele mogelijkheden tot het samenvoegen op basis van vergelijkbaarheid tussen zorgvraagtypes en opsplitsen van zorgvraagtypes op basis van 1) de aanvullende kenmerken rond gebruik van acute ggz, verblijf of Wvvgz/Wzd in het afgelopen jaar, 2) scores op individuele vragen of 3) subschalen van de HoNOS+. Deze analyses laten zien dat het mogelijk kan zijn (sommige) zorgvraagtypes op te splitsen, waarbij groepen ontstaan die gemiddeld verschillende kosten met zich mee brengen. In combinatie met samenvoegingen kan dit leiden tot een model met vergelijkbare prestaties op de kwaliteitsparameters en een (iets) kleiner aantal zorgvraagtypes. De adviescommissie zal moeten adviseren of het veranderen van de indeling wenselijk is en zo ja, welke samenvoegingen en splitsingen hierbij moeten worden aangehouden. Het is belangrijk hierbij rekening te houden met de hanteerbaarheid van het systeem (aantal zorgvraagtypes), de grootte van de groep patiënten met de 'nieuwe' zorgvraagtypes ten opzichte van de sector en de groep waarvan wordt afgesplitst en de grotere privacy-inbreuk voor de patiënt die een verfijning zou kunnen inhouden. Deze afwegingen worden verder uitgediept in het *afwegingskader subclusters* mee te nemen dat eerder is besproken in het programma Zorgprestatie-model (8 januari 2021, in bijlage). Naast deze bezwaren tegen aanpassingen aan de zorgvraagtype-indeling staan ook overwegingen rond de acceptatie van het model (wéér iets anders), meer administratie (registratie van de aanvullende elementen) en kosten die volgen uit het aanpassen van de ICT-systemen.



## 2 Zorgvraagtyperingsgegevens

Dit hoofdstuk is gebaseerd op analyses van de gegevens die in de zomer en het najaar van 2023 van zorgaanbieders zijn ontvangen, de zorgvraagtyperingsgegevens. Zoals ook hierboven is beschreven kunnen deze gegevens worden gebruikt voor het onderzoeken van de link tussen de scorelijst en het zorgvraagtype en veranderingen aan de zorgvraagtyperingsindeling.

### 2.1 Verkenning ontvangen gegevens

In het eerste deel van dit document beschrijven we verkennende analyses die inzicht moeten geven in de representativiteit van de gegevens en de spreiding van de zorgvraagtypes bij verschillende type aanbieders.

#### 2.1.1 Aantal ontvangen scorelijsten per methode

Tabel 1: Aantal

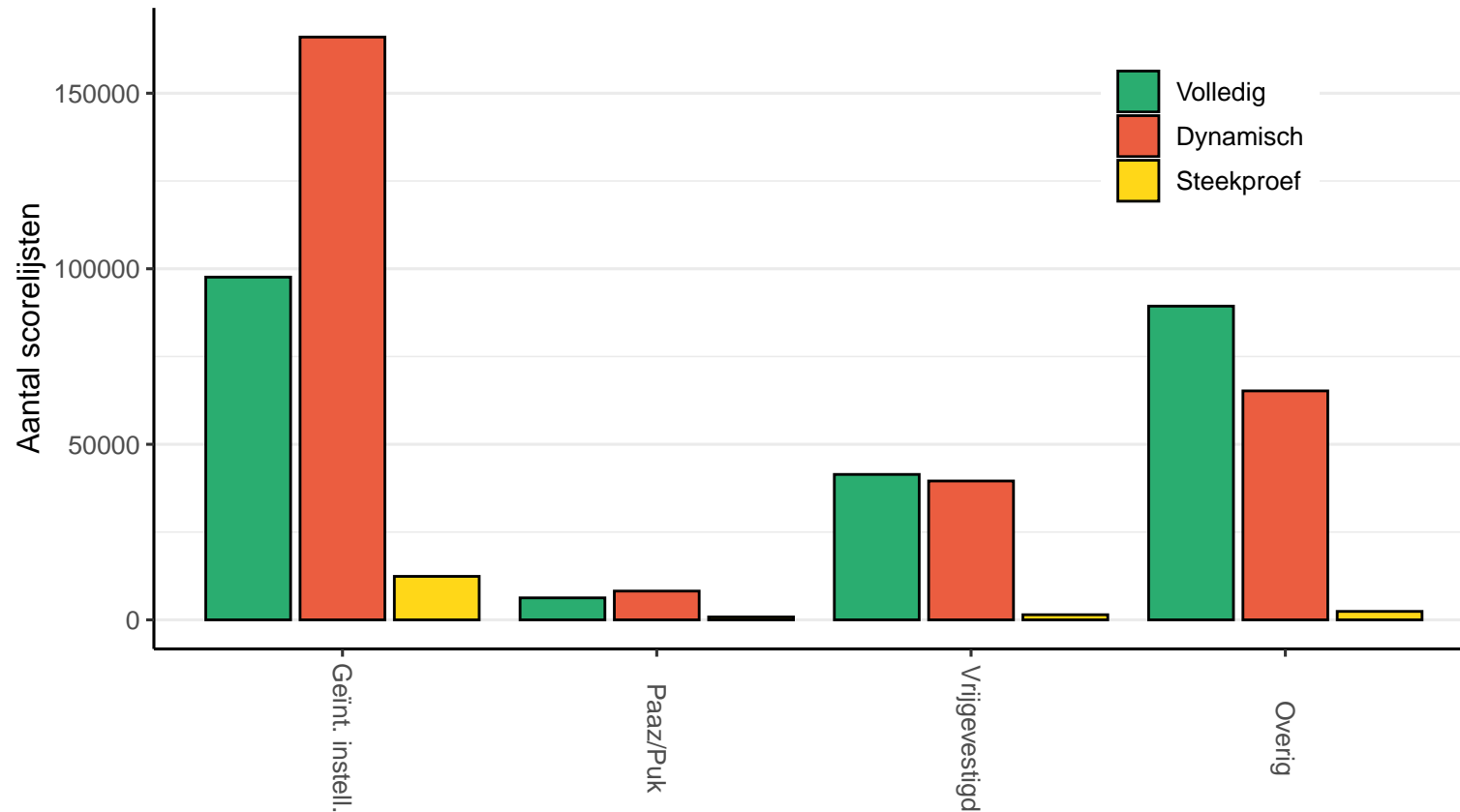
Methode	Aantal scorelijsten
Volledig	234667
Dynamisch	278955
Steekproef	17063
Totaal	530685

In tabel 1 is het aantal ontvangen scorelijsten te zien per methode van typeren. Er zijn bij benadering evenveel volledige en dynamische (verkorte) zorgvraagtyperingen uitgevoerd. Zoals te verwachten valt ongeveer 5% van de dynamische scorelijsten in de steekproef, zodat er voor deze patiënten ook een volledige zorgvraagtypering is uitgevoerd.

### 2.1.2 Aantal per type aanbieder

Het is van belang te weten bij welk type zorgaanbieders de gegevens vandaan komen. De aanbieders zijn voor deze analyse ingedeeld in vier types, gebaseerd op de begincijfers van AGB-code waarop de gegevens zijn aangeleverd. De rest van de AGB-code versleutelt de NZa om herleidbaarheid van de gegevens nog verder te verkleinen. De Geïntegreerde instellingen hebben AGB-codes beginnend met '0629', de vrijgevestigde aanbieders hebben AGB-codes beginnend met '03' of '94', Paaz/Puk'en hebben AGB-codes beginnend met '5454' en de overige aanbieders zijn alle overige AGB-codes.

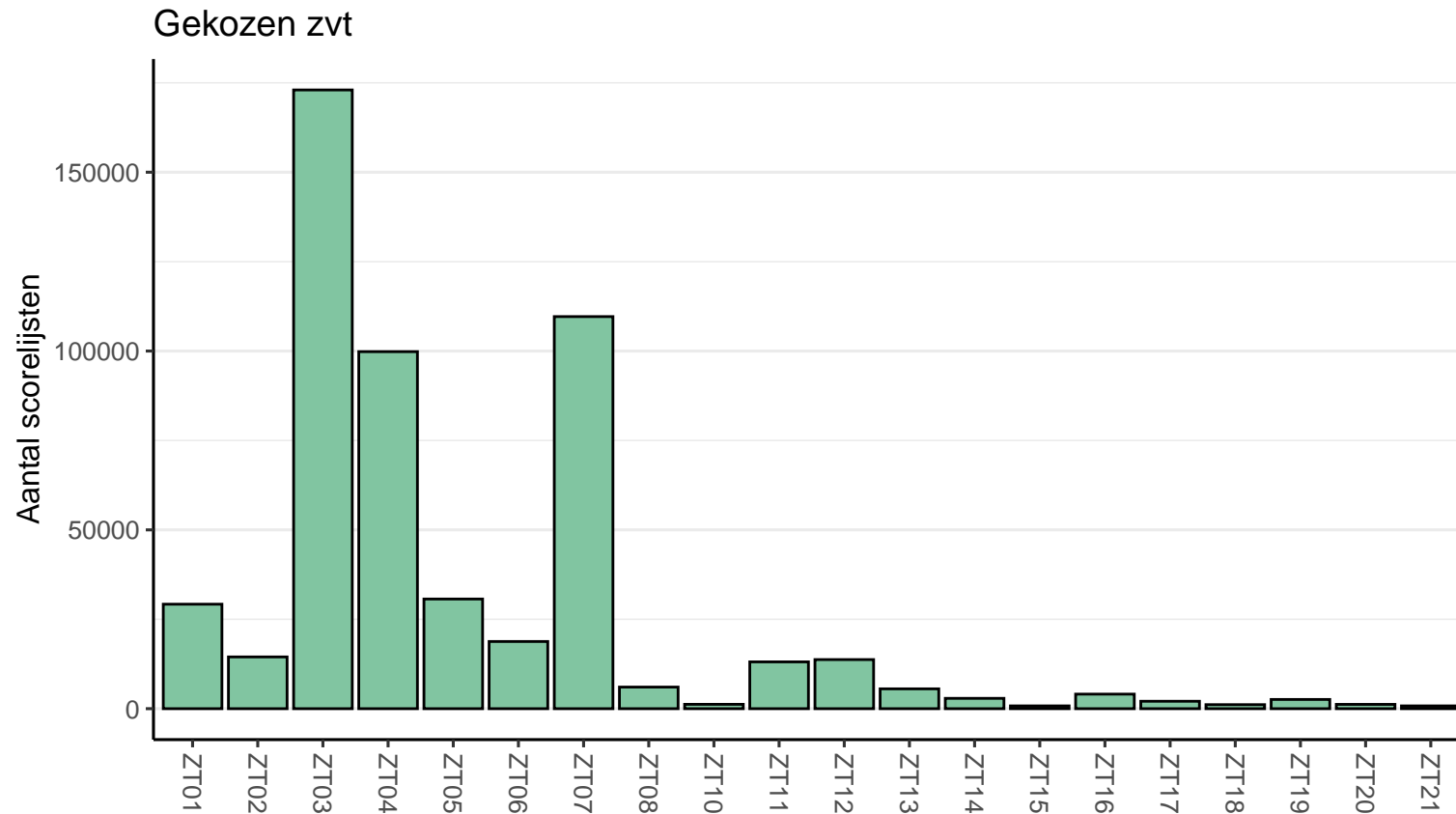
Figuur 1: Aantal scorelijsten per methode en type aanbieder



### 2.1.3 Aantal per gekozen zorgvraagtype

Sommige zorgvraagtypes komen vaker voor dan andere. Vooral zorgvraagtype ZT03, ZT04 en ZT07 komen vaak voor. Dit kan een reflectie zijn van de patiëntenpopulatie (common mental disorders komen vaak voor), maar kan ook (deels) het gevolg zijn van de voorspellingen van de algoritmes of doordat patiënten met voorheen complexe problematiek en die zijn gestabiliseerd na (her)typering worden doorbehandeld met het label van een relatief licht zorgvraagtype.

Figuur 2: Aantal voorspeld en gekozen zorgvraagtypes



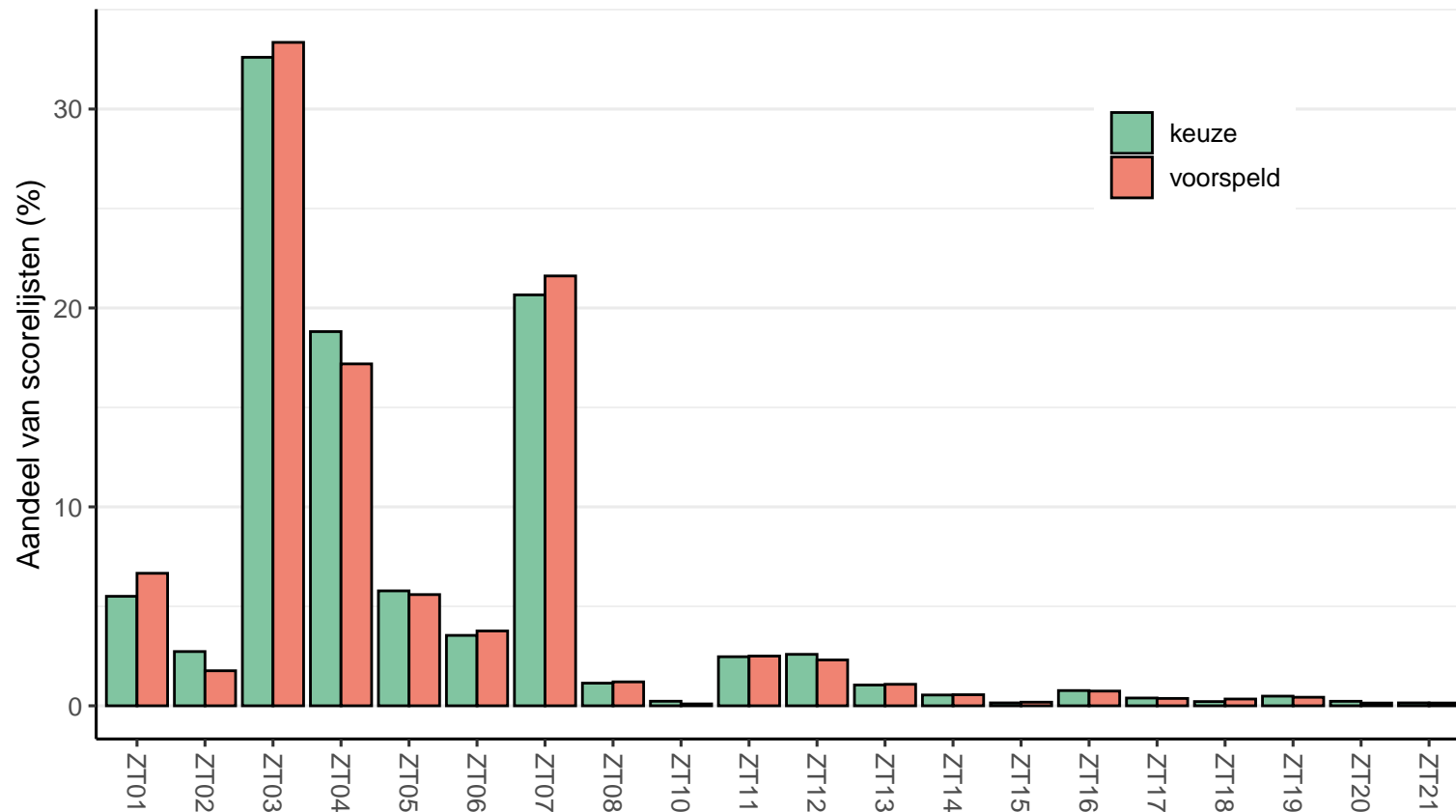
## 2.2 Vergelijken keuze en voorspelling

### 2.2.1 Aandeel voorspeld en gekozen zorgvraagtype

Welk aandeel van de scorelijsten valt in een zorgvraagtype ten opzichte van het totaal? Dit figuur laat zien welke aandeel van de zorgvraagtyperingen wordt voorspeld en gekozen in een bepaald zorgvraagtype.

NB Er is niet altijd een voorspelling te doen door de algoritmes, bijvoorbeeld wanneer alle zorgvraagtypes zijn uitgesloten door de rode regels, of wanneer de beslisboom onder dynamische zorgvragtypering niet voldoende diepte heeft. In het totaal zijn de onvoorspelde scorelijsten uitgesloten.

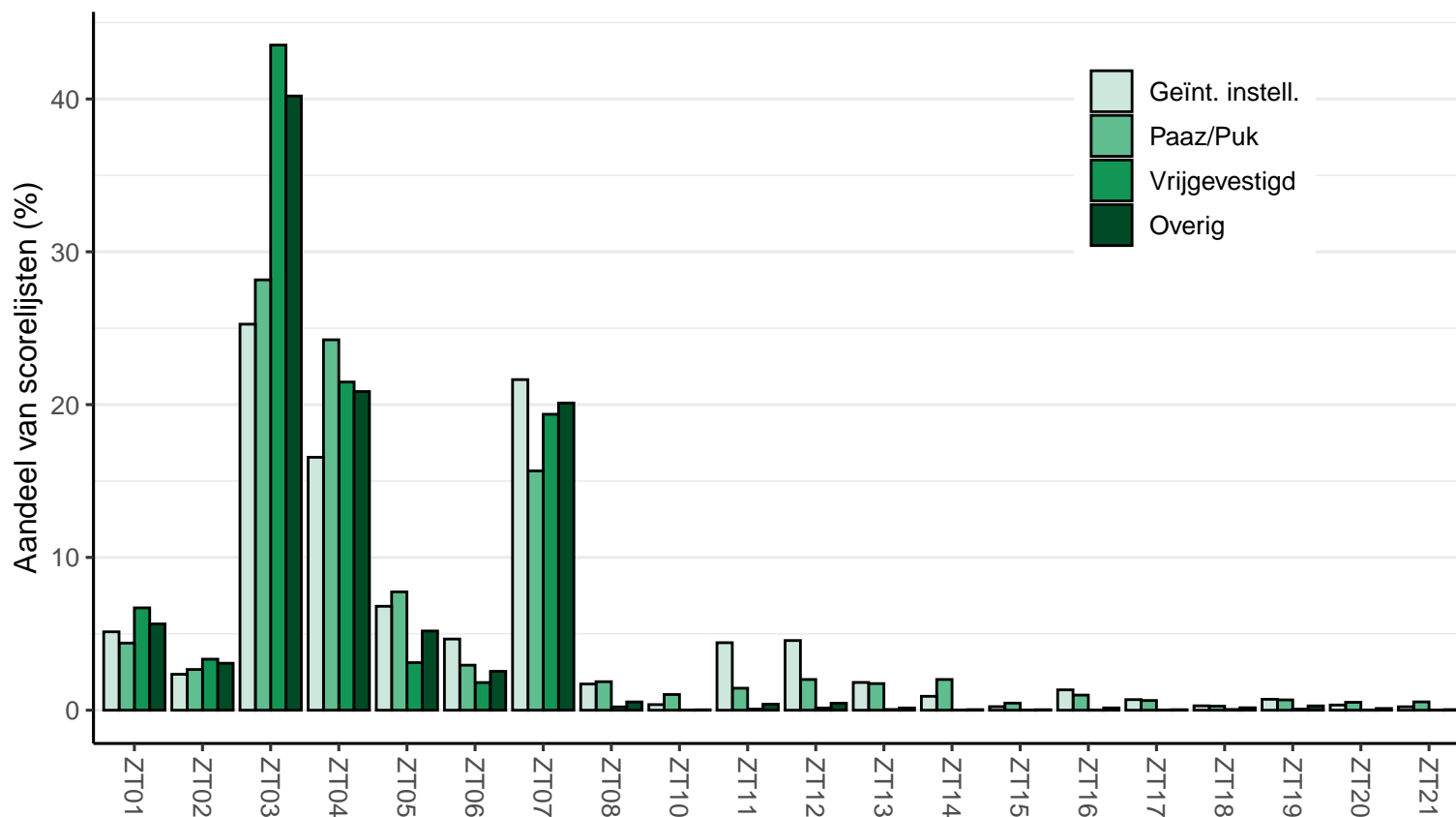
Figuur 3: Aandeel voorspeld en gekozen zorgvraagtypes



### 2.2.2 Aandeel gekozen per zorgvraagtype en type zorgaanbieder

Het is te verwachten dat de zorgvraagtypes niet uniform verdeeld zijn over de zorgaanbieders. Lichtere problematiek (Zorgvraagtype 1-4) zouden meer voor kunnen komen bij vrijgevestigde en overige aanbieders, terwijl de patiënten met psychotische problemen juist meer te verwachten zijn bij geïntegreerde instellingen en Paaz/Puk'en. Dit is ook terug te zien in de verdeling van de zorgvraagtypes tussen de types aanbieders hieronder.

Figuur 4: Aandeel zorgvraagtypes per type aanbieder



### 2.2.3 Relatieve keuze tov voorspelling

Het is mogelijk dat bepaalde zorgvraagtypes meer worden voorspeld dan dat ze worden gekozen, bijvoorbeeld omdat behandelaars niet graag een licht zorgvraagtype willen kiezen of bepaalde zorgvraagtypes juist als stigmatiserend zien. Dit kunnen we bekijken door het aandeel dat een bepaald zorgvraagtype wordt voorspeld te vergelijken met het aandeel dat het zorgvraagtype wordt gekozen. We delen hier door het aandeel voorspeld om te corrigeren voor relatief voorkomen. Waardes boven 0 betekenen dat een zorgvraagtype vaker wordt gekozen dan voorspeld, terwijl waarden onder 0 betekenen dat bepaalde zorgvraagtypes juist vaker worden gekozen. Het valt op dat zorgvraagtypes ZT01, ZT15 en ZT18 opvallend minder vaak worden gekozen dan voorspeld, terwijl ZT02, ZT10 en ZT20 juist vaker worden gekozen.

Figuur 5: Relatieve keuze tov voorspelling



## 2.2.4 Overeenkomst voorspelling en keuze

Om in te schatten of het verband tussen zorgvraag (scorelijst) en zorgvraagtype goed is bekijken we hoe het is met de voorspellingen door de modellen/algorithmes van volledige en dynamische zorgvraagtypering. De twee modellen bestaan voor elke hoofdgroep. Het is belangrijk te weten of deze vaak het zorgvraagtype voorspellen dat door de behandelaar wordt gekozen en of deze niet te vaak geen voorspelling kunnen doen. In tabel 2 is te zien dat behandelaren zowel bij de dynamische als volledige methode in 77%-87% van de scorelijsten hetzelfde zorgvraagtype kiest als dat het algoritme voorspelt. De voorspelling en keuze komen iets vaker niet overeen in hoofdgroep Z dan in X en Y. Naast het aandeel correcte voorspellingen is het nog van belang dat het algoritme überhaupt een voorspelling kan doen. In 13-21% van de gevallen kunnen de dynamische algoritmes dit niet, terwijl slechts 0-4% van de zorgvraagtyperingen met de volledige zorgvraagtypering geen voorspelling hebben.

De volledige modellen kunnen zelden geen voorspelling doen, dit is alleen het geval wanneer er een opvallende combinatie van hoofdgroep en scores wordt opgegeven, zoals geen cognitieve problemen (score 0 op item 4) binnen hoofdgroep Z. De dynamische modellen kunnen vaker geen voorspelling doen dan de volledige modellen. Dit is het gevolg van de versimpeling die de dynamische beslisboom inhoudt ten opzichte van de volledige modellen. Met de keuze voor het beginnen met een set basisvragen en het verdiepen van de dynamische modellen is het mogelijk het aandeel missende voorspellingen te verlagen.

Tabel 2: Aandeel scorelijsten waarbij de behandelaar het voorspelde zorgvraagtype koos per hoofdgroep

hoofdgroep	methode zorgvraagtypering	Aandeel keuze is voorspelling	Aandeel geen voorspelling
X	Dynamisch	0.86	0.13
Y	Dynamisch	0.83	0.21
Z	Dynamisch	0.80	0.16
X	Volledig	0.87	0.03
Y	Volledig	0.85	0.00
Z	Volledig	0.77	0.04

Tabel 3: Confusion matrix van voorspelde en gekozen zorgvraagtypes

Keuze	Geen	ZT01	ZT02	ZT03	ZT04	ZT05	ZT06	ZT07	ZT08	ZT10	ZT11	ZT12	ZT13	ZT14	ZT15	ZT16	ZT17	ZT18	ZT19	ZT20	ZT21
ZT01	4168	22709	590	1336	155	28	14	207	4	0	0	0	0	0	0	0	0	0	0	0	0
ZT02	949	4797	6682	1551	175	28	9	276	5	0	0	0	0	0	0	0	0	0	0	0	0
ZT03	8955	3162	987	148340	4799	287	100	6325	43	0	0	0	0	0	0	0	0	0	0	0	0
ZT04	13307	643	129	4748	73072	1593	637	5451	245	0	0	0	0	0	0	0	0	0	0	0	0
ZT05	3572	124	13	431	834	23606	625	1308	149	0	0	0	0	0	0	0	0	0	0	0	0
ZT06	2097	61	9	102	108	224	14590	1553	43	0	0	0	0	0	0	0	0	0	0	0	0
ZT07	6874	718	112	4827	3994	1162	2042	89318	577	0	0	0	0	0	0	0	0	0	0	0	0
ZT08	789	16	2	27	30	125	183	116	4751	0	0	0	0	0	0	0	0	0	0	0	0
ZT10	100	0	0	0	0	0	0	0	0	146	197	367	233	33	24	89	30	0	0	0	0
ZT11	1657	0	0	0	0	0	0	0	0	22	10701	562	43	17	10	22	55	0	0	0	0
ZT12	1636	0	0	0	0	0	0	0	0	121	991	9856	333	349	125	117	199	0	0	0	0
ZT13	556	0	0	0	0	0	0	0	0	55	21	77	4300	342	58	84	54	0	0	0	0
ZT14	379	0	0	0	0	0	0	0	0	28	61	137	268	1717	28	188	82	0	0	0	0
ZT15	78	0	0	0	0	0	0	0	0	3	21	24	18	12	596	17	8	0	0	0	0
ZT16	851	0	0	0	0	0	0	0	0	70	64	57	9	36	14	2933	33	0	0	0	0
ZT17	257	0	0	0	0	0	0	0	0	14	32	67	30	190	19	136	1332	0	0	0	0
ZT18	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1030	16	0	2
ZT19	230	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	540	1751	24	30
ZT20	236	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	262	606	47
ZT21	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	50	42	598

### 2.2.5 Keuze tov voorspelling - welk zorgvraagtype

De behandelaar kan een ander zorgvraagtype kiezen dan het model heeft voorspeld. Hoe de keuzes afwijken van de voorspelling kunnen we inzichtelijk maken met de confusion matrix in tabel 3. De rijen geven het gekozen zorgvraagtype. De kolommen laten de voorspelde zorgvraagtypes zien. Soms kon er geen voorspelling gedaan worden, dit is kolom 'Geen'. De velden bevatten het aantal aangeleverde scorelijsten waarvoor de combinatie van voorspelling en keuze geldt. Het is te zien dat in het geval van afwijking van de voorspelling meestal een zorgvraagtype 'in de buurt' van het voorspelde zorgvraagtype wordt gekozen, al komt het (zelden) voor dat bijvoorbeeld ZT08 wordt gekozen terwijl ZT01 werd voorspeld.



### 2.2.6 Relatieve keuze tov voorspelling

Met de absolute aantallen en grote verschillen in aantallen tussen de hoofdgroepen hierboven is het moeilijk in te schatten of er per zorgvraagtype veel wordt afgeweken. Daarom laten we hieronder per hoofdgroep ook het procentuele aandeel zien van de voorspellingen waarop wordt afgeweken. Het valt op dat voor de meeste zorgvraagtypes in meer dan 80% van de voorspellingen hetzelfde zorgvraagtype ook daadwerkelijk wordt gekozen. Dit geldt niet voor alle voorspelde zorgvraagtypes. ZT01 (69%) en ZT18 (62%) worden als lichtste mogelijkheid binnen de hoofdgroep minder vaak gekozen. In hoofdgroep Y worden vooral de zorgvraagtypes die een psychotische episode beschrijven minder vaak gekozen, wanneer deze worden voorspeld. Dit geldt voor ZT14 (64%), ZT15 (68%) en het sterkst voor ZT10 (32% gekozen in het geval van voorspelling). Anders dan item 6 wordt er in de zorgvraagtyperingsscorelijst ook niet gevraagd of er sprake is van een psychotische episode.

Tabel 4: Afwijking van voorspelling hoofdgroep X

gekozen zorgvraagtype	ZT01	ZT02	ZT03	ZT04	ZT05	ZT06	ZT07	ZT08
ZT01	0.70	0.07	0.01	0.00	0.00	0.00	0.00	0.00
ZT02	0.15	0.78	0.01	0.00	0.00	0.00	0.00	0.00
ZT03	0.10	0.12	0.92	0.06	0.01	0.01	0.06	0.01
ZT04	0.02	0.02	0.03	0.88	0.06	0.04	0.05	0.04
ZT05	0.00	0.00	0.00	0.01	0.87	0.03	0.01	0.03
ZT06	0.00	0.00	0.00	0.00	0.01	0.80	0.01	0.01
ZT07	0.02	0.01	0.03	0.05	0.04	0.11	0.85	0.10
ZT08	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.82

### Hoofdgroep X

Tabel 5: Afwijking van voorspelling hoofdgroep Y

gekozen zorgvraagtype	ZT10	ZT11	ZT12	ZT13	ZT14	ZT15	ZT16	ZT17
ZT10	0.32	0.02	0.03	0.04	0.01	0.03	0.02	0.02
ZT11	0.05	0.89	0.05	0.01	0.01	0.01	0.01	0.03
ZT12	0.26	0.08	0.88	0.06	0.13	0.14	0.03	0.11
ZT13	0.12	0.00	0.01	0.82	0.13	0.07	0.02	0.03
ZT14	0.06	0.01	0.01	0.05	0.64	0.03	0.05	0.05
ZT15	0.01	0.00	0.00	0.00	0.00	0.68	0.00	0.00
ZT16	0.15	0.01	0.01	0.00	0.01	0.02	0.82	0.02
ZT17	0.03	0.00	0.01	0.01	0.07	0.02	0.04	0.74

## Hoofdgroep Y

Tabel 6: Procentuele afwijking van voorspelling hoofdgroep Z

gekozen zorgvraagtype	ZT18	ZT19	ZT20	ZT21
ZT18	0.62	0.01	0.00	0.00
ZT19	0.33	0.84	0.04	0.04
ZT20	0.03	0.13	0.90	0.07
ZT21	0.02	0.02	0.06	0.88

## Hoofdgroep Z

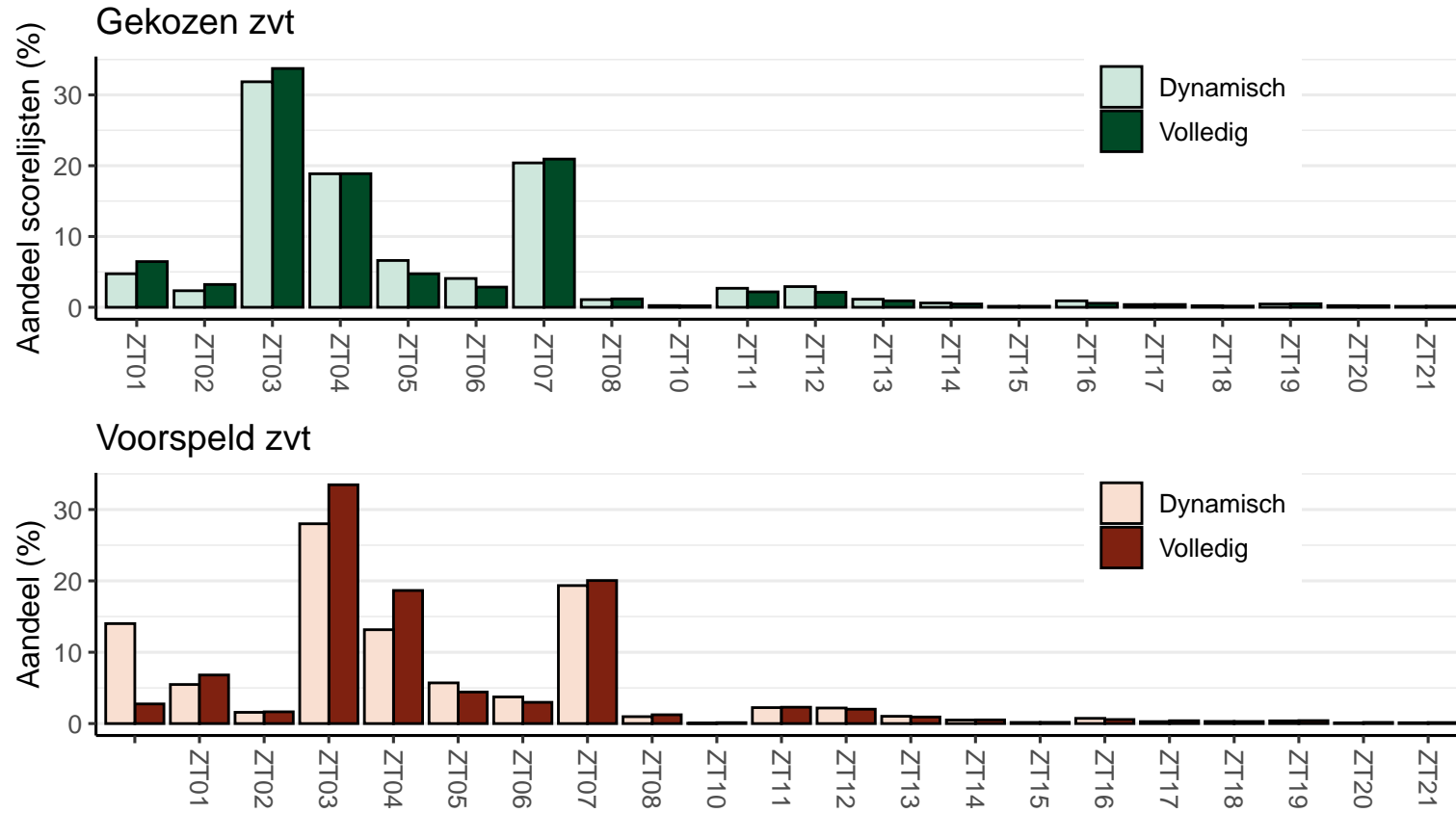
## 2.3 Volledig en dynamisch vergelijken

Een belangrijke vraag rond het vergelijken van de twee methodes van zorgvraagtypering was of de dynamische zorgvraagtypering goed genoeg functioneert om te blijven bestaan. Deze vraag is in de werkgroep data-analyse en de adviescommissie beantwoord met 'ja'. Hieronder volgen de analyses die hebben geleid tot deze conclusie.

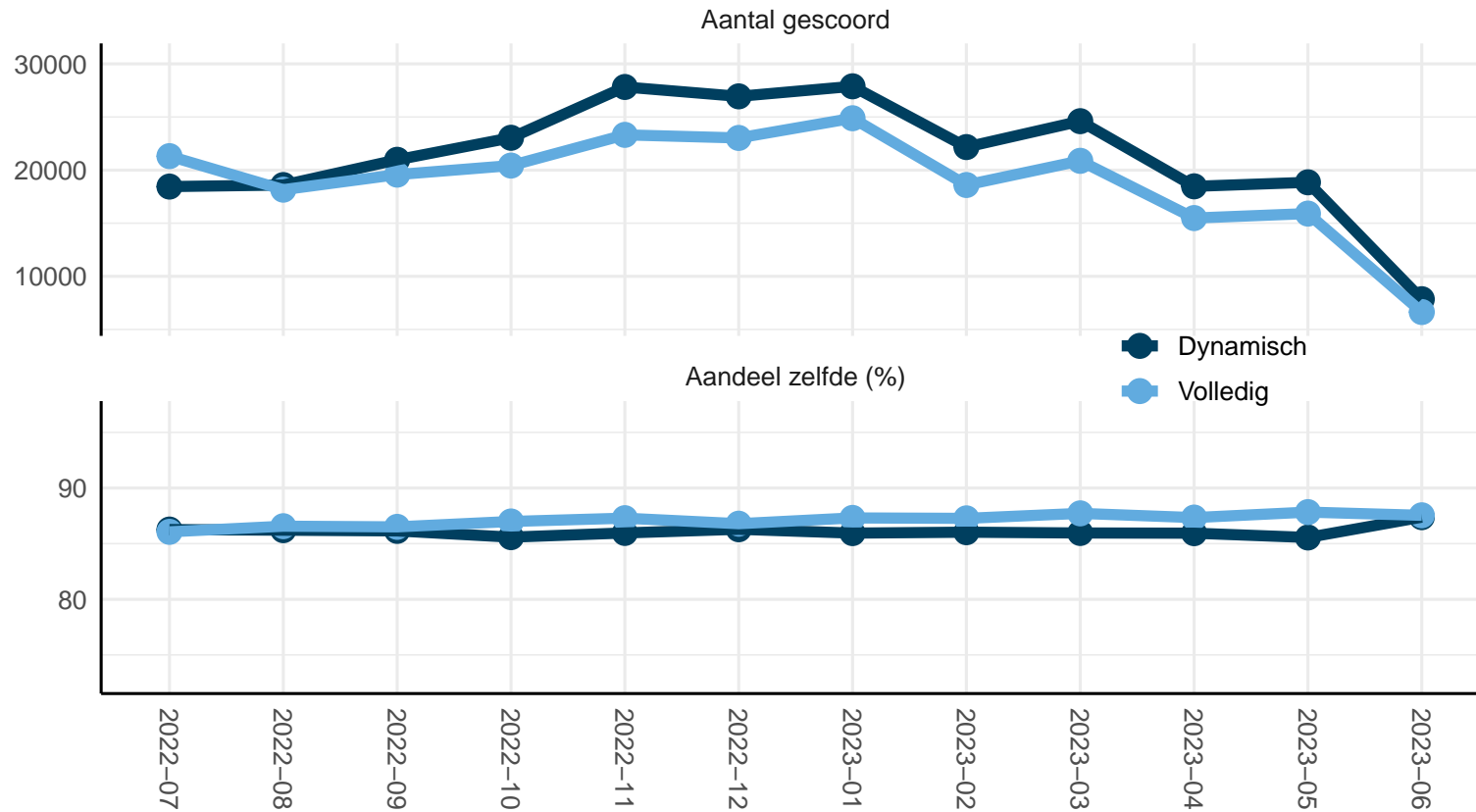
### 2.3.1 Keuze en voorspeld zorgvraagtype per methode

De vraag rees of bepaalde zorgvraagtypes vaker worden voorspeld met de dynamische of met de volledige zorgvraagtypering. In het figuur hieronder visualiseren we het aandeel van de aangeleverde scorelijsten waarvoor een bepaald zorgvraagtype is voorspeld of gekozen. Het is te zien dat bijvoorbeeld ZT03 en ZT04 een groter aandeel uitmaakt van de voorspellingen met de volledige dan met de dynamische zorgvraagtypering. Dit zou het gevolg kunnen zijn van een verschil in patiëntenpopulatie en beleid tussen verschillende zorgaanbieders. Het signaal uit de werkgroep praktijk en in figuur 1 is dat de minder complexe zorgvraagtypes vaker voorkomen bij vrijevestigde zorgaanbieders en overige instellingen en dat bij de Paaz/Puk en geïntegreerde instellingen vaker complexe en psychotische problematiek voorkomt. Het is aannemelijk dat verschillende soorten aanbieders er ook verschillend beleid op nahouden voor het uitvoeren van volledige of dynamische zorgvraagtypering.

Figuur 6: Aandeel voorspellingen en keuzes per zorgvraagtype en methode



Figuur 7: Ontwikkeling van aantal scorelijsten en aandeel correct over tijd

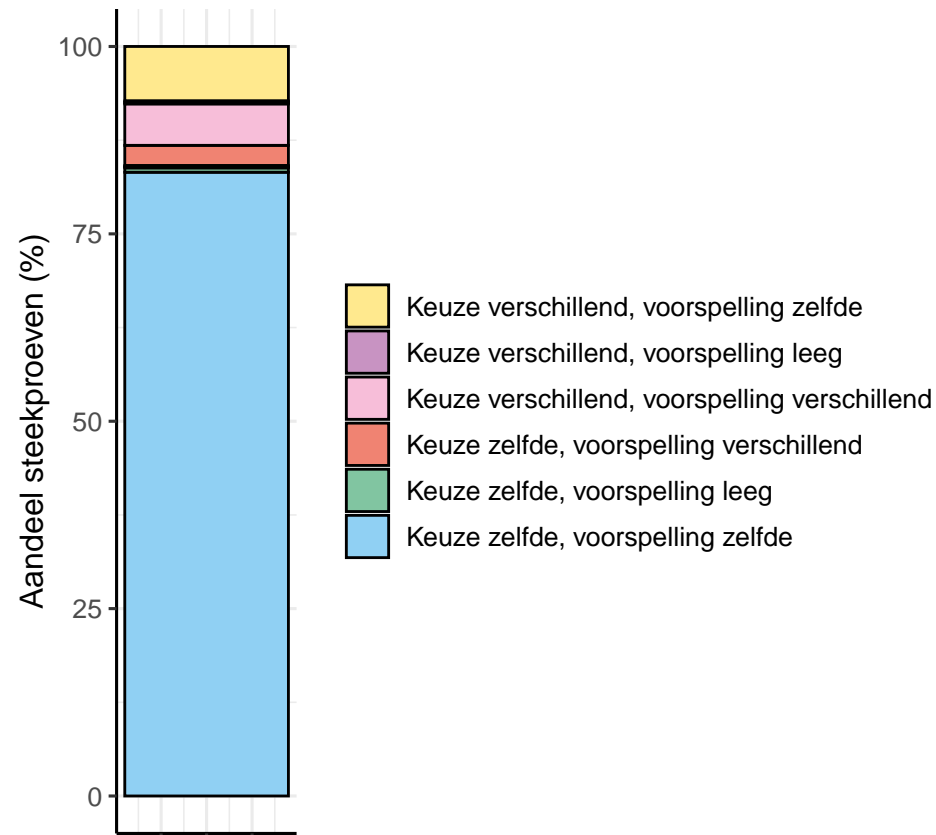


### 2.3.2 Leereffect

De verwachting uit de werkgroepen data en praktijk was dat aanbieders mogelijk over de tijd meer dezelfde keuze zouden gaan maken als de algoritmes voorspellen. Om dit te onderzoeken hebben we per maand waarin de zorgvraagtypering is uitgevoerd berekend hoe vaak de behandelaar van het algoritme af wijkt. In het bovenste figuur laten we zien hoeveel scorelijsten er per maand en per methode zijn aangeleverd en in het onderste figuur visualiseren we het percentage van de scorelijsten waarvoor de keuze hetzelfde was als de voorspelling. Uit deze analyses komt geen opvallend leereffect naar voren.

### 2.3.3 Steekproeven dynamische methode

Figuur 8



Voor 5% van de dynamische scorelijsten is ook een koppelbare volledige scorelijst ingevuld. We kunnen de uitkomsten van deze twee methodes vergelijken voor dezelfde zorgvraag. Het is hierbij interessant om te vergelijken of de twee methodes dezelfde voorspelling doen en of de behandelaar bij beide scorelijsten hetzelfde zorgvraagtype kiest. In veruit de meeste gevallen voorspellen het dynamische en volledige model hetzelfde zorgvraagtype, al komt het in 8.2 % van de gevallen voor dat er een verschil is tussen de twee voorspellingen. De keuze van de behandelaar is ook in veruit de meeste gevallen hetzelfde. Het is opvallend dat in 7.3 % van de gevallen de voorspelling bij de dynamische en volledige scorelijsten hetzelfde is, maar de keuze van de behandelaar verschilt. Dit kan bijvoorbeeld het gevolg zijn van een verminderde motivatie voor het invullen wanneer ook de volledige scorelijst



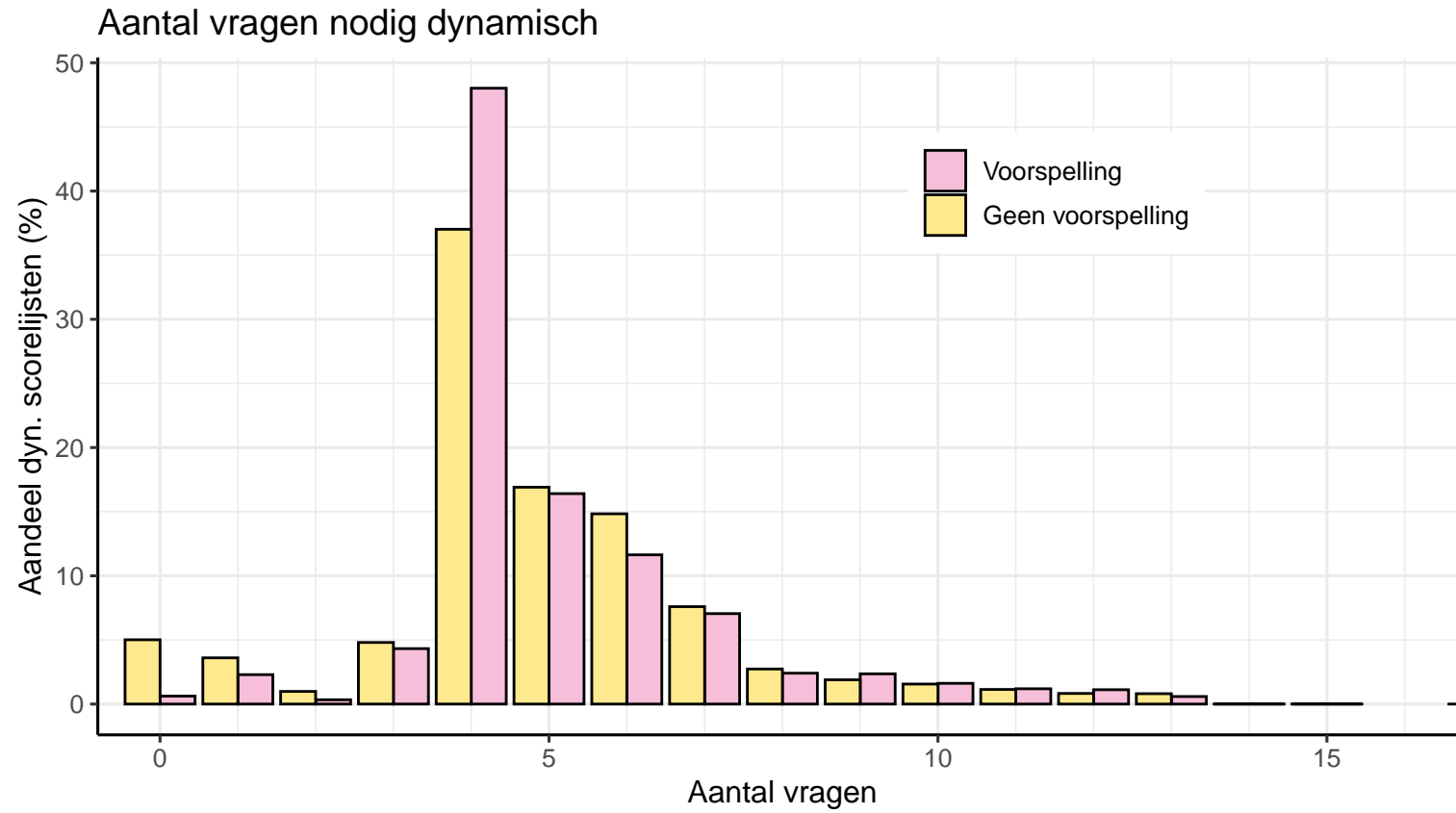
moet worden uitgevoerd, of het kan komen door nieuw inzicht na het denken over alnog ingevulde scores die ontbraken bij de dynamische zorgvraagtypering.

#### 2.3.4 Aantal items

Dynamische zorgvraagtypering is ontwikkeld om de administratieve lasten als gevolg van de zorgvraagtypering te verlagen. Dit is gedaan vanuit de gedachte dat het scoren van minder items gelijk staat aan minder administratie. Hieronder is te zien bij hoeveel items de dynamische scorelijst tot een resultaat is gekomen, ofwel met een voorspelling, ofwel zonder voorspelling. Het is te zien dat in de meeste gevallen 5-8 items voldoende is om tot een voorspelling te komen en dat een vergelijkbaar beeld te zien is voor de gevallen waarbij geen voorspelling gedaan kan worden.

NB Het komt voor dat er soms al na één of enkele items geen voorspelling gedaan kan worden. Dit is het gevolg van de rode regels (zie ook verderop) als een behandelaar bijvoorbeeld ernstige psychotische problemen (item 6) scoort in hoofdgroep X worden alle zorgvraagtypes uit die hoofdgroep op basis van de rode regels uitgesloten.

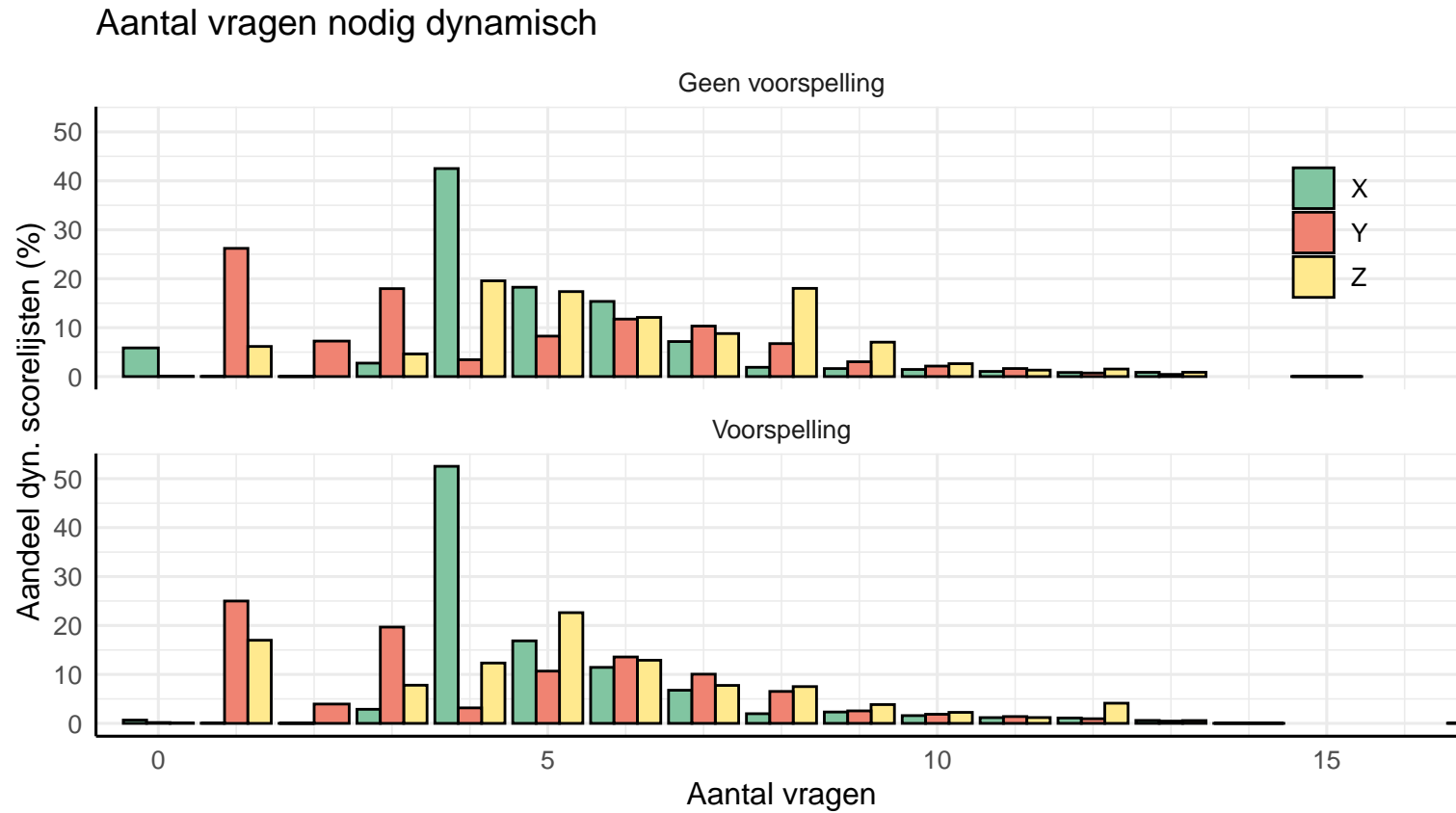
Figuur 9



### 2.3.5 Aantal items per hoofdgroep

Het is mogelijk het aantal items per hoofdgroep te bekijken. Hierbij is te zien dat er kleine verschillen zijn tussen de hoofdgroepen, maar ook dat voor alledrie de hoofdgroepen het waarschijnlijk is dat er een voorspelling kan worden gedaan na 5-8 items en dat het ook voor alle hoofdgroepen regelmatig voorkomt dat er geen voorspelling kan worden gedaan, soms al met weinig items.

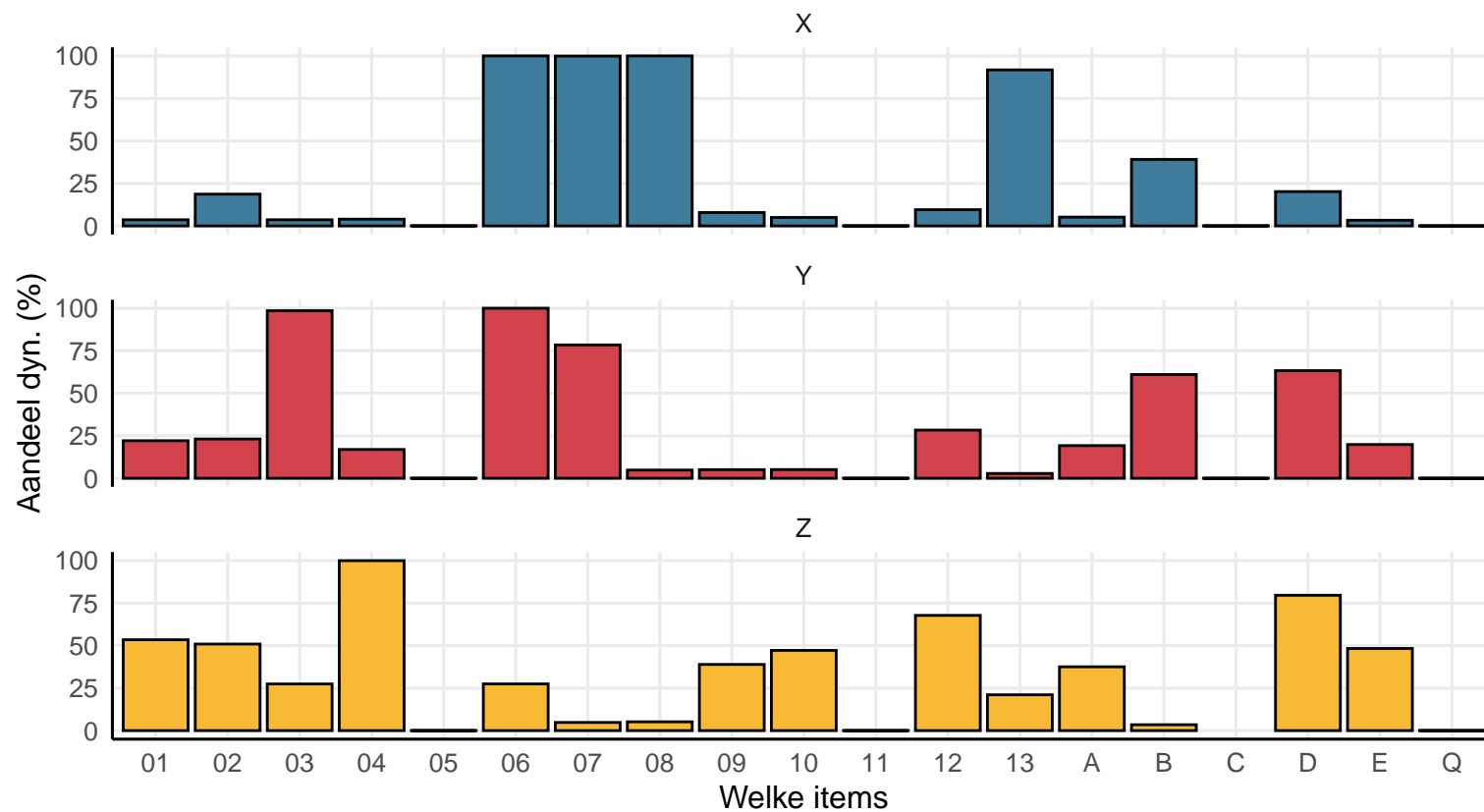
Figuur 10



### 2.3.6 Specifieke items dynamische methode

Het is belangrijk dat behandelaren het gevoel hebben dat de juiste items worden uitgevraagd bij dynamische zorgvraagtypering. In de werkgroepen praktijk en data zijn signalen besproken dat niet altijd de juiste items worden uitgevraagd. Daarnaast kan het zo zijn dat het dynamische model beter wordt als bepaalde items bij voor een hoofdgroep altijd worden gescoord. Daarom maken we hieronder inzichtelijk welke items vaak worden uitgevraagd. Het is te zien dat in hoofdgroep X altijd items 6 (hallucinaties en waanvoorstellingen), 7 (depressieve stemmingen) en 8 (overige psychische en gedragsproblemen) worden gescoord en dat vrijwel altijd item 13 (sterk onredelijke gedachten, niet psychotisch) wordt gescoord. Voor hoofdgroep Y wordt wederom altijd item 6 gescoord en in veel gevallen ook items 3 (alcohol- of middelengebruik), 7, B (zelfverwonding, voorgaand) en D (betrokkenheid/motivatie, voorgaand) gescoord. Voor hoofdgroep Z wordt altijd gevraagd naar item 4 (cognitieve problemen) en worden vaak ook items 12 (verbeteren van vaardigheden) en D gescoord. Het valt ook op dat items 5, 11, C en Q niet worden gescoord in dynamische zorgvraagtypering. Dit is te verklaren doordat deze items niet gebruikt worden in het voorspellen van het zorgvraagtype in het huidige model onder volledige zorgvraagtypering en het volledige model wordt zo goed mogelijk nagebootst door de dynamische zorgvraagtypering. De adviescommissie heeft besloten dat het opnieuw opbouwen van de dynamische zorgvraagtypering met per hoofdgroep een basisset vragen wenselijk is. De NZa gaat hiermee aan de slag.

Figuur 11



## 2.4 Zorgvraagtypering en behandelinzet

Bij de zorgvraagtyperingsgegevens is behandelinzet aangeleverd, per maand geaggregeerd in brede categorieën, om de herleidbaarheid van de gegevens zo klein mogelijk te maken. De gegevens zijn bijvoorbeeld teruggebracht tot het aantal kwartieren zorg dat een beroepscategorie in een setting heeft geleverd of het aantal verblijfsdagen met verzorgingsgraad A/B of C/D in een maand. Het is goed om te realiseren dat er veel zorgvraagtyperingen met in ieder geval één maand aan behandelinzet in de gegevens zitten, maar van elke maand extra er steeds minder zorgvraagtyperingen in de aangeleverde gegevens aanwezig zijn. Dit heeft twee oorzaken, ten eerste is het voor een zorgvraagtypering niet mogelijk om gegevens te bevatten van na de datum dat de gegevens zijn aangeleverd. Dit zorgt voor een geleidelijke terugloop, voor weinig trajecten is 11-12 maanden zorg aanwezig (zorgvraagtypering uitgevoerd in juli 2022) en dit loopt op tot (vrijwel) alle zorgvraagtyperingen waarvoor minimaal één maand aan behandelinzet aanwezig is (zorgvraagtypering uitgevoerd 30 dagen voor aanleveren of eerder). Ten tweede kan het zijn dat een nieuwe zorgvraagtypering is uitgevoerd, bijvoorbeeld na evaluatie van het zorgplan. Wanneer dit gebeurt wordt een nieuwe zorgvraagtypering aangemaakt en aangeleverd. Deze twee redenen zijn niet van elkaar te onderscheiden. Gegevens van herhaald uitgevoerde zorgvraagtypering zijn niet aan elkaar te koppelen.

De kosten bij de categorieën in de behandelinzet baseren we op het gewogen gemiddelde van de tarieven van alle gedeclareerde prestaties binnen de categorie op basis van de vektis-declaraties (peildatum 8 november 2023).

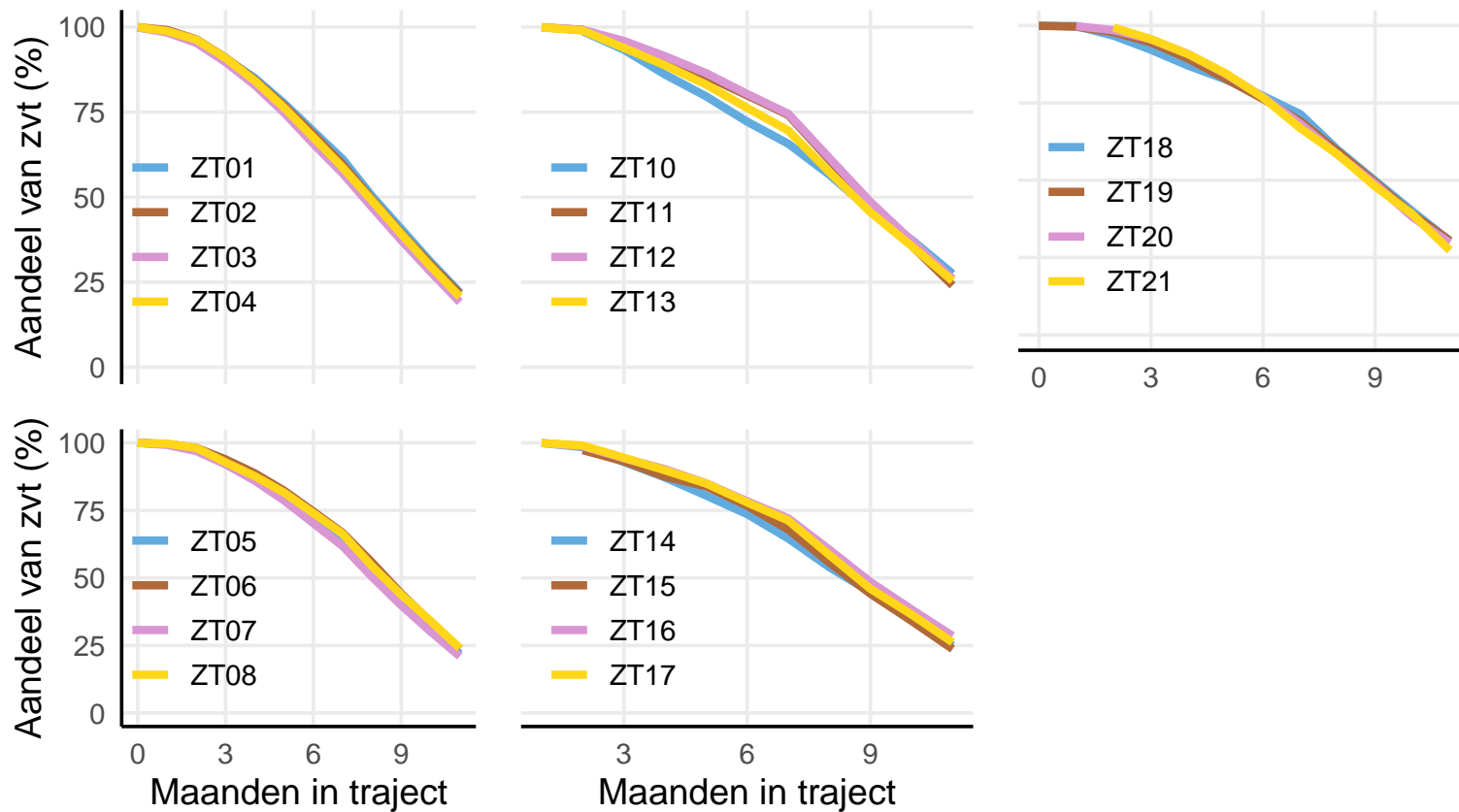
#### 2.4.1 Maanden in data

Zoals ook hierboven genoemd is voor de meeste zorgvraagtyperingen geen behandelinzet voor het volledige zorgtraject bekend. We kunnen alleen gebruik maken van de gegevens die zijn aangeleverd. Hieronder maken we inzichtelijk welk aandeel van de zorgvraagtyperingen nog behandelinzet uit een bepaalde maand heeft. We zien dat voor alle zorgvraagtypes minstens 50% van de scorelijsten behandelinzet over zes maanden na het scoren bevatten. Om robuuste conclusies te kunnen trekken op basis van voldoende gegevens selecteren we de eerste zes maanden van scorelijsten met minstens 6 maanden aan gegevens, specifiek moet er minimaal 6 maanden zitten tussen het uitvoeren van de zorgvraagtypering en het aanleveren van de gegevens.

NB niet altijd is er voor de volledige 6 maanden behandelinzet aanwezig, dit kan komen door afronding van de behandeling binnen 6 maanden, maar kan ook het geval zijn door het opnieuw uitvoeren van zorgvraagtypering.



Figuur 12



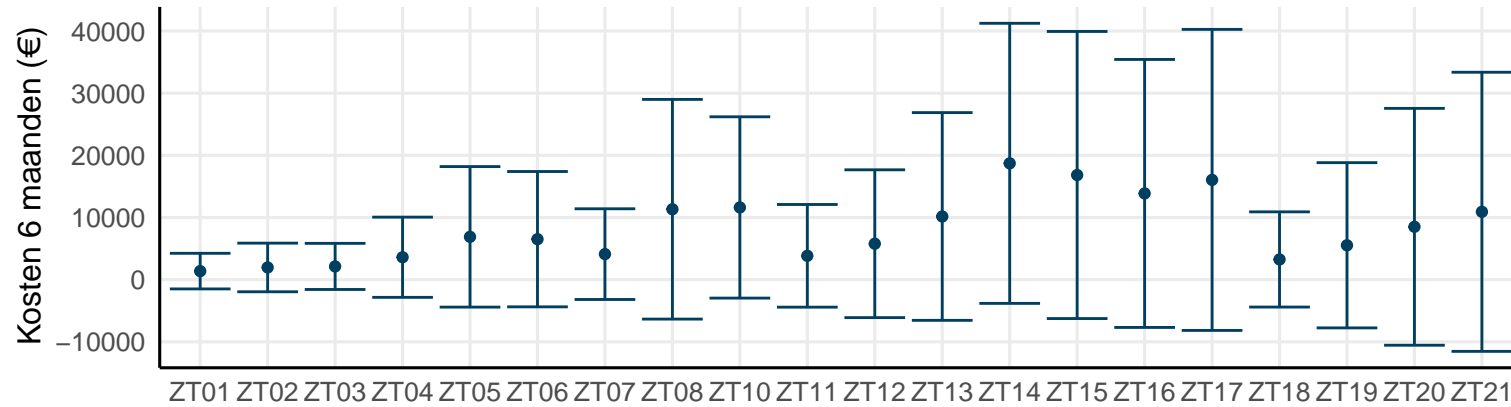
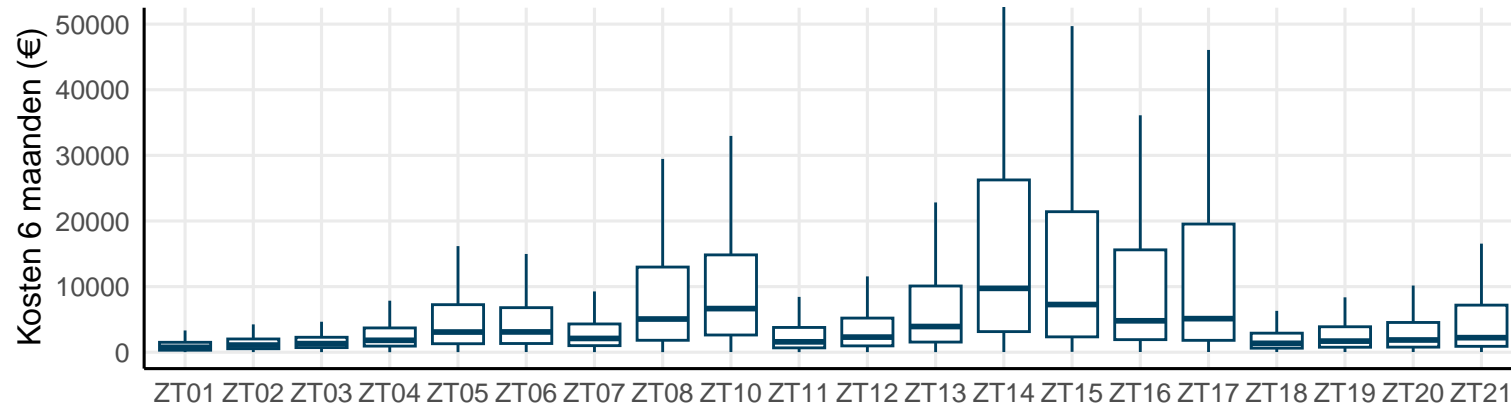
### 2.4.2 Verdeling kosten per zorgvraagtype

Per zorgvraagtype bekijken we de totale kosten over de eerste zes maanden. Hieronder doen we dat in twee vormen om meer detail in de verdeling te laten zien dan slechts het gemiddelde en standaardafwijking. In het bovenste paneel zien we de een boxplot, waarbij de middelste horizontale lijn de mediaan laat zien, de bovenste en onderste lijn het 25e en 75e percentiel, terwijl de verticale 'whiskers' 1,5 keer de interkwartielrange. In het onderste paneel zien we de gemiddelde kosten en de standaardafwijking. Het is te zien dat sommige zorgvraagtypes beduidend hogere gemiddelde kosten hebben dan andere zorgvraagtypes, maar het is ook duidelijk dat er voor elk zorgvraagtype grote spreiding van de kosten is.

In hoofdstuk 4 beschrijven we twee kwaliteitsparameters die we kunnen berekenen om het model van zorgvraagtypering te evalueren, de gemiddelde absolute fout en de verklaarde basisfout. Deze parameters zijn ook op deze gegevens te berekenen, doordat we een (onherleidbaar) versleutelde identicator hebben waarmee we kunnen zien welke trajecten door dezelfde zorgaanbieder zijn aangeleverd. Kijk voor de technische details van de berekeningen in hoofdstuk 4. De gemiddelde absolute fout is €1270 in deze dataset. De verklaarde basisfout is 39.1%, dit is het percentage van de basisfout (voorspelling op basis van gemiddelde) verklaard wordt met voorspelling op basis van het zorgvraagtype. Deze waardes houden we in het achterhoofd wanneer we kijken naar verfijning van het model door splitsing of samenvoeging van zorgvraagtypes.

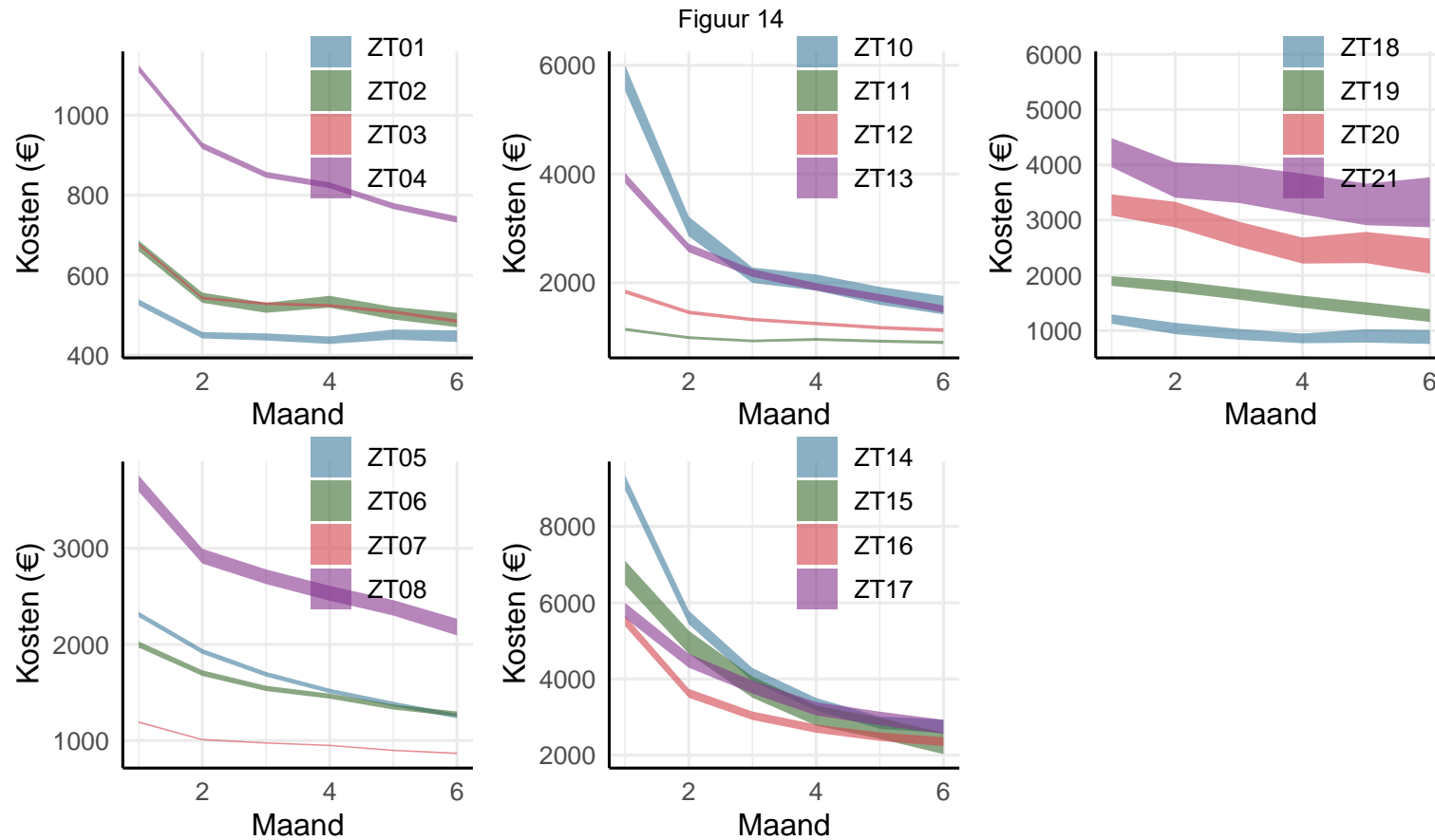
Uit privacyoverwegingen visualiseren we de individuele zorgtrajecten niet in de boxplots.

Figuur 13



### 2.4.3 Kosten per maand

Het is aannemelijk dat voor sommige patiënten de intensiteit van de behandeling varieert over tijd. Het verloop van kosten over tijd kunnen we bekijken voor de verschillende zorgvraagtypes. Dit doen we door per zorgvraagtype en per maand de spreiding op basis van de standaardfout van het gemiddelde te laten zien. Dit laat zien hoe goed het gemiddelde geschat is op basis van de gegevens, de schatting van het gemiddelde wordt beter met meer gegevens, dit weerspiegelt zich in de smalle gearceerde gebieden rond het gemiddelde van de zorgvraagtypes met grote aantallen.



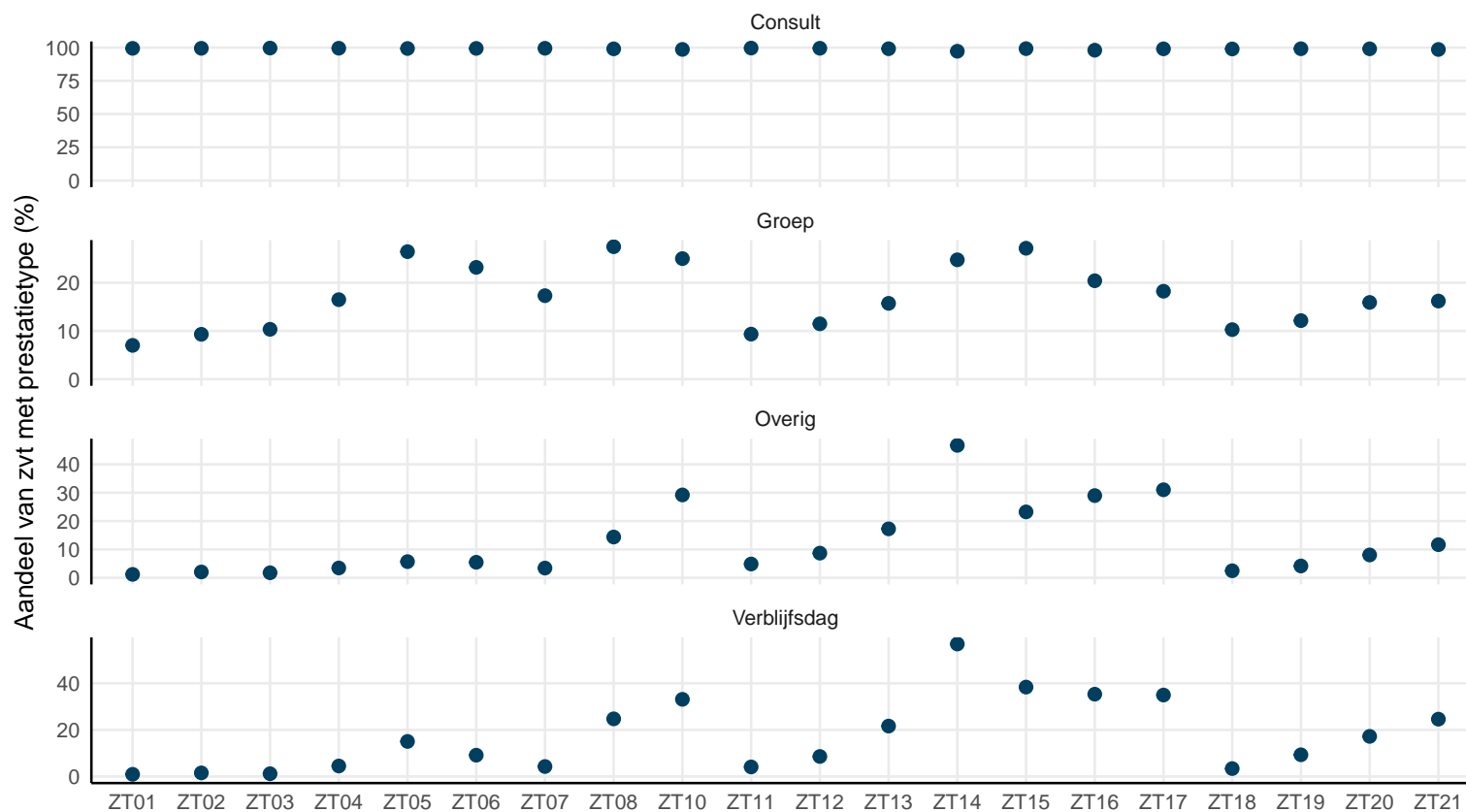
## 2.5 Zorgvraagtype en type prestatie

Welke prestaties worden geleverd per zorgvraagtype? De zorgvraag bepaalt, als het goed is, welke zorg passend is en welke zorg niet goed past. Dit leidt ertoe dat niet alle patiënten elk soort prestatie geleverd krijgen. Het is ook niet logisch dat een patiënt met een lichte zorgvraag, bijvoorbeeld zorgvraagtype ZT01-ZT04 klinisch zou worden opgenomen. We maken hieronder inzichtelijk welk percentage van de zorgvraagtyperingingen binnen een zorgvraagtype een bepaald type prestatie heeft ontvangen. We onderscheiden hierin voor de overzichtelijkheid vier verschillende categorieën, namelijk consulten, groepsconsulten, verblijfsdagen en een categorie overige prestaties. De kosten van toeslagen op consulten en verblijfsdagen schrijven we toe aan de bijbehorende prestatie categorie.

### 2.5.1 Aandeel zorgvraagtyperingen met prestatietype

Bij vrijwel alle zorgvraagtyperingen zijn consulten geleverd, terwijl maximaal ongeveer 25% van de zorgvraagtyperingen in een zorgvraagtype groepsconsulten ontvangen. Verblijfsdagen komen het meest voor bij de acuut psychotische zorgvraagtyperingen. De minst complexe zorgvraagtyperingen hebben het laagste aandeel groepsconsulten en verblijfsdagen. De overige prestaties zijn grotendeels acute ggz-prestaties en komen logischerwijs vooral voor bij de meest complexe zorgvraagtyperingen. Bij de minst complexe zorgvraagtyperingen komt af en toe ook een overige prestatie voor, in deze gevallen is dat vrijwel altijd MDO.

Figuur 15

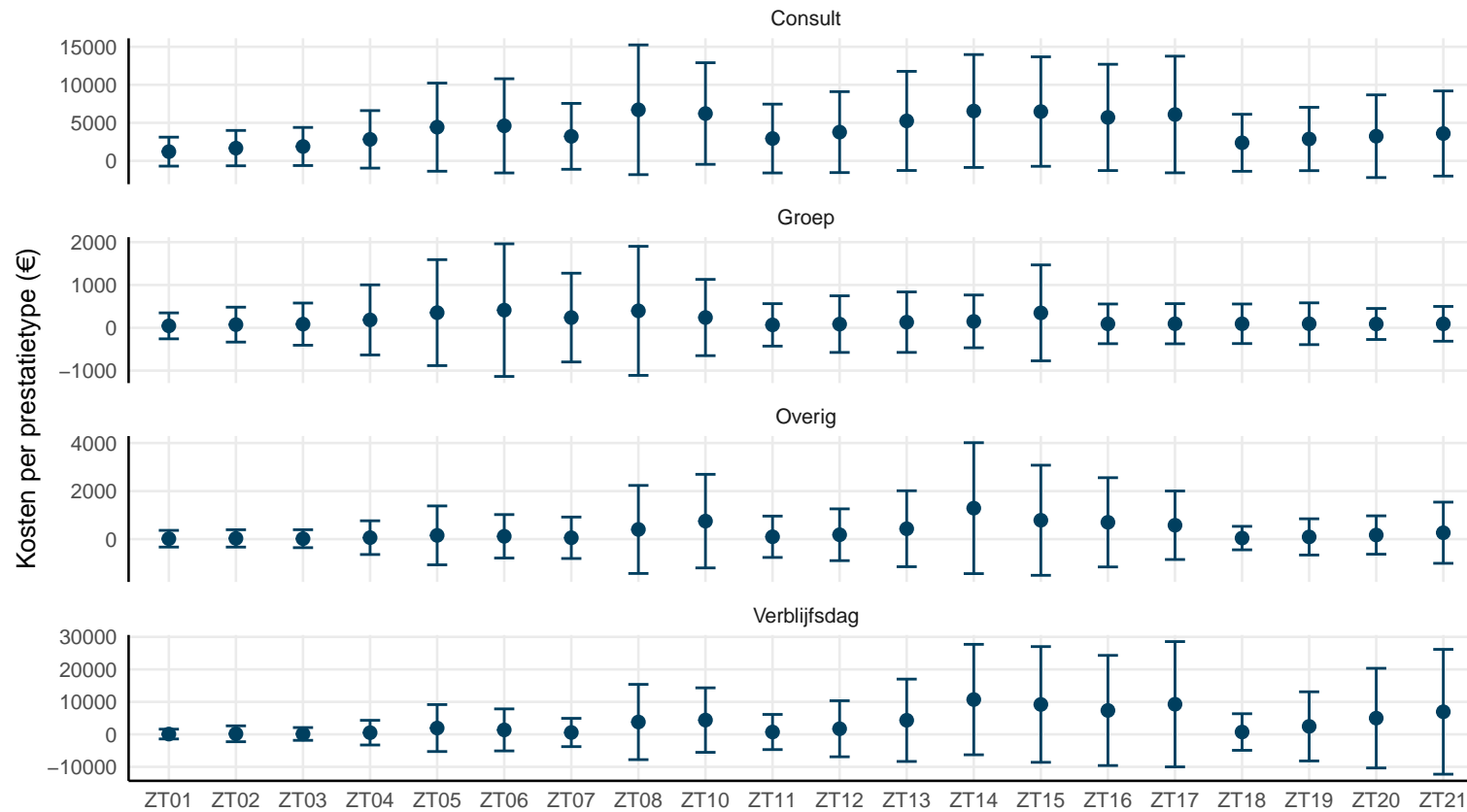


### 2.5.2 Kosten per prestatietype

De geleverde prestaties brengen kosten met zich mee, deze rekenen we toe naar de verschillende categorieën. Dat kan op twee manieren, namelijk door de gemiddelde kosten te nemen van zorgvraagtyperingen waarin de prestatie voorkomt óf door de gemiddelde kosten te laten zien voor alle zorgvraagtyperingen.

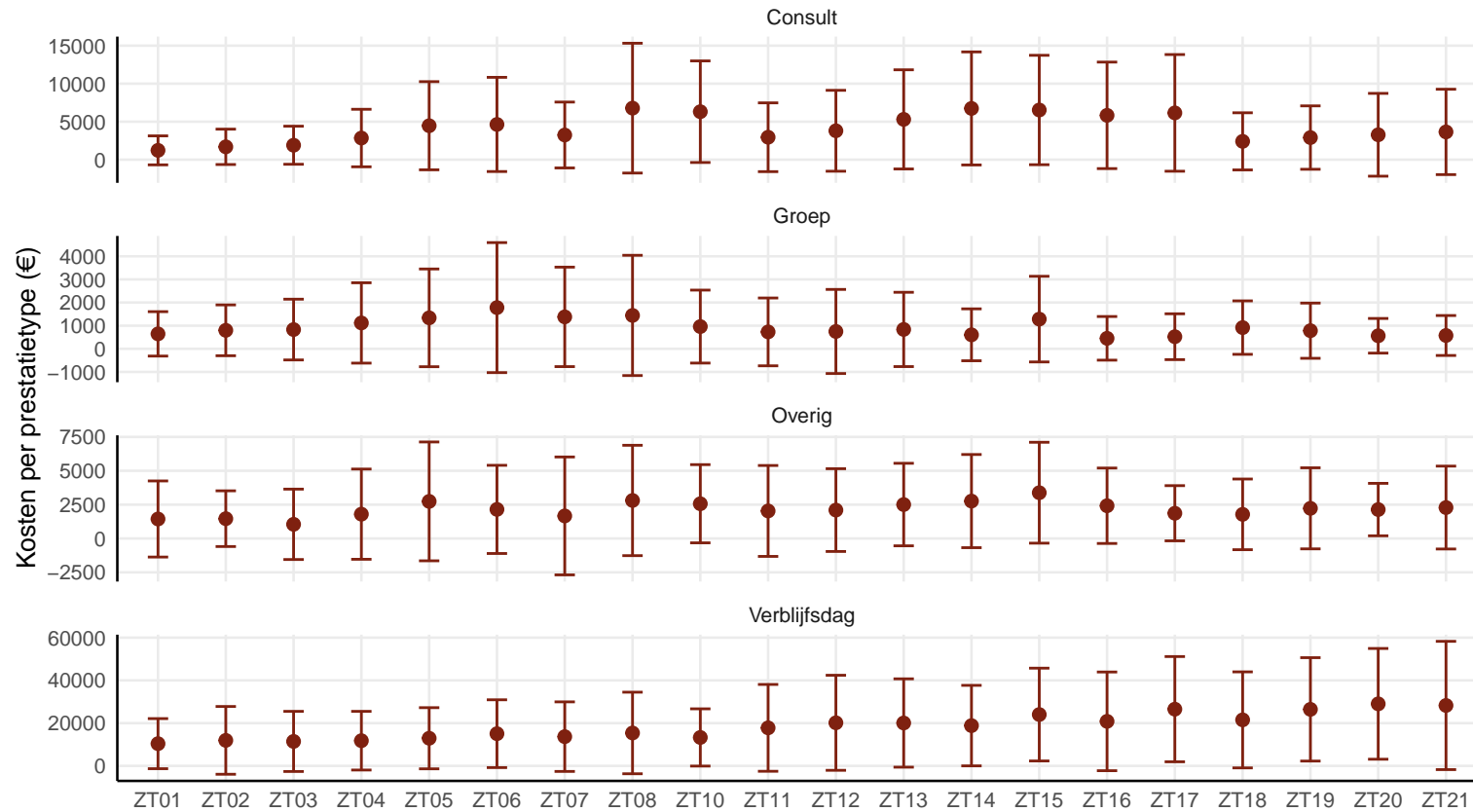
**Kosten per prestatietype - alle** Hieronder laten we de gemiddelde kosten zien, waarin alle zorgvraagtyperingen binnen een zorgvraagtype zijn meegenomen. Dit betekent dat er veel zorgvraagtyperingen voorkomen met kosten van 0 voor alle prestatietypes anders dan consulten.

Figuur 16



**Kosten per prestatietype - met prestatietype** Hieronder laten we de gemiddelde kosten zien alleen van zorgvraagtyperingen waarin het prestatietype voorkomt. Op deze manier kan worden ingeschat hoeveel kosten worden gemaakt wanneer een prestatietype (bijvoorbeeld verblijf) wordt ingezet. Dit betekent dat de gemiddelde kosten in dit figuur hoger zijn dan wanneer alle zorgvraagtyperingen binnen het zorgvraagtype worden meegenomen, zoals in 16.

Figuur 17





## 2.6 Verbeteren algoritmes

De wens is om de algoritmes onder de volledige zorgvraagtypering aan te passen aan de Nederlandse situatie. Het is hierbij belangrijk dat de methodiek niet verandert, om grote ICT-implementatiekosten te voorkomen. Daarom schatten we de volledige algoritmes op dezelfde manier als dat in Engeland is gebeurd de huidige algoritmes, namelijk met een Linear Discriminant Analysis (LDA). Voorwaarde voor implementatie van vernieuwde algoritmes is dat de kwaliteit van de voorspellingen goed blijft. Bij herschatten van de volledige algoritmes moeten uiteraard ook de dynamische algoritmes worden vernieuwd, zodat de twee methodes goed op elkaar aangesloten blijven. Wanneer er aanpassingen worden doorgevoerd aan de indeling in zorgvraagtypes moeten de algoritmes ook opnieuw worden geschat.

### 2.6.1 Afwijking voorspelling van keuze

In tabellen 2, 4, 5 en 6 is te zien hoe het oude algoritme voorspelt. De keuze en voorspelling zijn vaker gelijk dan verwacht op basis van de pilots met het zorgclustermodel. Waarschijnlijk zijn de behandelaren beïnvloed door het advies van het algoritme in het EPD. De goede prestatie van de oude algoritmes maken het moeilijk om deze veel te verbeteren. Het is vooral belangrijk dat de algoritmes niet veel slechter presteren en niet vaak grote afwijkingen hebben van het gekozen zorgvraagtype.

Het valt op dat de het percentage correcte voorspellingen iets beter is met het oude model voor de meeste zorgvraagtypes. Opvallend is de slechtere prestatie van het nieuwe model voor ZT01, ZT06, ZT14, ZT15, ZT17 en ZT18. Het lijkt niet te lukken met het selecteren van andere items . Het is de verwachting dat de prestatie van het algoritme verbetert, wanneer behandelaren de voorspelling van het nieuwe algoritme als advies krijgen.

Het is voor de NZa belangrijk advies van de adviescommissie te krijgen of de partijen de nieuwe algoritmes voldoende goed vinden om deze te implementeren.

Tabel 7: Prestatie voorspelling oude en nieuwe model per zorgvraagtype

ZVT	Correct nieuw ( )	Correct oud ( )	Vershil oud en nieuw (procentpunt)
ZT01	58.0	92.3	-34.3
ZT02	65.1	44.1	21.0
ZT03	89.0	92.1	-3.1
ZT04	85.9	88.6	-2.7
ZT05	84.2	87.7	-3.5
ZT06	63.2	87.1	-23.9
ZT07	81.8	83.4	-1.6
ZT08	94.4	93.8	0.6
ZT10	20.3	21.5	-1.2
ZT11	81.2	94.0	-12.8
ZT12	83.8	84.1	-0.3
ZT13	85.3	84.3	1.0
ZT14	58.0	72.4	-14.4
ZT15	49.7	83.2	-33.5
ZT16	91.7	89.0	2.7
ZT17	69.1	76.0	-6.9
ZT18	66.8	97.6	-30.8
ZT19	82.0	72.1	9.9
ZT20	76.8	71.7	5.1
ZT21	77.5	78.4	-0.9

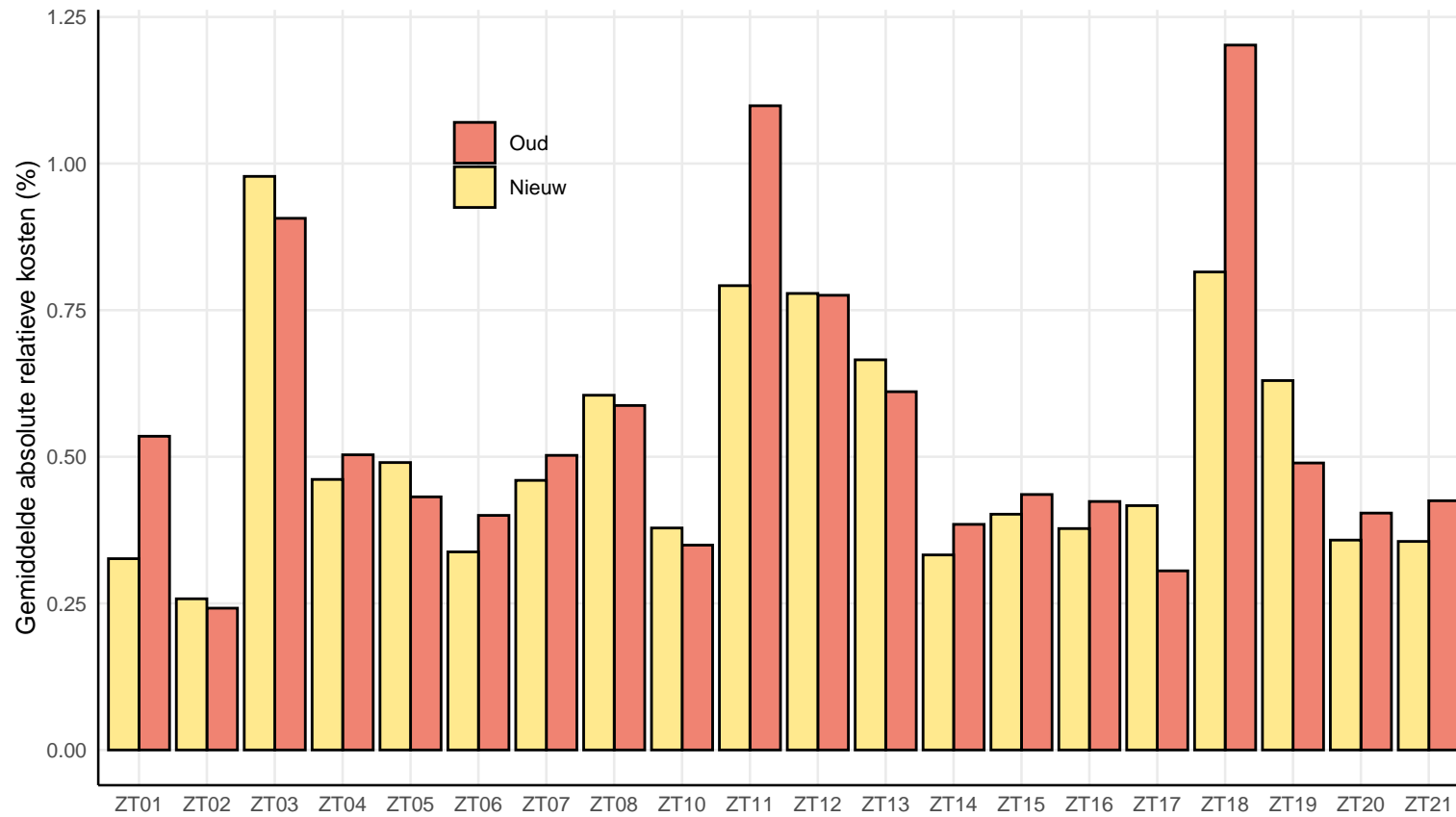
### 2.6.2 Mate van afwijking

De absolute afwijking tussen keuze en algoritme is mogelijk niet de meest relevante uitleeswaarde. Het gaat namelijk ook om hoe groot die afwijking is. Twijfel tussen ZT02 en ZT03 is immers eerder te verwachten dan twijfel tussen ZT02 en ZT08. Daarom gebruiken we de absolute relatieve kosten van het voorspelde en gekozen zorgvraagtype (absolute waarde van de gemiddelde kosten van gekozen zorgvraagtype min de gemiddelde kosten van voorspelde zorgvraagtype gedeeld door de gemiddelde kosten van het gekozen zorgvraagtype). De kosten halen we uit de declaraties en zoals gebruikt in de analyses verderop in dit document, waaronder figuur 68. Na koppelen van de ordening van kosten we het aantal zorgvraagtypes dat een afwijking omhoog of omlaag is gegaan. Deze methode van werken heeft geen oog voor de subtiliteit van de indeling in zorgvraagtypering op andere assen dan de kosten, helaas is het ons niet gelukt de volledige complexiteit goed te vatten in een begrijpelijke analyse.

Het komt voor dat zowel het oude als het nieuwe algoritme niet kan voorspellen wat de behandelaar kiest. Het is mogelijk dat de subtiliteit van de keuze van de behandelaar in deze gevallen niet goed door het rekenkundig model kan worden gevat, of het is mogelijk dat de registratie niet de kwaliteit heeft die we zouden willen. In beide gevallen is het niet zinvol de algoritmes op deze registraties te beoordelen. We sluiten deze dan ook uit en kijken naar de registraties waarbij één algoritme (oud of nieuw) wel goed voorspelt en het andere algoritme niet.

De gemiddelde absolute relatieve kosten zijn in veel gevallen vergelijkbaar, maar in sommige gevallen zijn de gemiddelde absolute relatieve kosten een stuk hoger, specifiek voor ZT01, ZT11 en ZT18. Het valt daarnaast op dat de zorgvraagtypes waarvoor het nieuwe algoritme duidelijk minder goed voorspelt, op ZT17 na, hogere gemiddelde absolute relatieve kosten hebben met het oudere algoritme. De grotere overeenkomst tussen voorspelling en keuze van de behandelaar door het oude algoritme lijkt dus ten koste te zijn gegaan van een grotere afwijking wanneer de keuze en voorspelling verschillen.

Figuur 18

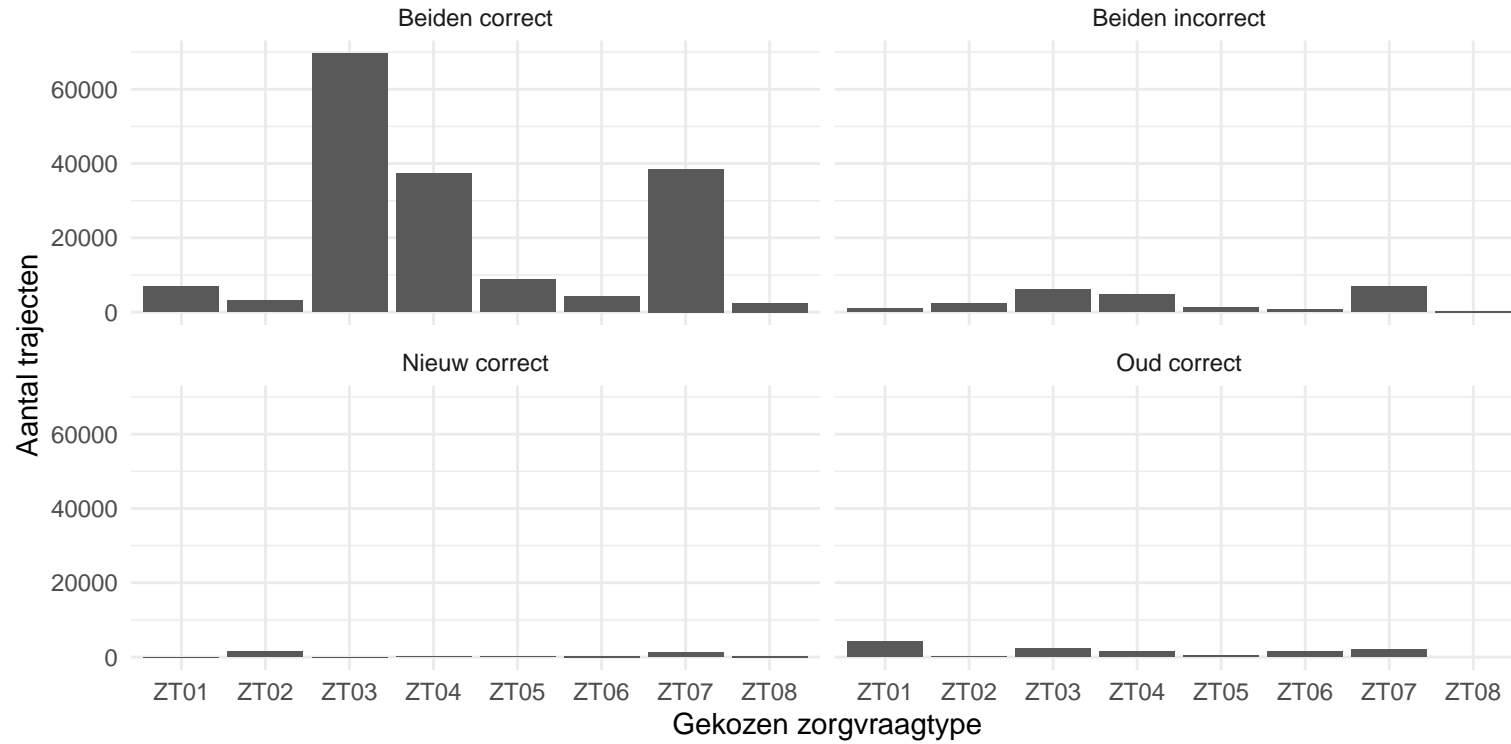


**hoofdgroep X: psychische stoornissen** Meestal geven het oude en het nieuwe algoritme dezelfde voorspellingen. In het merendeel komt dit overeen met de keuze van de behandelaar. Maar in onderstaande grafieken is zichtbaar dat er ook duidelijke verschillen zijn. Elke zorgvraagtypering is voor onderstaande figuren ingedeeld op basis van de oude en de nieuwe voorspelling en de daadwerkelijke keuze. De zorgvraagtypering is in het balke geplaatst van het zorgvraagtype dat is gekozen. Als zowel het oude als het nieuwe algoritme ook dit zorgvraagtype hadden voorspeld is de zorgvraagtypering geplaatst in het paneel 'Beiden correct'. Wanneer beide algoritmes iets anders voorspellen is de zorgvraagtypering geplaatst in het paneel 'Beiden incorrect'. Wanneer het nieuwe algoritme de juiste voorspelling heeft gedaan, maar het oude niet, dan is de zorgvraagtypering geplaatst in het paneel 'Nieuw correct'. Wanneer het omgekeerde het geval is, namelijk correcte voorspelling door het oude algoritme en incorrecte voorspelling door het nieuwe algoritme, dan is de zorgvraagtypering in het paneel 'Oud correct' geplaatst.

Wanneer we dieper de resultaten in duiken blijkt dit in absolute aantallen vaak voor te komen bij de grootste groepen en vooral wanneer ZT07 gekozen is, terwijl het algoritme ZT03 of ZT04 heeft voorspeld.

Figuur 19

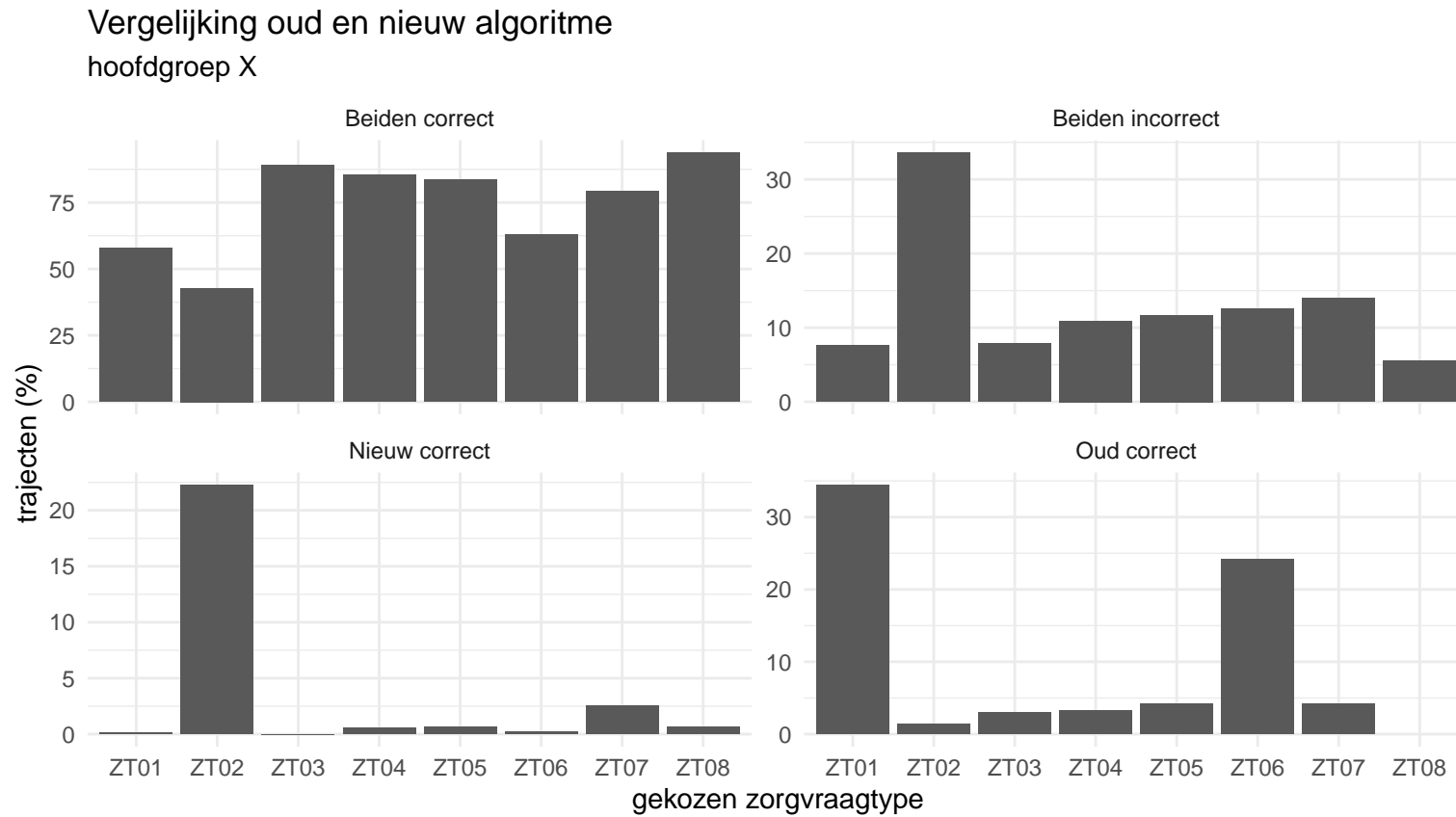
Vergelijking oud en nieuw algoritme  
hoofdgroep X



Wanneer we de gegevens bekijken als percentage van de daadwerkelijk gekozen zorgvraagtypes, dan scoort het nieuwe algoritme beter bij ZT02, terwijl het oude algoritme beter scoort in ZT01 en ZT06. Dat één algoritme beter scoort op ZT01 en het ander op ZT02 is niet onverwacht, omdat de (rekenkundige) centra van deze groepen vlakbij elkaar liggen. Het nieuwe algoritme adviseert sneller ZT02 op basis van de dezelfde scores, en neemt op deze manier ook veel ZT01 mee, terwijl het oude algoritme dit andersom lijkt te doen.

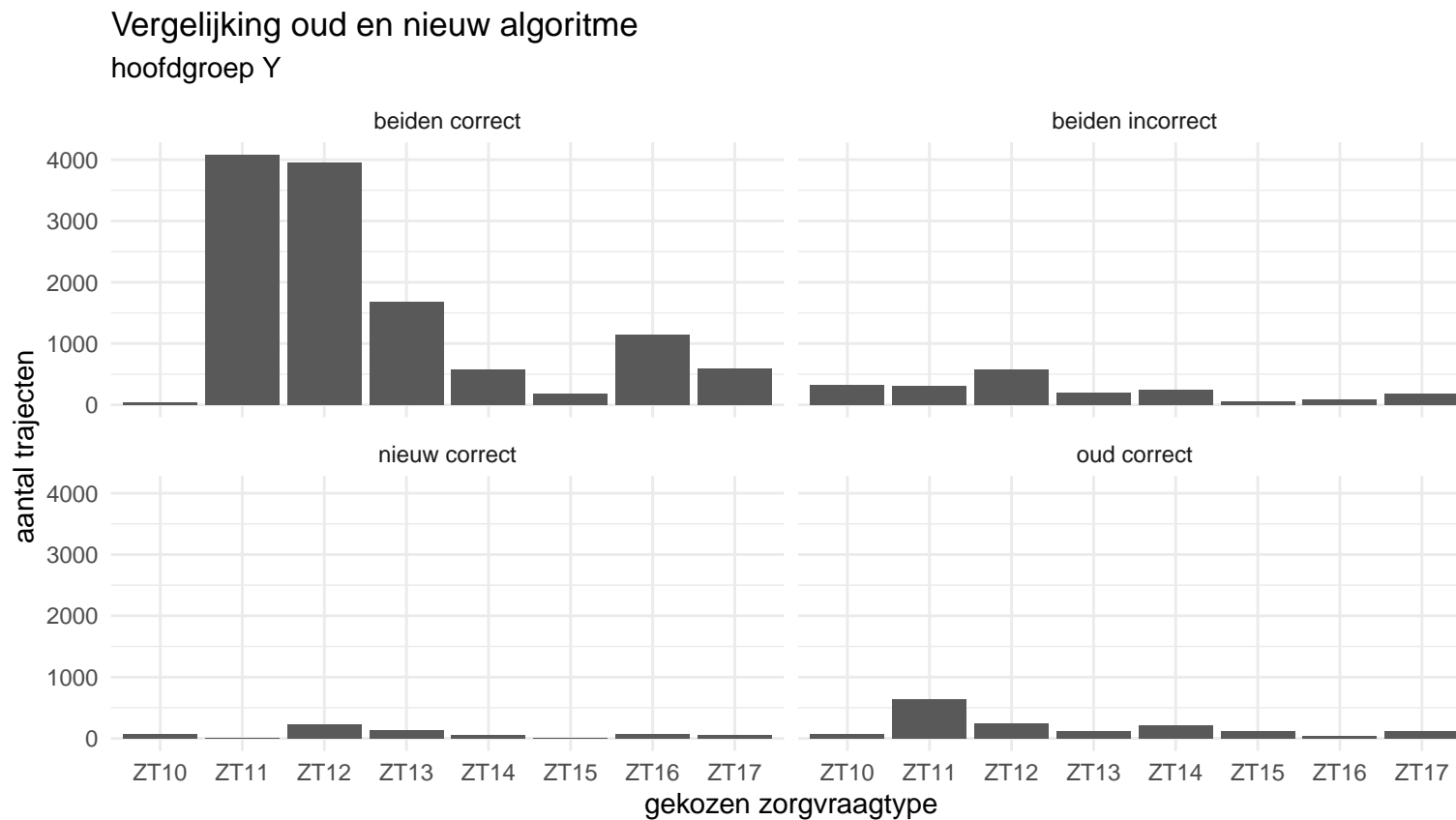
ZT06 komt met het nieuwe algoritme vaker terecht bij ZT07 dan voorheen. Aangezien ZT07 veel meer trajecten bevat dan ZT06, is dit niet meteen zichtbaar in de verminderde prestatie voor ZT07.

Figuur 20



**hoofdgroep Y: psychotische stoornissen** Meestal doen het oude en het nieuwe algoritme dezelfde voorspellingen. In het merendeel komt dit overeen met de keuze van de behandelaar, behalve voor ZT10. Maar in onderstaande grafieken is zichtbaar dat er ook duidelijke verschillen zijn. In absoluut aantal komt dit vaak bij de grootste groepen en vooral wanneer ZT11 gekozen is, terwijl het nieuwe algoritme ZT12 adviseerd.

Figuur 21

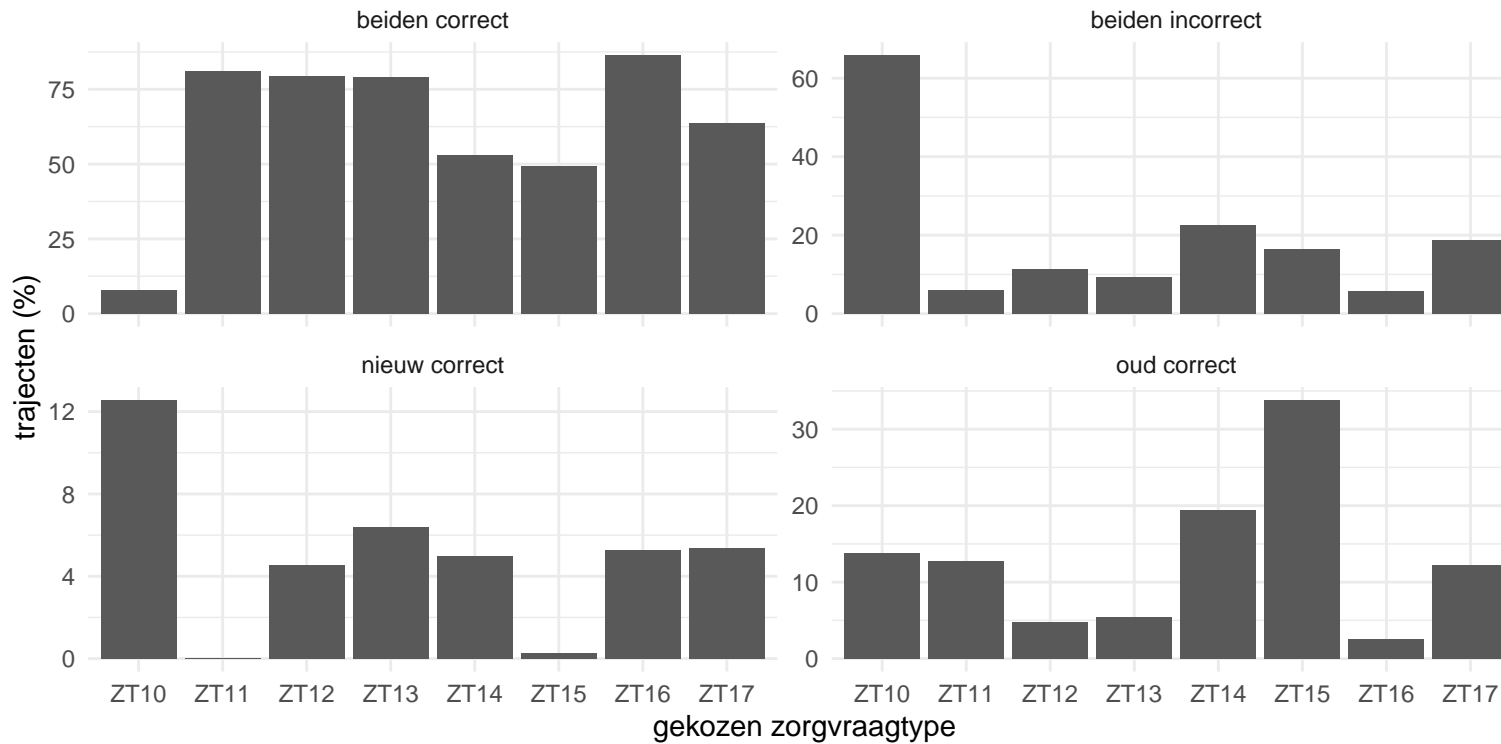




Procentueel ten opzichte van het aantal keuzes voor het zorgvraagtype scoort het oude algoritme beter voor zorgvraagtypes met psychotische episode (ZT14, ZT15) en ZT11 en ZT17.

Wanneer we dieper de gegevens induiken blijkt het volgende. ZT10 wordt door de algoritmes verward met allerlei niet-psychotische types (ZT11, 12, 13, 16). Ook ZT14 en ZT15 worden vaak bij niet acute-zorgvraagtypes ingedeeld door het nieuwe algoritme.

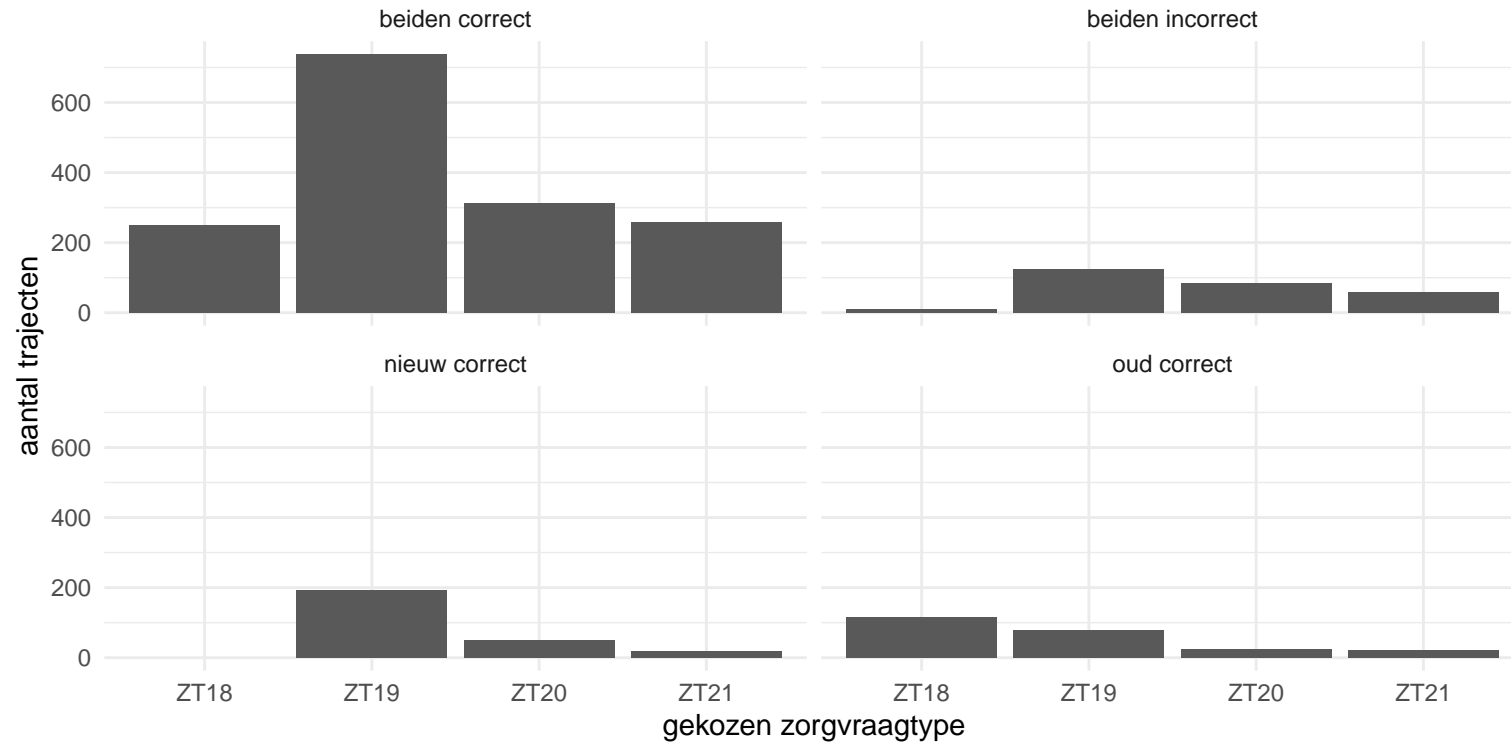
### Vergelijking oud en nieuw algoritme hoofdgroep Y



**hoofdgroep Z: cognitieve beperkingen** Meestal doen het oude en het nieuwe algoritme dezelfde voorspellingen. In het merendeel komt dit overeen met de keuze van de behandelaar. Maar in onderstaande grafieken is zichtbaar dat er ook duidelijke verschillen zijn.

Figuur 22

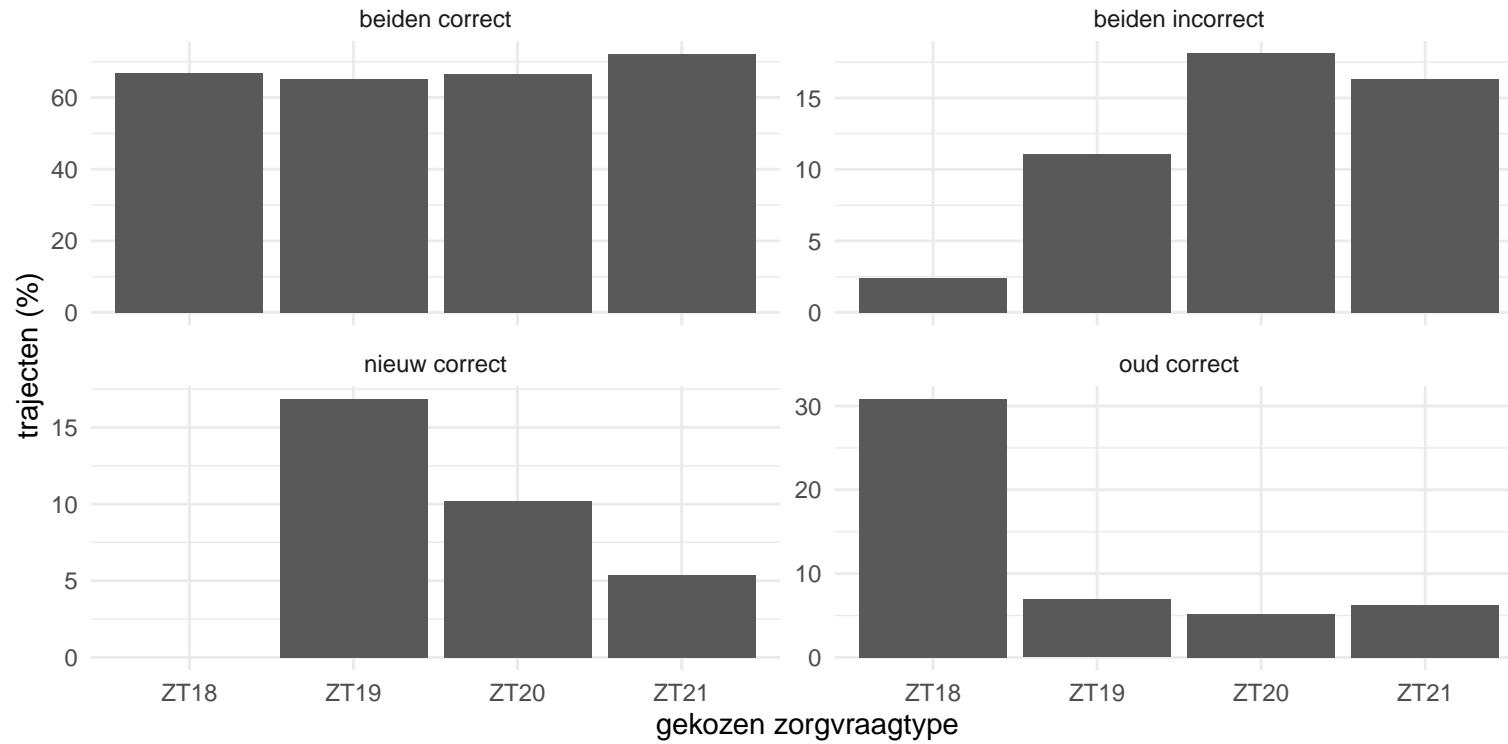
Vergelijking oud en nieuw algoritme  
hoofdgroep Z



Relatief ten opzichte van het totale aantal gekozen zorgvraagtyperingen deelt het oude algoritme ZT18 vaker en beter in, terwijl het nieuwe algoritme ZT19 en ZT20 beter indeelt.

Figuur 23

### Vergelijking oud en nieuw algoritme hoofdgroep Z



### 2.6.3 Rode regels

De rode regels exclusie regels, sommige combinaties van antwoorden op HoNOS+ vragen met een zorgvraagtype worden niet geaccepteerd. De rode regels zijn opgesteld door experts uit Engeland. Hieronder volgen visualisaties van deze regels voor de hoofdgroepen. Op de x-as staan de mogelijke scores op de items, op de y-as het zorgvraagtype. Wanneer een hokje rood is gekleurd wordt het zorgvraagtype uitgesloten (niet voorspeld) als de score is gegeven.

Een patroon in elke groep is dat ongeveer de helft van de zorgvraagtypes weinig variatie mogen hebben in de zwaarte van de symptomen (bijvoorbeeld ZT04 waarbij de hoogste score op items 7 of 8 de waarde 3 moet hebben). Bij de andere helft zijn er restricties uit andere vragen (ZT06, ZT08, ZT15) De grenzen tussen de andere zorgvraagtypen worden door het algoritme bepaald.

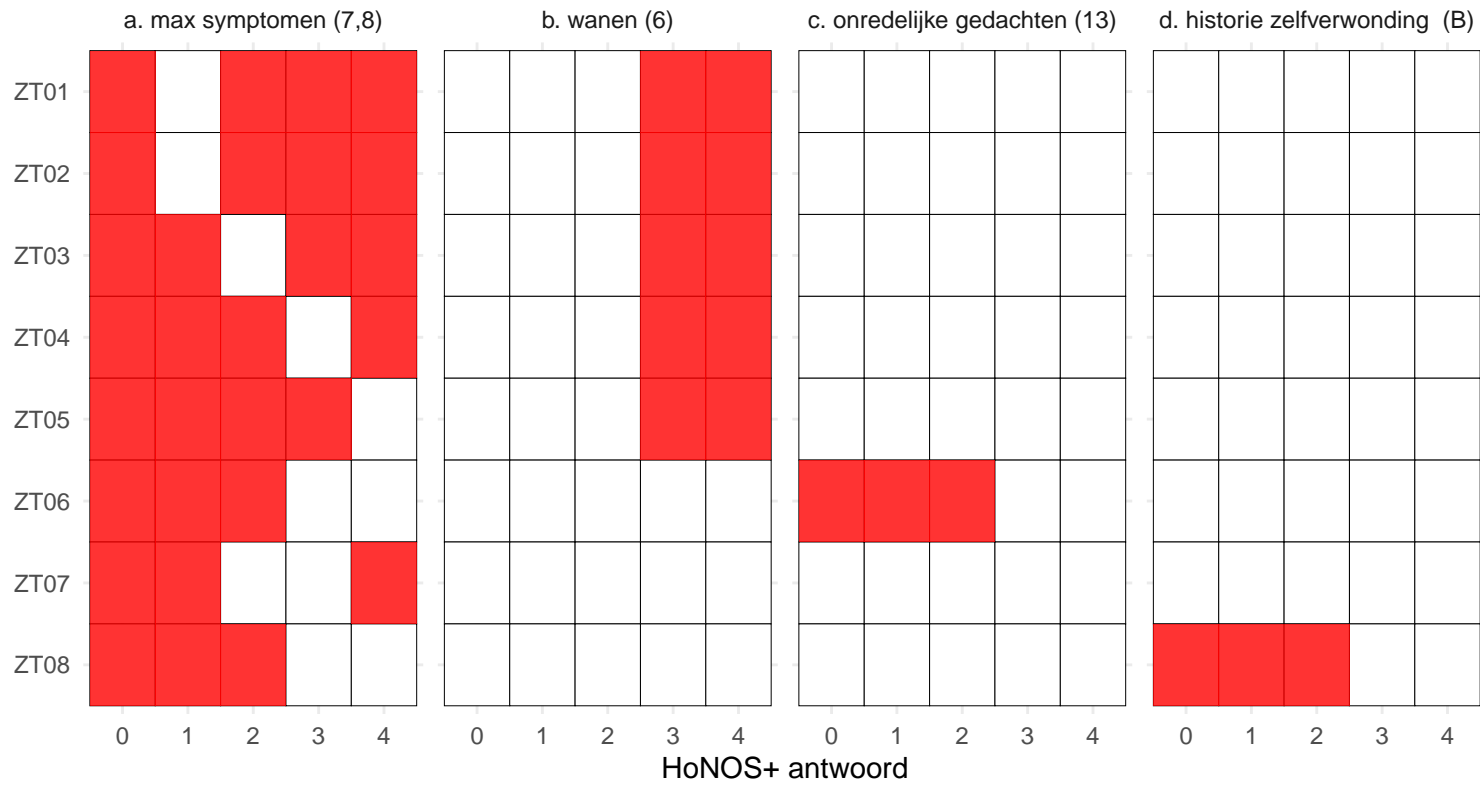
Bij X (psychische stoornissen) betreft het regels voor items 7 en 8 (respectievelijk depressieve symptomen of overige psychische symptomen), 6 (wanen en hallucinaties), 13 (sterk onredelijke gedachten) en B (historie van zelfverwonding.) Bij zorgvraagtype ZT01 tot en met ZT05 wordt de voorspelling van het zorgvraagtype uitgesloten op basis van de maximale score op items 7 of 8.

Bij Y (psychotische problemen) betreft het regels voor item 6 (wanen en hallucinaties), en vragen 7 (symptomen depressie). ZT11 wordt alleen voorspeld bij ontbrekende of lichte klachten rond wanen en hallucinaties, voor de andere zorgvraagtypes is lichte of zwaardere wanen juist een vereiste. ZT15 wordt niet voorspeld als er niet minimaal 3 is gescoord op item 7 (depressie).

Bij Z (cognitieve beperkingen) betreft het regels voor 4 (cognitieve beperkingen).

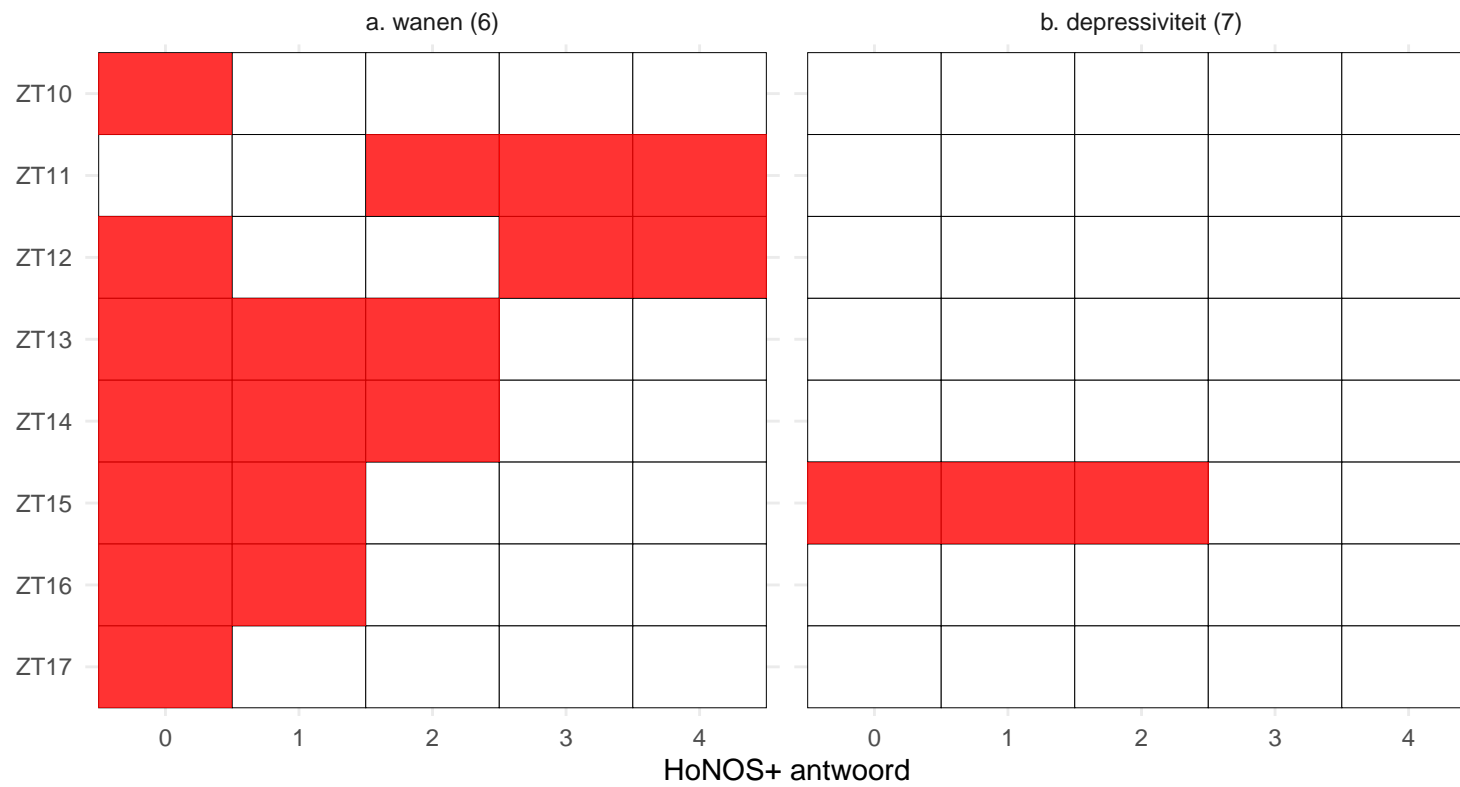
Figuur 24

Rode regels voor X – psychische stoornissen (rood is niet toegestaan)



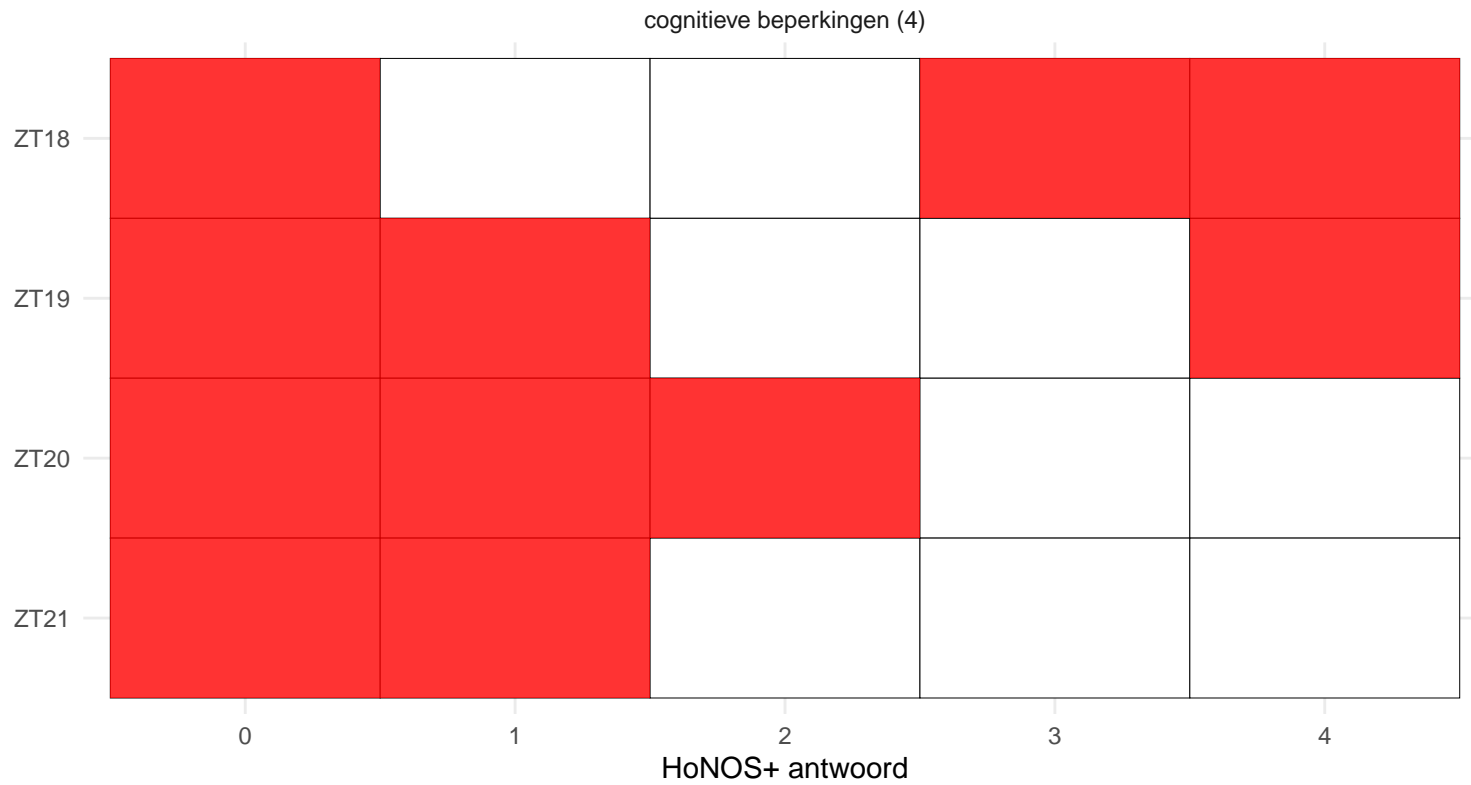
Figuur 25

Rode regels voor Y – Psychotische problemen (rood is niet toegestaan)



Figuur 26

Rode regels voor Z – cognitieve beperkingen (rood is niet toegestaan)



#### 2.6.4 Behandelaar volgt algoritme

Het oude algoritme kan de keuze van de behandelaar beïnvloeden; dat kan een nieuw algoritme niet in de huidige gegevens. Het is onbekend hoe sterk deze beïnvloeding is geweest en wat daarvan het meetbare resultaat is. Naar verwachting is er een mengeling van zorgvraagtyperingen die wel en niet zijn beïnvloed in de dataset. Om een gevoel te krijgen hoe de beïnvloeding van de behandelaar invloed heeft op de correcte voorspellingen van het nieuwe algoritme hebben we onderstaande simulatie uitgevoerd.

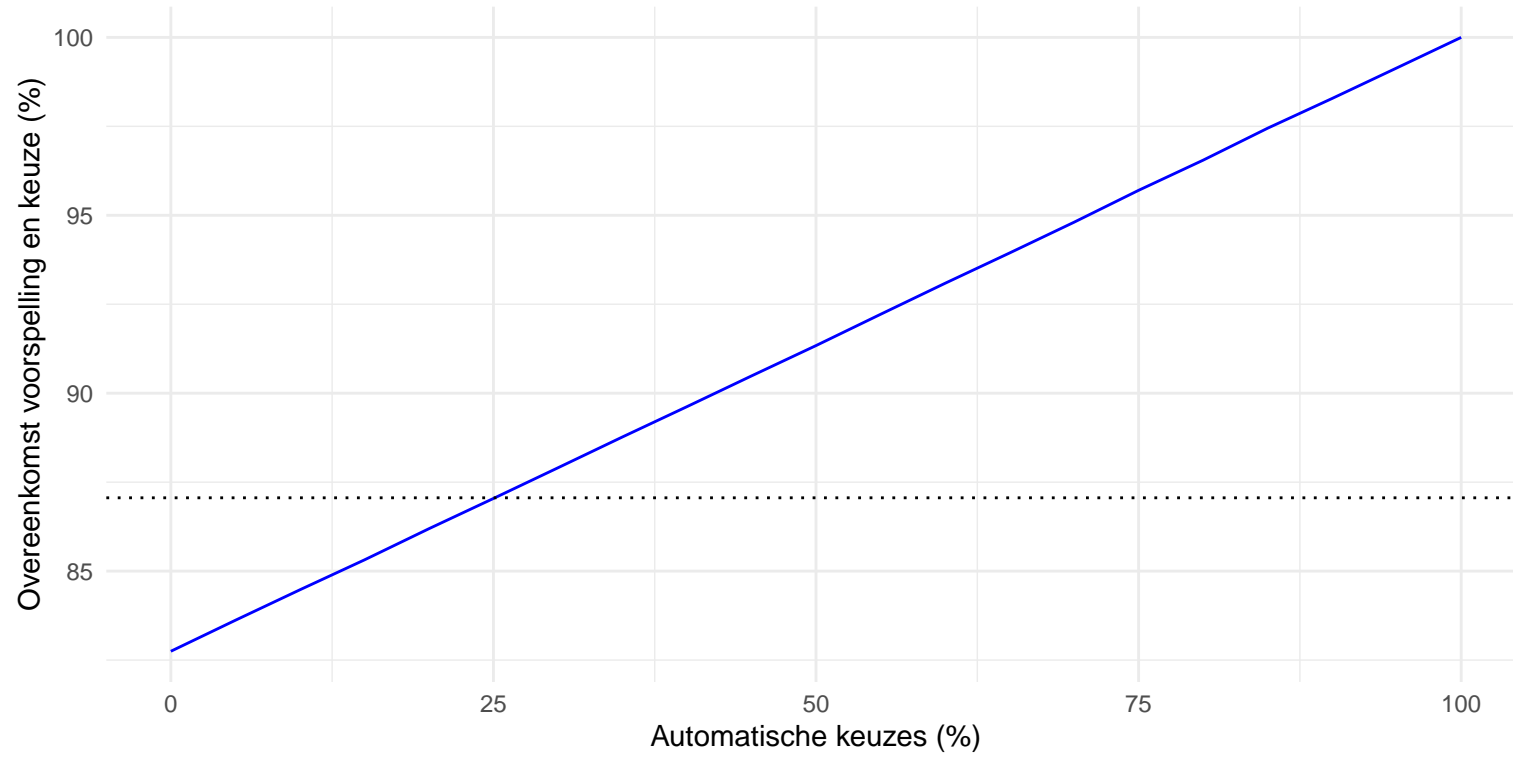
Op de y-as staat het percentage voorspellingen waarop de behandelaar het voorspelde zorgvraagtype heeft gekozen (87%). De stippelijijn geeft het percentage voorspellingen waarbij de behandelaar het voorspelde zorgvraagtype (oude algoritme) heeft gekozen, zoals gemeten in de gegevens. In de simulatie (blauwe lijn) laten we voor een bepaald percentage van zorgvraagtyperingen de 'behandelaar' (computer) automatisch kiezen voor de voorspelling. Het percentage automatische keuzes zien we op de x-as. Logischerwijs is bij 0% beïnvloeding het aandeel overeenkomsten tussen voorspelling en keuze even groot als hierboven gegeven (83%) en is bij 100% beïnvloeding de overeenkomst tussen voorspelling en keuze ook 100%. Uit deze simulatie blijkt dat wanneer in 25% van de twijfelgevallen het algoritme wordt gevolgd, dat het nieuwe algoritme dan een vergelijkbaar percentage overeenkomsten tussen voorspelling en keuze heeft als het oude algoritme.



Figuur 27

Effect van volgen van algoritme bij twijfel

stippelij: oude algoritme



### 2.6.5 Verfijnen trainen LDA

Er zijn enkele mogelijkheden om het trainen van de LDA te verfijnen. Hieronder laten we de uitkomsten zien van het werken met minder variabelen. Andere opties die we hebben geprobeerd zijn het balanceren van de verdeling over de zorgvraagtypes en het gebruiken van aanvullende informatie, zoals een variabele die expliciet maakt of een zorgvraag in een psychotische episode valt. Deze verfijningen laten we niet zien, omdat deze niet tot betere resultaten hebben geleid.

### 2.6.6 Minder variabelen

Mogelijk krijgen de LDA-algoritmes te veel informatie mee en leren ze daardoor van items die eigenlijk geen invloed horen te hebben op het zorgvraagtype van de patiënt. Het is mogelijk dat de algoritmes beter gaan voorspellen wanneer we minder variabelen (items) meegeven tijdens het 'trainen' van het algoritme. Dit onderzoeken we door verschillende selecties te maken van variabelen om in de training mee te geven aan het algoritme.

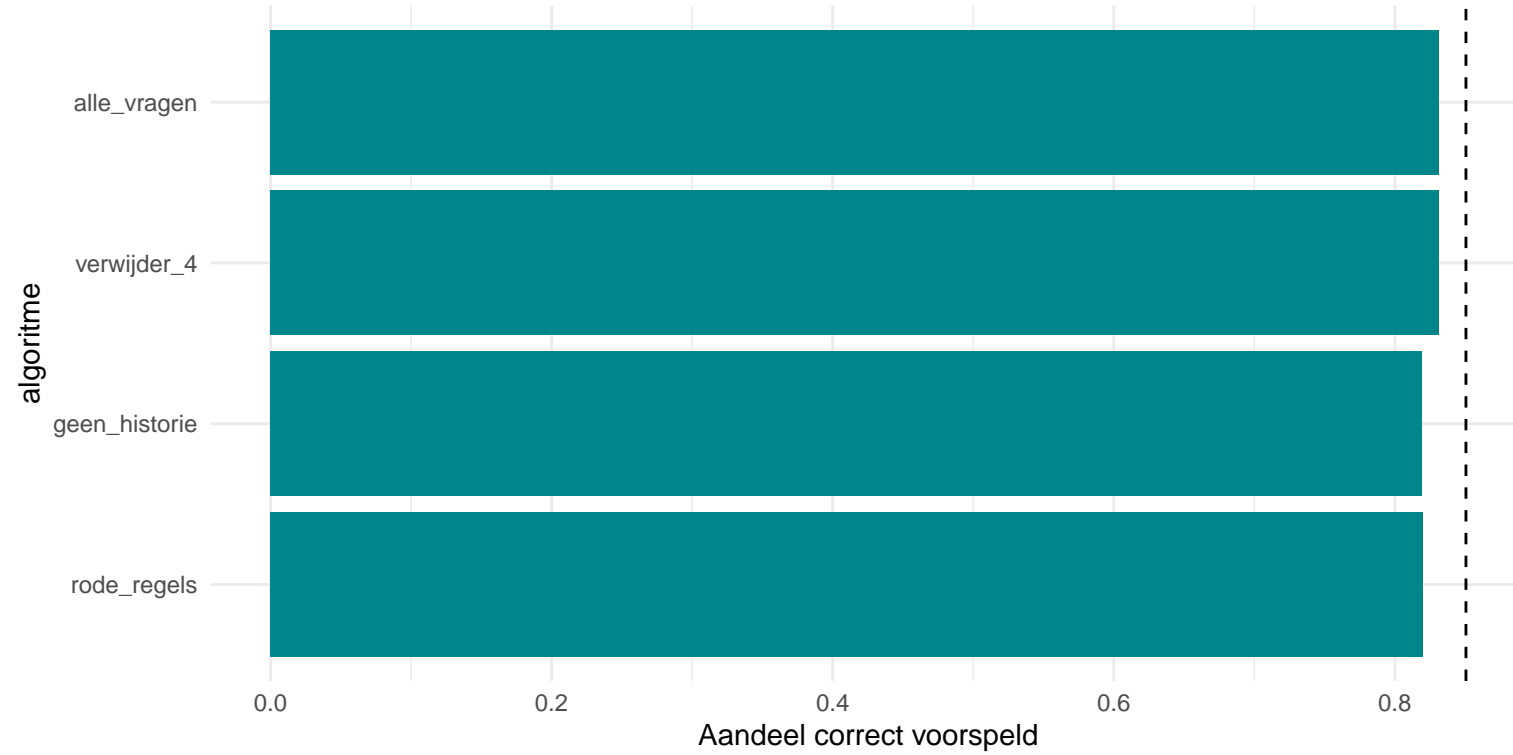
De kleinste selectie is met alleen de honosvragen waar rode regels voor bestaan. Deze zijn verschillend per hoofdgroep. Daarnaast zijn er algoritmes met alle honosvragen over gedrag, symptomen en sociale omgeving (vraag 1 t/m 4 en 6 t/m 13); met alle honosvragen min 4 vragen (1 t/m 4, 6 t/m 11, 13, A, B, D, E); en alle honosvragen. In onderstaande figuur geeft de stippellijn de correcte voorspellingen door het oude algoritme. Het is te zien dat het weglaten van items niet leidt tot betere prestaties. Het weglaten van andere items dan 5, 12, C en Q, zoals in het Engelse algoritme levert minimale verlaging van de prestaties. Het weglaten van meer items heeft iets meer effect, maar met alleen de rode regel-items kan het algoritme al een heel eind komen.

Hieronder splitsen we de resultaten ook uit per zorgvraagtype.

Figuur 28

Resultaat algoritmes met verschillend aantal items

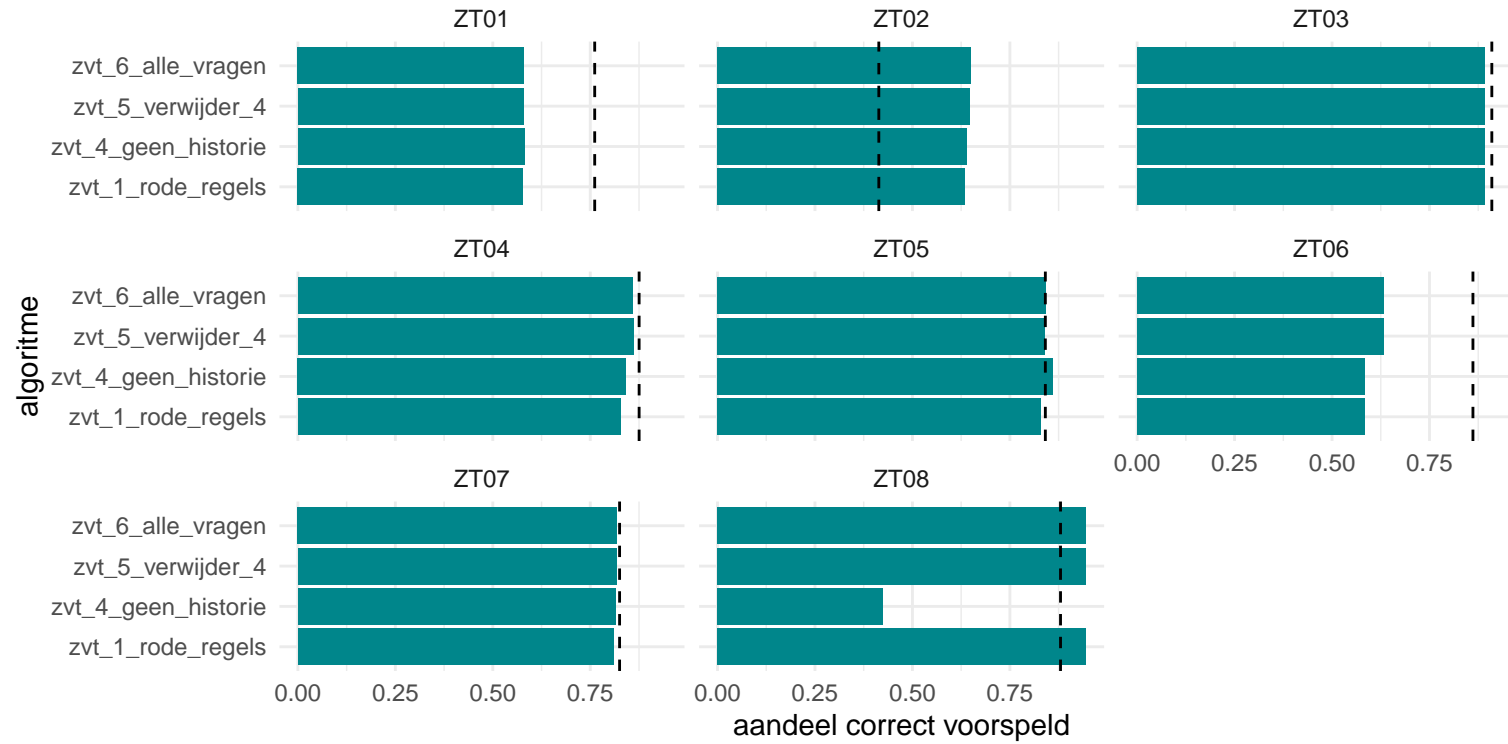
Stippelijijn: oud algoritme



Figuur 29

### Resultaat verschillende algoritmes per ZT

stippelijijn: oud algoritme

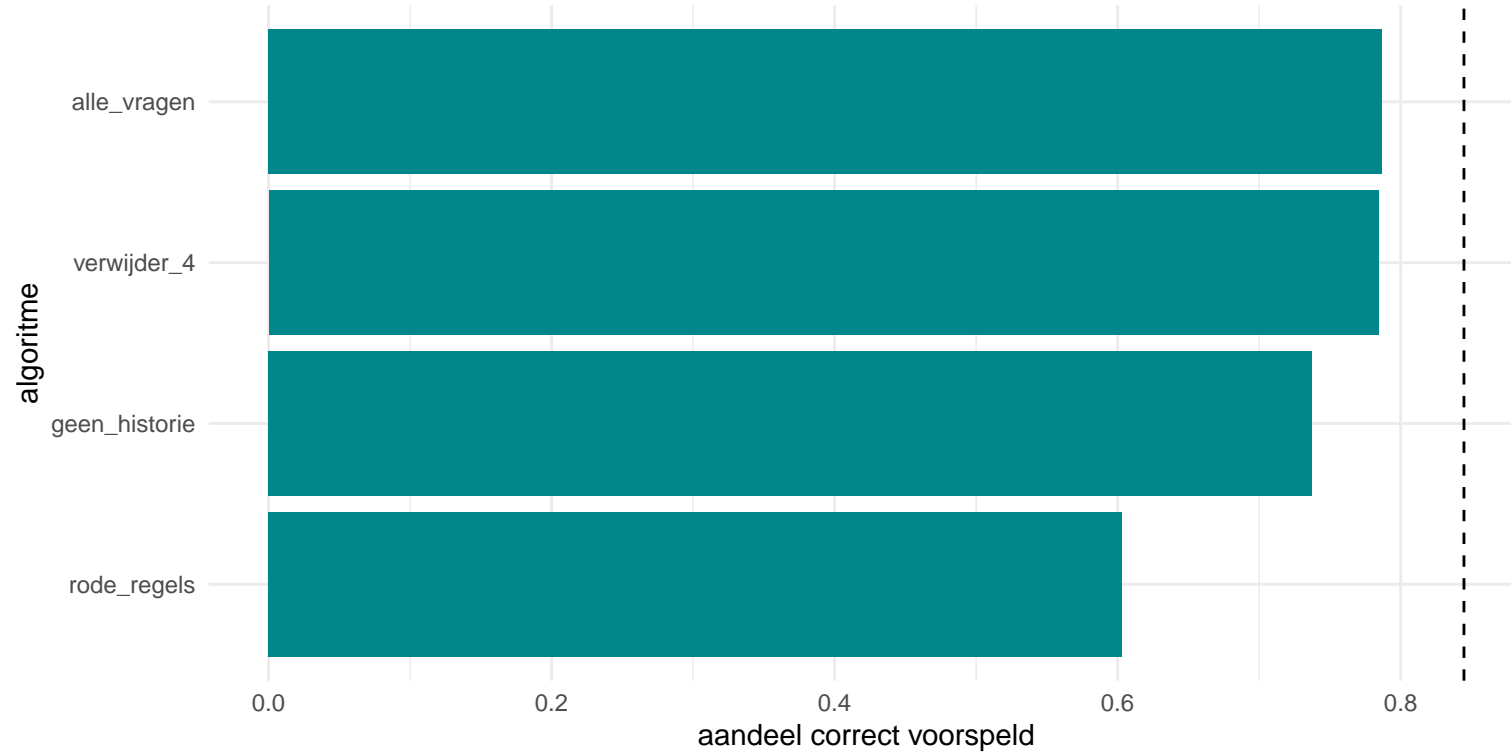


hoofdgroep X

voor hoofdgroep Y

Figuur 30

Resultaat algoritmes met verschillend aantal vragen  
stippellijn: oud algoritme

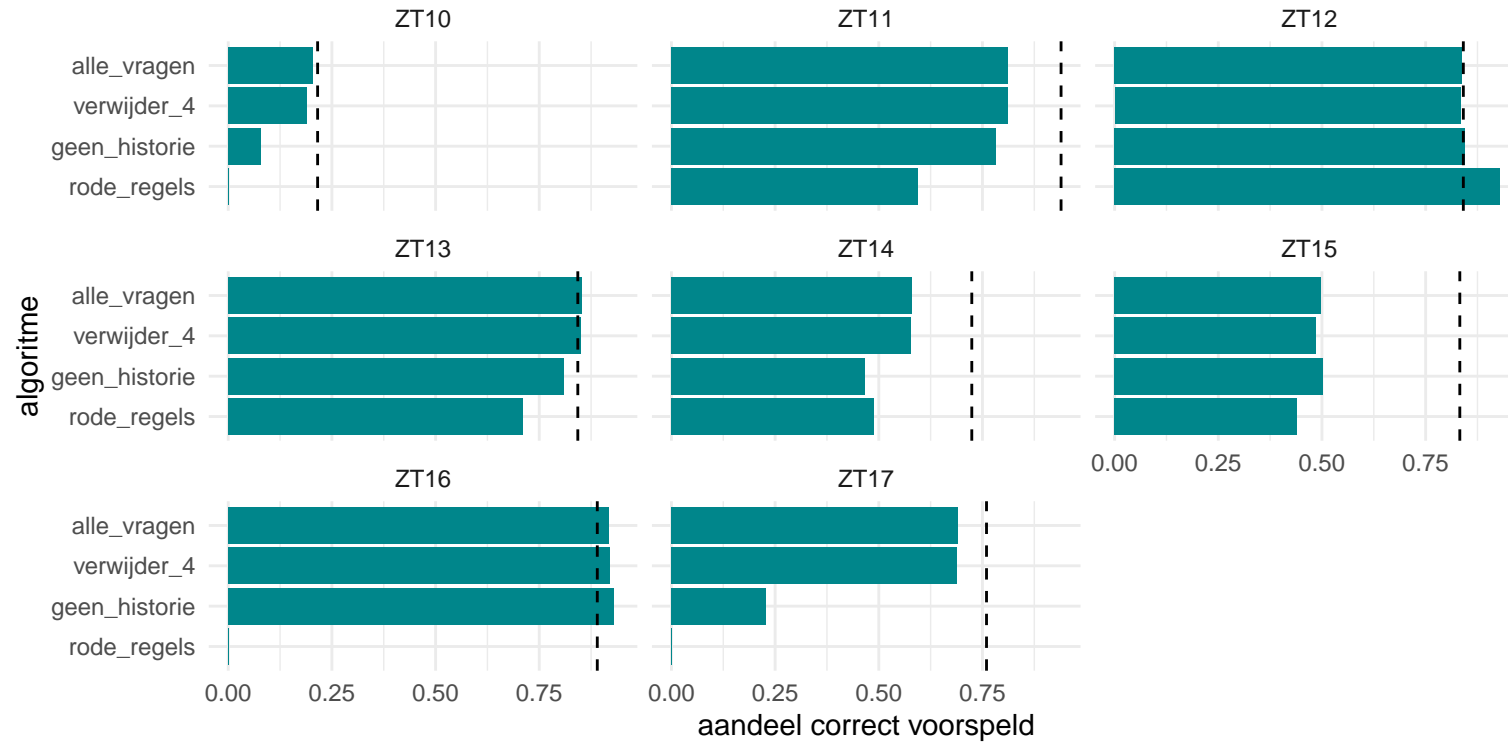


per gekozen zorgvraagtype

Figuur 31

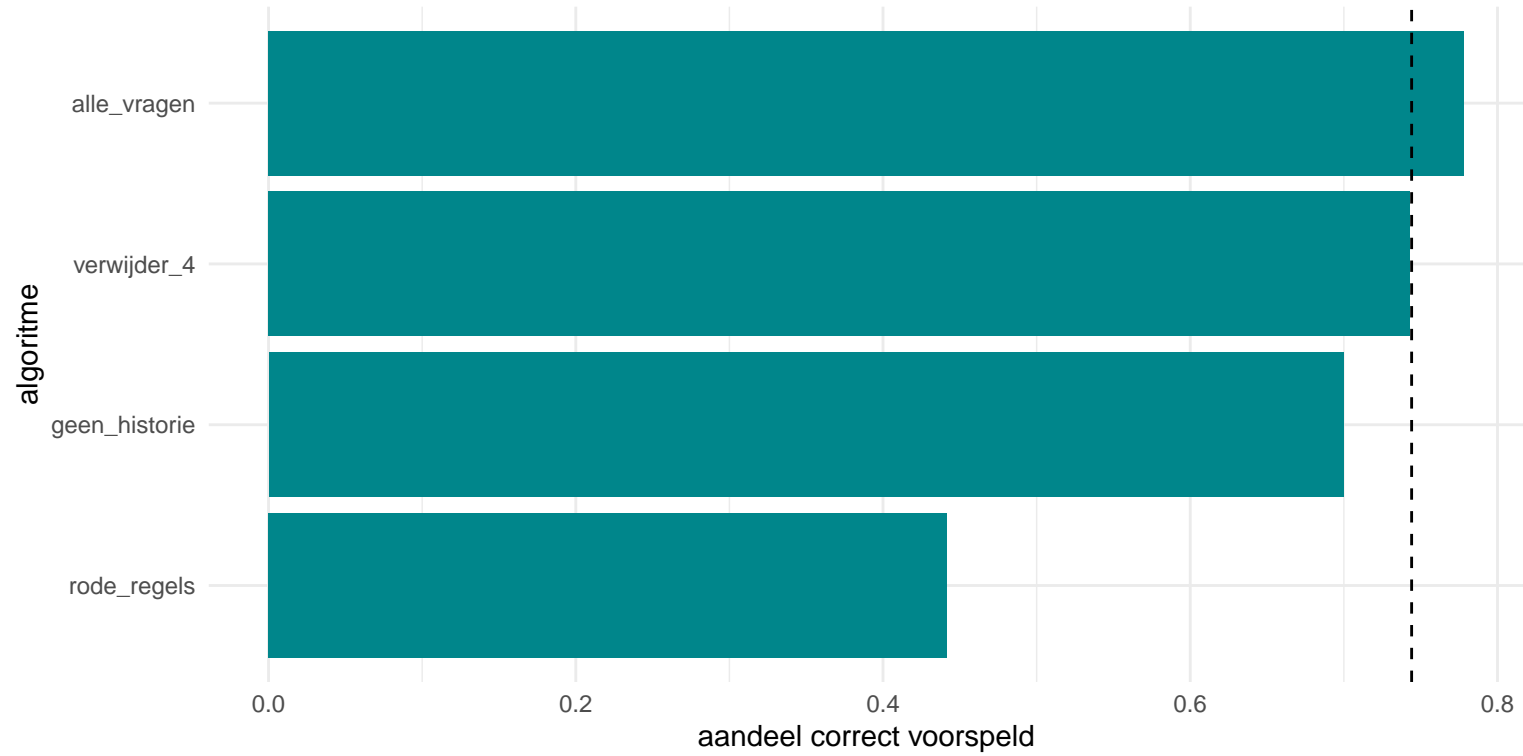
### Resultaat verschillende algoritmes per ZT

stippellijn: oud algoritme



Figuur 32

Resultaat algoritmes met verschillend aantal vragen  
stippelijijn: oud algoritme



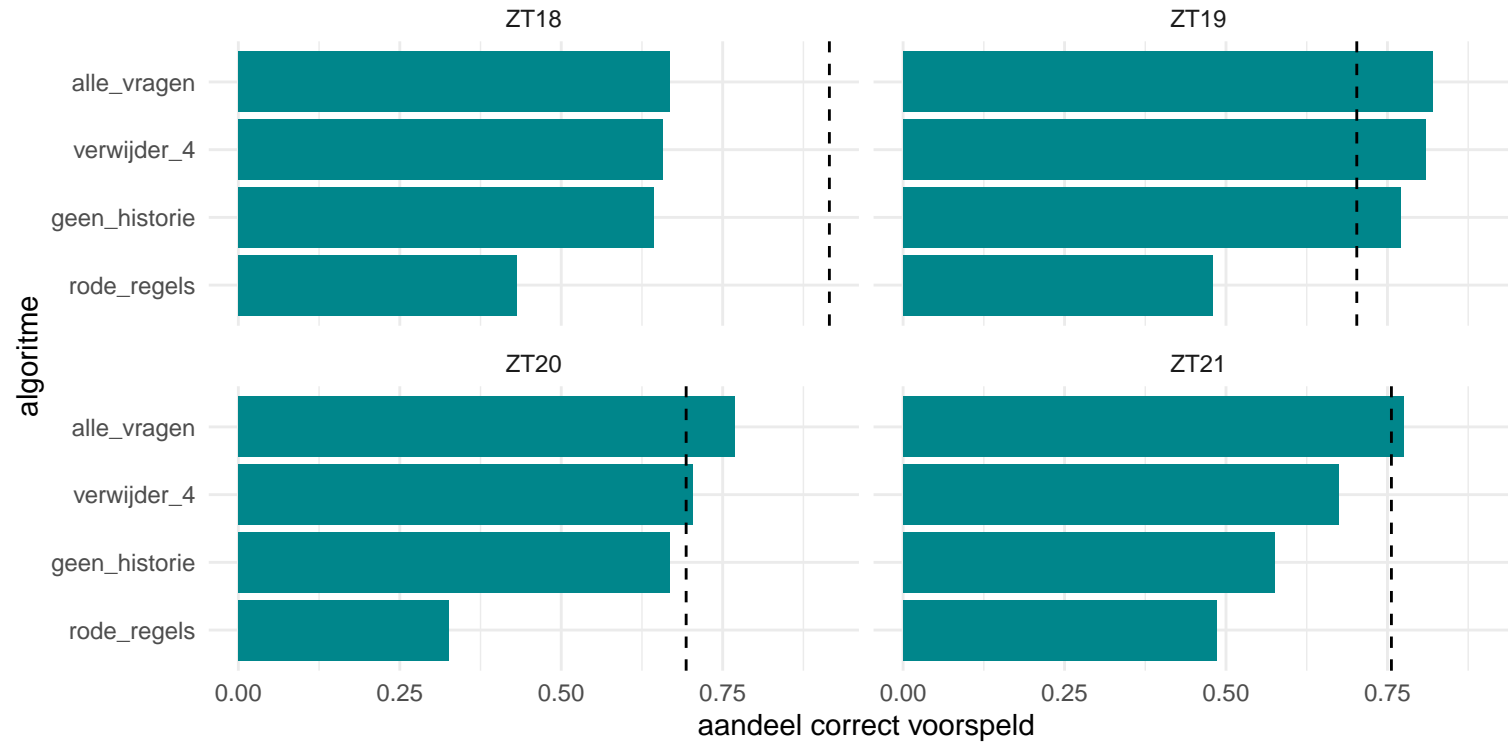
hoofdgroep Z



resultaat per gekozen ZVT

Figuur 33

Resultaat verschillende algoritmes per ZT  
stippellijn: oud algoritme

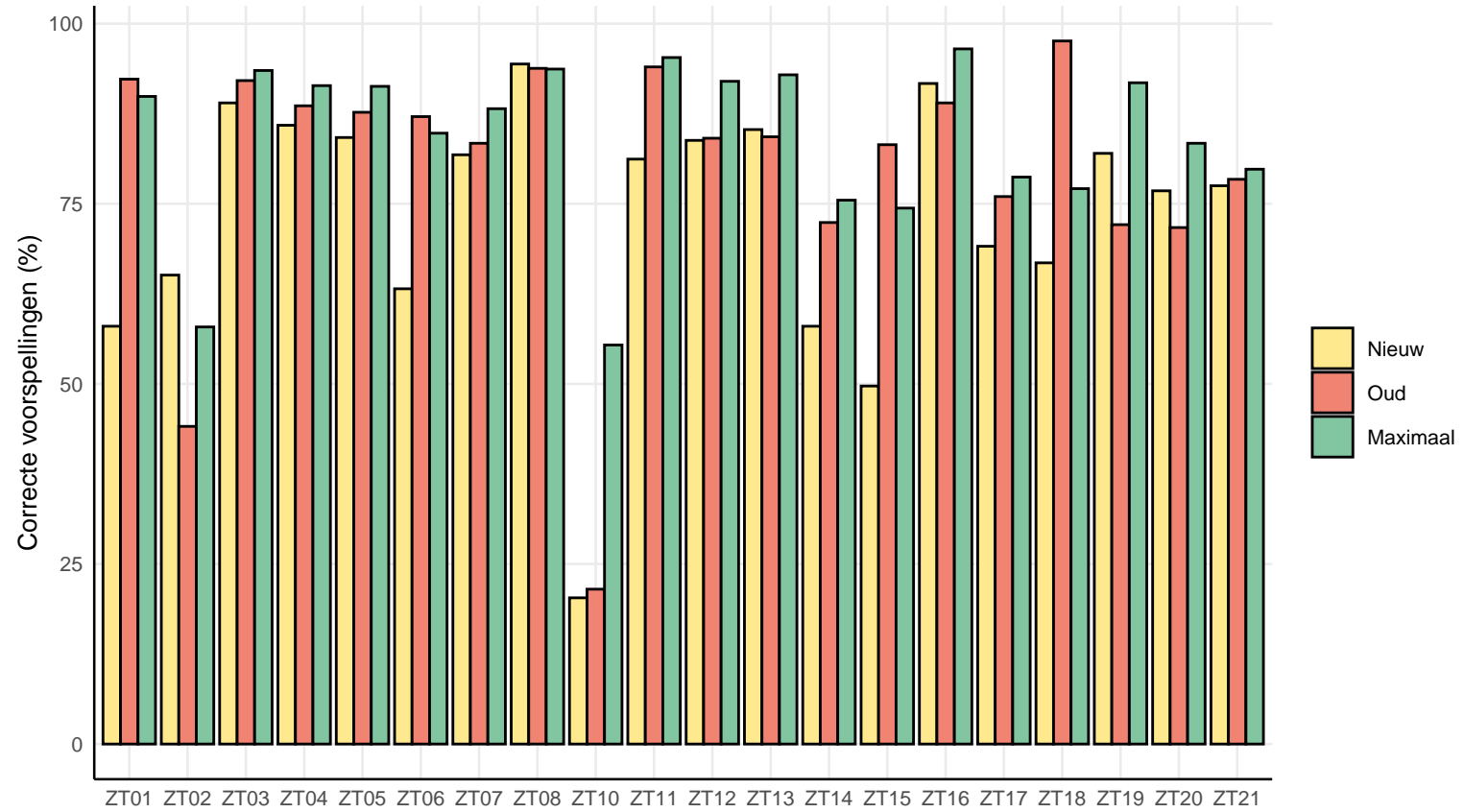


### 2.6.7 Hoeveel structuur in honosvragen?

In vrijwel elk proces dat gegevens produceert is ruis te vinden, die zorgt voor variatie in de gegevens. Met het proces van zorgvraagtypering is het net zo. Deze ruis is te verwachten en zorgt ervoor dat het niet mogelijk is perfecte voorspellingen te doen. Hoe hoog de lat wél kan liggen voor de voorspelling van het zorgvraagtype kunnen we onderzoeken door een andere methode voor het indelen van zorgvraagtypes te gebruiken.

De algoritmes zoeken naar structuur in de data en koppelen die aan de uitkomsten. Voor het algoritme voor de zorgvraagtypering is de structuur gezocht met een linear discriminant analysis (LDA). Er bestaan andere methoden, die andere aannames doen. Het random forest algoritme heeft minder aannames en is daarom flexibler in het vinden van een structuur. Het random forest algoritme zou daarom anders meer kunnen leren dan een LDA en betere prestaties leveren. In onderstaande figuur geven we de prestaties van de oude en de nieuwe LDA naast de prestaties van het random forest (groen, Maximaal) als benchmark. Het is te zien dat een random forest algoritme het in veel gevallen beter doet dan het nieuwe algoritme en veelal ook beter dan het oude algoritme, terwijl er geen beïnvloeding is geweest doordat de behandelaar de uitkomst van het algoritme heeft gezien. Vooral in hoofdgroep Y valt op dat een random forest algoritme beter kan voorspellen. Voor de adviescommissie is het goed om bij de keuze voor het al dan niet herschatten van de algoritmes rekening te houden met de mogelijkheid tot het veranderen van de technische werking van de algoritmes. Dit kan leiden tot betere voorspellingen, maar heeft ook implicaties voor de tijdlijnen, omdat het vraagt om (veel) ICT-implementatie.

Figuur 34



## 2.7 Wvggz/Wzd, verblijf of acute ggz in laatste jaar

Met de scorelijsten zijn drie aanvullende kenmerken aangeleverd over het voorkomen van acute ggz, verblijf of Wvggz/Wzd gedurende het jaar voorafgaand aan het uitvoeren van zorgvraagtypering. Deze kenmerken volgen uit een verkenning van een adviesbureau of zorgvraagtypes gesplitst kunnen worden om te leiden tot groepen met meer homogene kosten. Per kenmerk kan worden aangegeven of het kenmerk voor de patiënt geldt, of het niet geldt of dat het onbekend is.

Voor alle kenmerken interpreteren we scores die aangeven dat het kenmerk onbekend is, alsof het kenmerk niet aanwezig is. Voor Wvggz/Wzd is aangegeven of er sprake is van ofwel Wet verplichte ggz, ofwel Wet zorg en dwang ofwel beiden, alle drie deze scores interpreteren we dat er sprake is van het kenmerk. Het kan voorkomen dat meerdere van de kenmerken gelden voor een zorgvraagtypering.

NB De kwaliteit van de registraties van deze kenmerken is niet bekend. Verschillende EPD-leveranciers hebben de registratie van de kenmerken op verschillende manieren geïmplementeerd. Dit kan gevolgen hebben voor het aantal patiënten waarvoor de kenmerken als aanwezig is geregistreerd. Tegelijk weten we ook dat veel EPD-leveranciers een specifieke groep klanten hebben, bijvoorbeeld de grote instellingen (met gemiddeld hoge kosten) of vrijgevestigde aanbieders (met gemiddeld lagere kosten). Het is op basis van de huidige gegevens niet uit te sluiten dat verschillen tussen kosten en de kenmerken een gevolg is van de relatie tussen registratie-implementatie en EPD-submarkten. Dit betekent niet dat onderstaande analyses geen waarde hebben, maar het is belangrijk te realiseren dat de analyses mogelijk over- of onderschattingen van de daadwerkelijke situatie zijn en het is mogelijk dat ze een onvolledig beeld laten zien.

### 2.7.1 Voorkomen van de kenmerken

Hoe vaak zijn de verschillende kenmerken in de gegevens aanwezig?

Tabel 8: Aantal zorgvraagtyperingen met historisch kenmerk

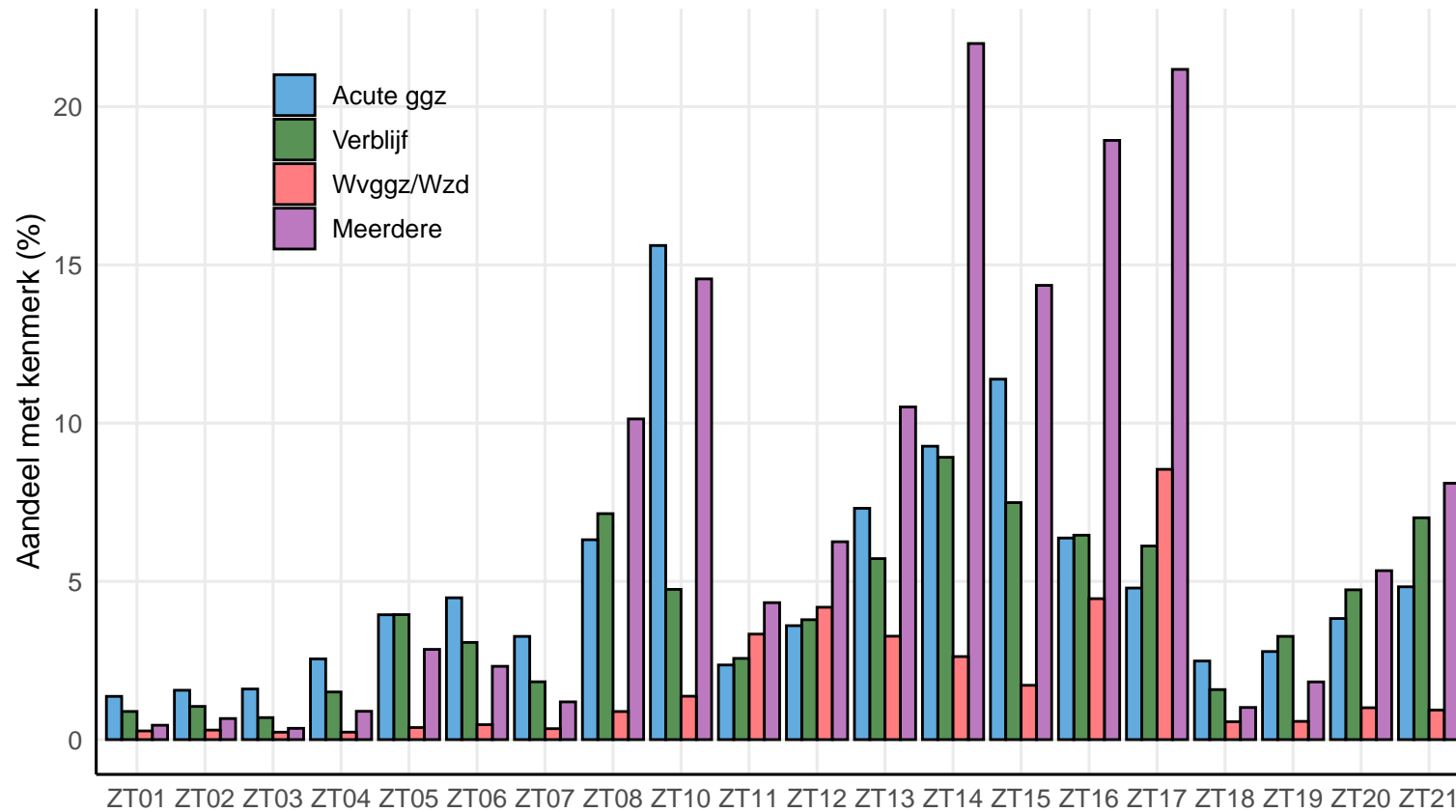
verzwaring	aantal
Geen	377242
Acute ggz	11024
Verblijf	7420
Wvggz/Wzd	2492
Meerdere	7418

### 2.7.2 Voorkomen kenmerken per zorgvraagtype

Bij welk aandeel van de patiënten in een zorgvraagtype komt het element voor? Zoals te verwachten komen de kenmerken niet vaker dan in 10% van de gevallen voor in hoofdgroep X, behalve bij ZT08. In hoofdgroep Y komen alle kenmerken relatief vaker voor, en vooral bij de complexe zorgvraagtypes en psychotische episodes komen de elementen relatief vaak voor.

NB een zorgvraagtypering valt in één van de staafjes in het staafdiagram. Een patiënt met zowel verblijf als acute ggz komt dus alleen voor in het paarse staafje (meerdere elementen) en niet in de afzonderlijke staafjes acuut of verblijf.

Figuur 35: Aandeel van de zorgvraagtyperingen met één of meerdere van de kenmerken



**Voorkomen kenmerken - aantallen** Om volgende analyses te interpreteren is het van belang te weten voor welk aantal zorgvraagtyperingen het kenmerk voorkomt. Hieronder een tabel met de aantallen. Het kan zo zijn dat meerdere kenmerken gelijktijdig gelden voor een zorgvraagtypering, dit is te zien in kolom 'Meerdere'.

Tabel 9: Aantal zorgvraagtyperingen met historisch kenmerk per zorgvraagtype

gekozen zorgvraagtype	Acute ggz	Verblijf	Wvggz/Wzd	Meerdere
ZT01	304	198	61	101
ZT02	171	115	33	73
ZT03	2036	885	299	454
ZT04	1915	1130	177	672
ZT05	941	942	91	680
ZT06	683	468	72	353
ZT07	2774	1554	294	1014
ZT08	306	346	43	491
ZT10	148	45	13	138
ZT11	261	284	369	479
ZT12	421	443	489	731
ZT13	331	259	148	476
ZT14	212	204	60	503
ZT15	73	48	11	92
ZT16	216	219	151	642
ZT17	83	106	148	367
ZT18	22	14	5	9
ZT19	58	68	12	38
ZT20	38	47	10	53
ZT21	31	45	6	52

### 2.7.3 Verschil kosten bij kenmerk

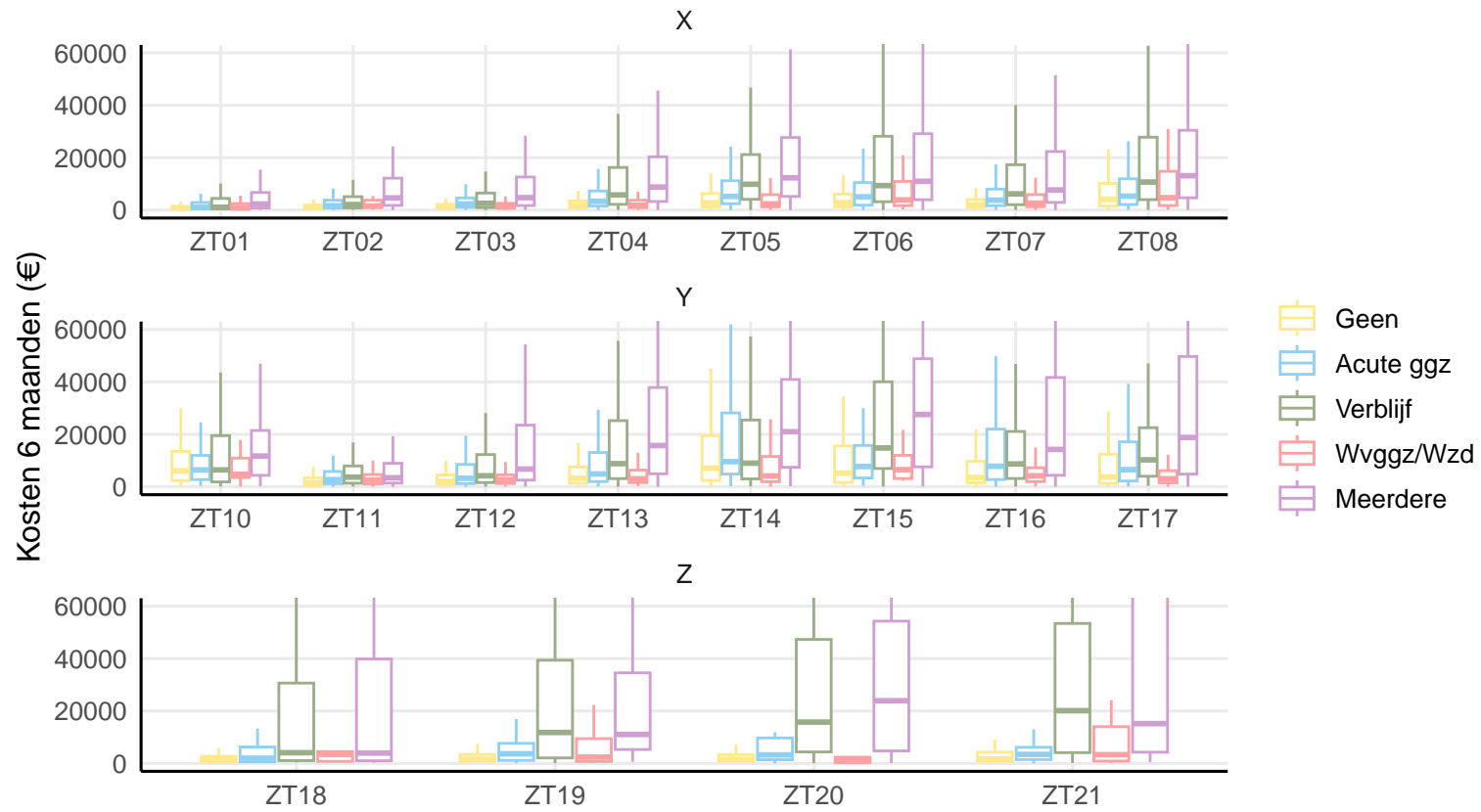
Uit het rapport dat SiRM in opdracht van enkele partijen heeft opgeleverd blijkt dat de kosten voor patiënten met één of meerdere van de kenmerken hoger kunnen liggen dan patiënten zonder de kenmerken. Dit onderzoeken wij ook door de verdeling van kosten binnen een zorgvraagtype te laten zien voor de groepen met elk van de kenmerken, de afwezigheid van de kenmerken of met meerdere kenmerken.

Er is grote verscheidenheid tussen de verschillen van kosten met en zonder kenmerken binnen de zorgvraagtypes. Het valt op dat vooral de aanwezigheid van verblijf in het afgelopen jaar of meerdere kenmerken lijken te leiden tot hogere kosten in vergelijking met de groep zorgvraagtyperingen zonder deze kenmerken.

Het is een punt om te bespreken hoe deze bevindingen in de zorgvraagtypering gevat kunnen worden, waarbij het belangrijk is ook het afwegingskader subclusters mee te nemen dat eerder is besproken in het programma Zorgprestatie-model (8 januari 2021).

**Kosten bij kenmerk - boxplot** De zorgkosten zijn doorgaans niet normaal verdeeld. Om de verdeling goed in te kunnen schatten is een boxplot geschikter. Daarom visualiseren we hier de kosten in een boxplot.

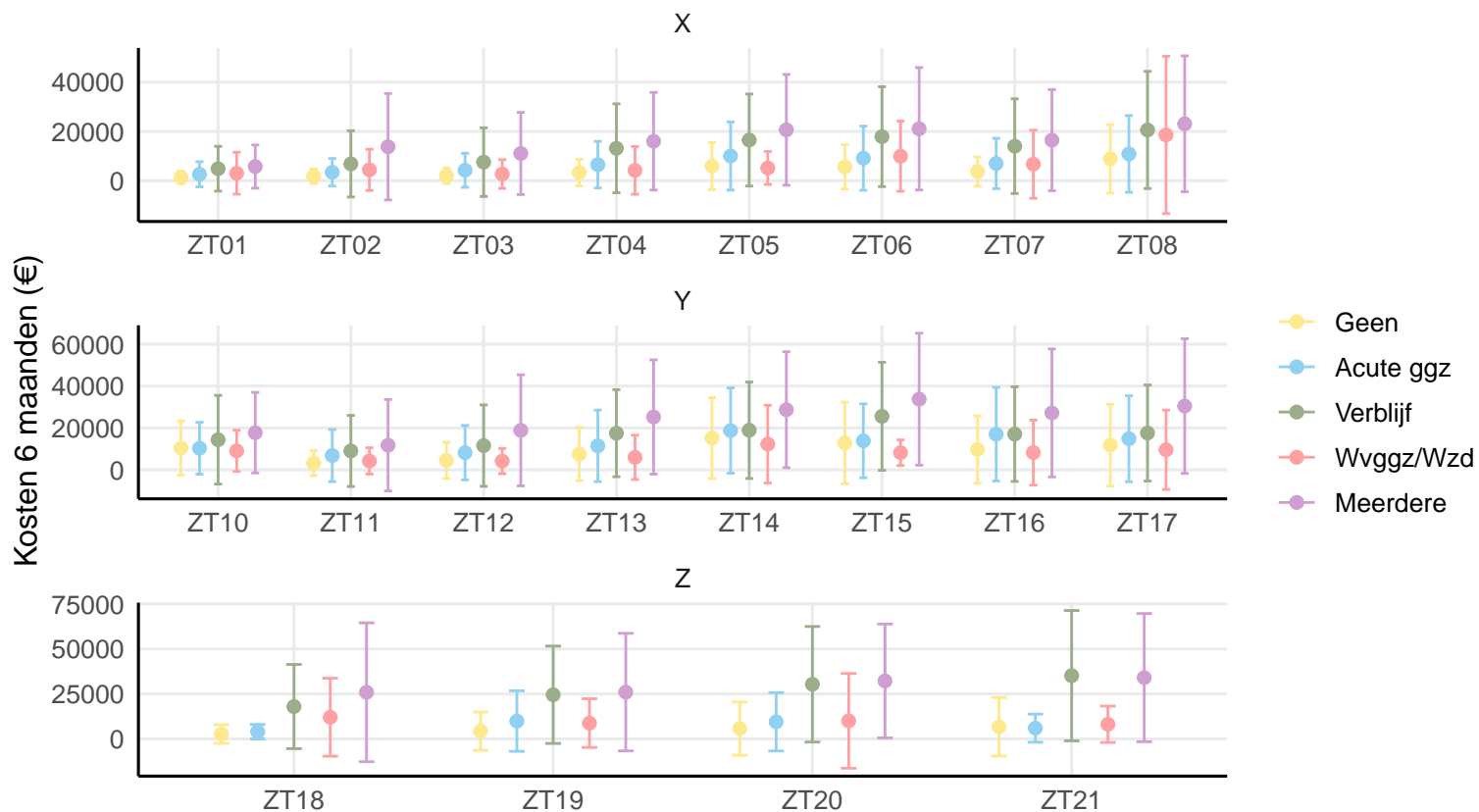
Figuur 36





**Kosten bij kenmerk - gemiddelde en afwijking** Om in te schatten of de gemiddelde kosten tussen de verschillende groepen ook daadwerkelijk verschillen is een gemiddelde met standaardafwijking makkelijker te interpreteren. Daarom visualiseren we deze ook.

Figuur 37: Gemiddelde kosten per zorgvraagtypering met of zonder kenmerk



**Kosten bij kenmerk - relatieve kosten** We bekijken de kenmerken bij de zorgvraagtypering vanuit het idee dat sommige zorgvraagtyperingen mogelijk te splitsen zijn op basis van één of enkele van de kenmerken. De verwachting is dat de groep met het kenmerk gemiddeld hogere kosten met zich mee brengt dan de groep zonder het kenmerk. Hoe groot dit verschil is kunnen we bepalen door de ratio te nemen tussen de gemiddelde kosten van de groep met het kenmerk (of kenmerken) en de gemiddelde kosten van de groep die dat niet heeft. Dit visualiseren we hieronder als de relatieve kosten. Als een bolletje op 1 ligt zijn de kosten voor beide groepen even hoog, als het bolletje op 2 ligt zijn de kosten twee keer zo hoog in de groep met het kenmerk enzovoorts. Er bestaan dus groepen, zoals de groep ZT02 met meerdere aanvullende kenmerken, waarvoor de kosten ruim zes keer zo hoog zijn in vergelijking met de groep met geen of één van de kenmerken. Het is goed hierbij te realiseren dat dit vaak om hele kleine groepen gaat van enkele of enkele tientallen zorgvraagtyperingen met het kenmerk. In deze gevallen lijkt het niet logisch een zorgvraagtype op basis van het kenmerk te splitsen, ookal zijn de relatieve kosten heel hoog. Bij figuur 43 en verder kijken we meer naar de relatieve kosten in relatie tot het voorkomen van de kenmerken in een zorgvraagtype.

Figuur 38: Kosten van zorgvraagtype met kenmerk ten opzichte van de groep zonder

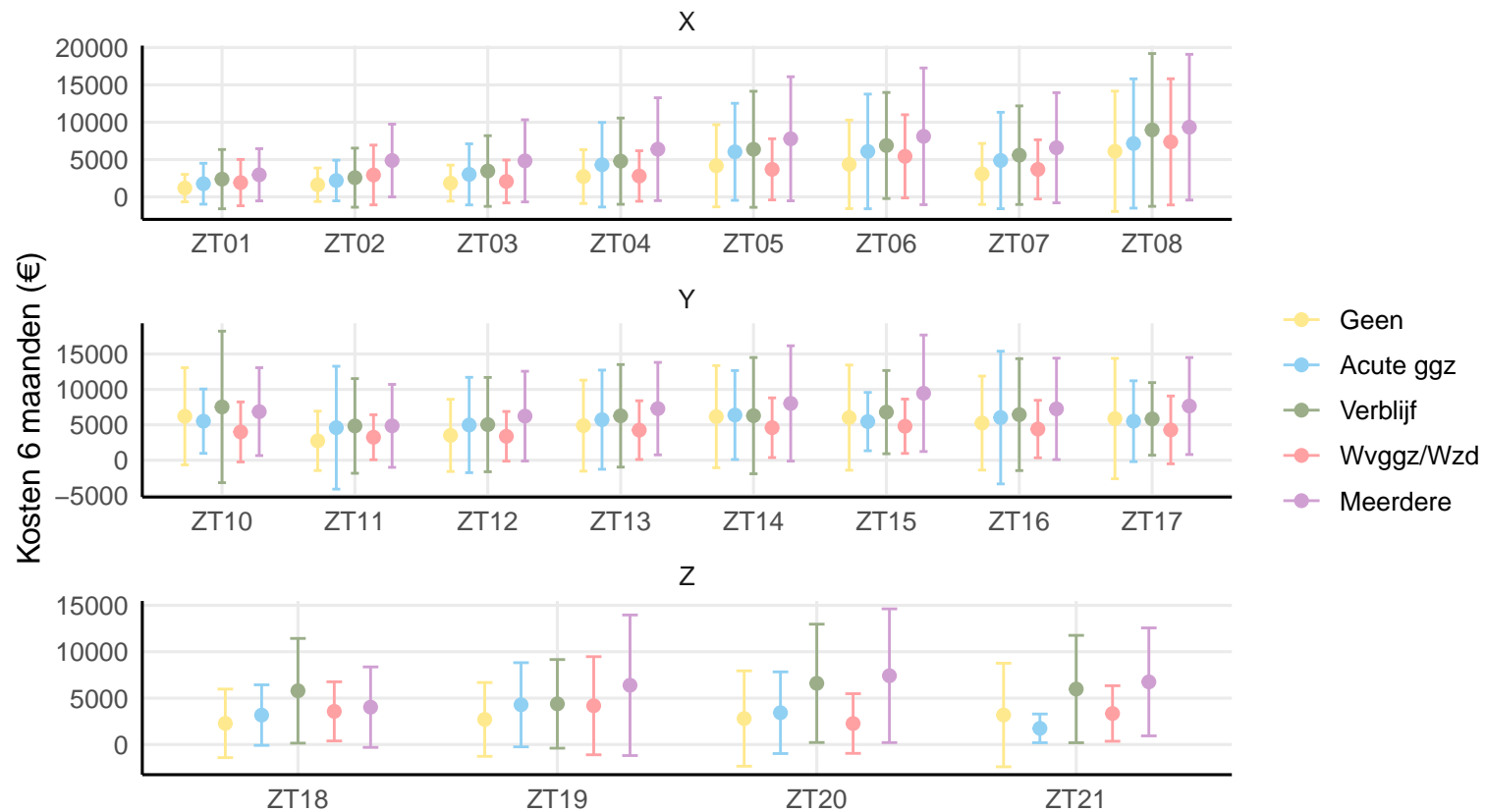


### 2.7.4 Kosten kenmerken per prestatietype

Het kan zijn dat specifieke verhoging van kosten toe te schrijven is aan een bepaald type prestaties. Daarom verkennen we deze één voor één.

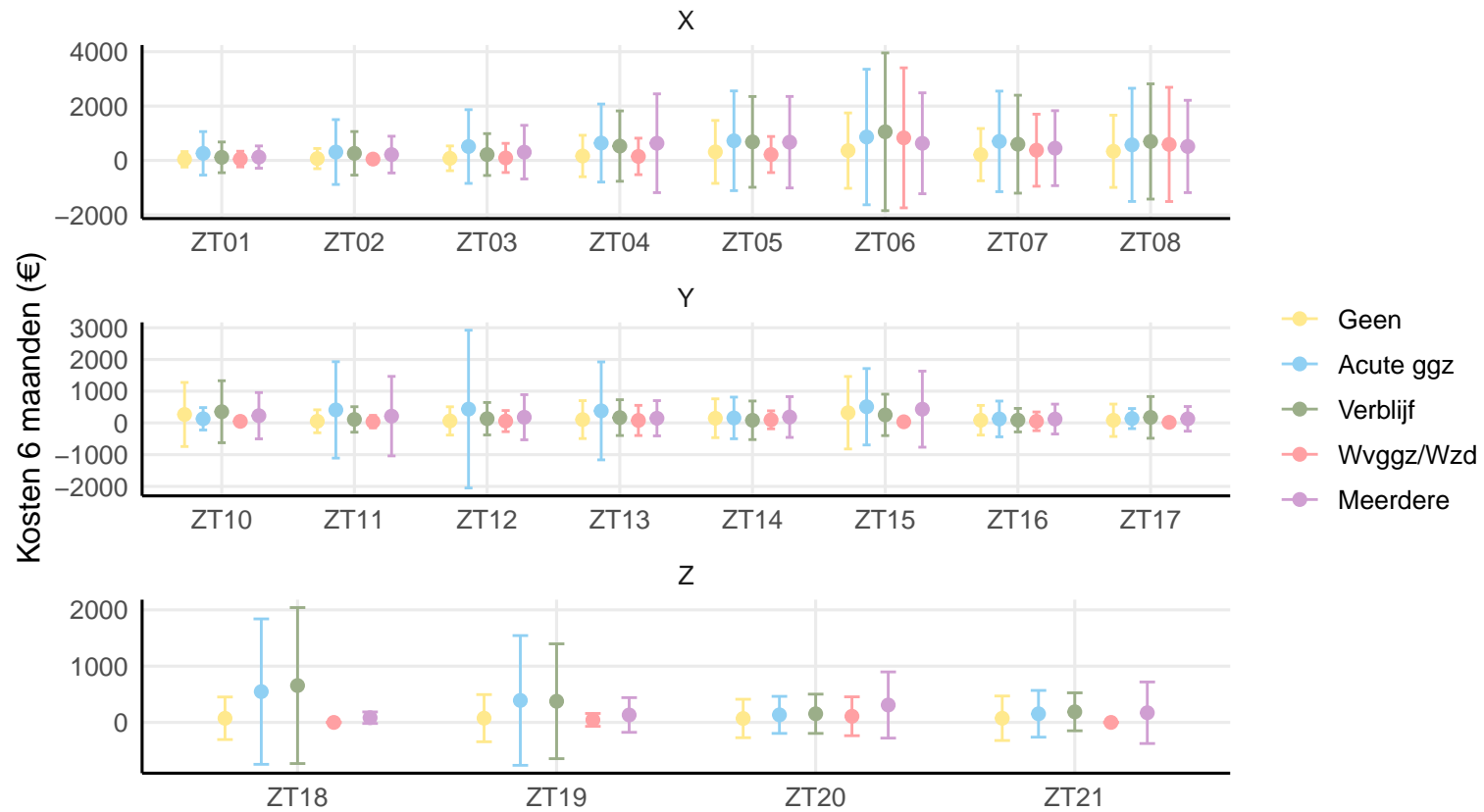
**Kosten verzwaren - consultants** Er is grote spreiding van de kosten. Dit maakt het moeilijk duidelijk aan te wijzen of de (gemiddelde) kosten van specifieke groepen van elkaar verschillen.

Figuur 39



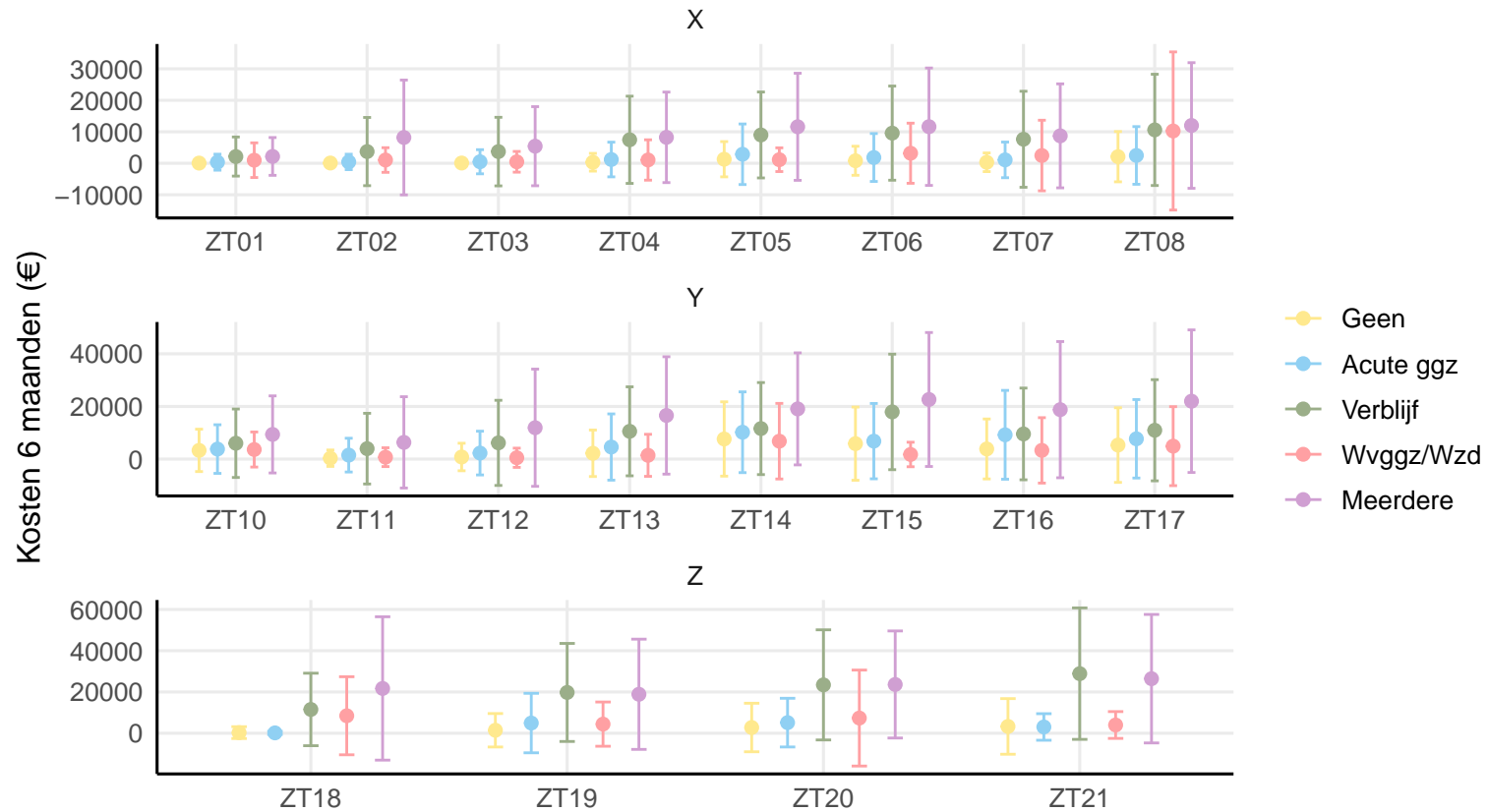
**Kosten verzvaring - Groepsconsulten** Voor de groepsconsulten valt vooral op dat de standaardafwijking (spreiding) vele malen groter is dan het gemiddelde. Daarom is het niet goed te bepalen of de groepen met kenmerken hogere groepskosten hebben dan de groep zonder.

Figuur 40



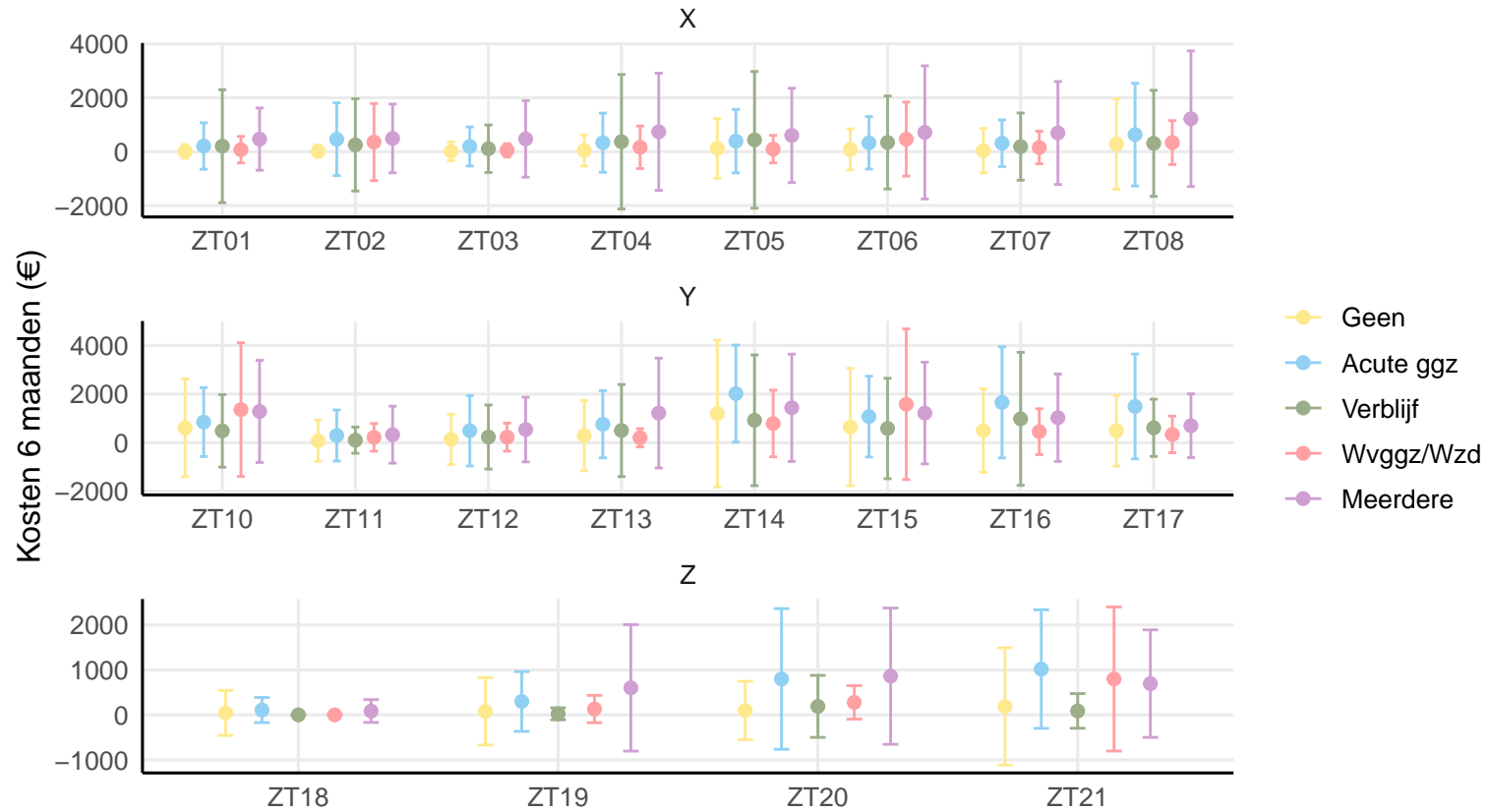
**Kosten verzwaren - Verblifsdagen** Bij de verblifsdagen valt logischerwijs op dat verblijf gedurende het afgelopen jaar bij veel zorgvraagtyperingen lijkt te leiden tot hogere gemiddelde verblifskosten. Voor de groep met meerdere kenmerken lijken de verblifskosten in veel gevallen ook hoger te liggen dan de verblifskosten van de groep zonder kenmerken, dit is vermoedelijk het geval wanneer er zowel sprake is van verblijf als acute ggz en/of Wvggz/Wzd.

Figuur 41



**Kosten verzwaren - Overig** Bij de overige kosten valt op dat dit niet sterk beïnvloed lijkt in hoofdgroep X, terwijl de kosten voor sommige zorgvraagtypes in hoofdgroep Y en Z hogere kosten laten zien wanneer er sprake is van acute zorg in het afgelopen jaar, of wanneer er sprake is geweest van Wvggz/Wzd of meerdere kenmerken. De relatie met de acute zorg is mogelijk te verklaren doordat dit crisisgevoelige patiënten betreft. Als er eerder acute ggz nodig is geweest kan dit weer gebeuren, deze zorg wordt als overige prestaties gedeclareerd en zit daarom in deze categorie.

Figuur 42



### 2.7.5 Voorkomen kenmerk en relatieve kosten

Als we overwegen zorgvraagtypes te splitsen zijn de kosten van de twee groepen niet de enige maatstaf om rekening te houden. Ook de relatieve omvang van elk van de splitsingen is van belang. In figuur 35 zien we in hoe groot aandeel van de zorgvraagtyperingen er sprake is van één of meerdere van de kenmerken en uit 37 en 38 is op te maken hoe groot de kosten van een zorgvraagtype met een kenmerk zijn ten opzichte van de groep zonder. We berekenen voor elke combinatie van zorgvraagtype en kenmerk wat de relatieve kosten zijn (kosten van groep mét gedeeld door kosten van groep zónder kenmerk(en)) en zetten dit af tegen het aandeel van de zorgvraagtyperingen waarvan het kenmerk voorkomt.

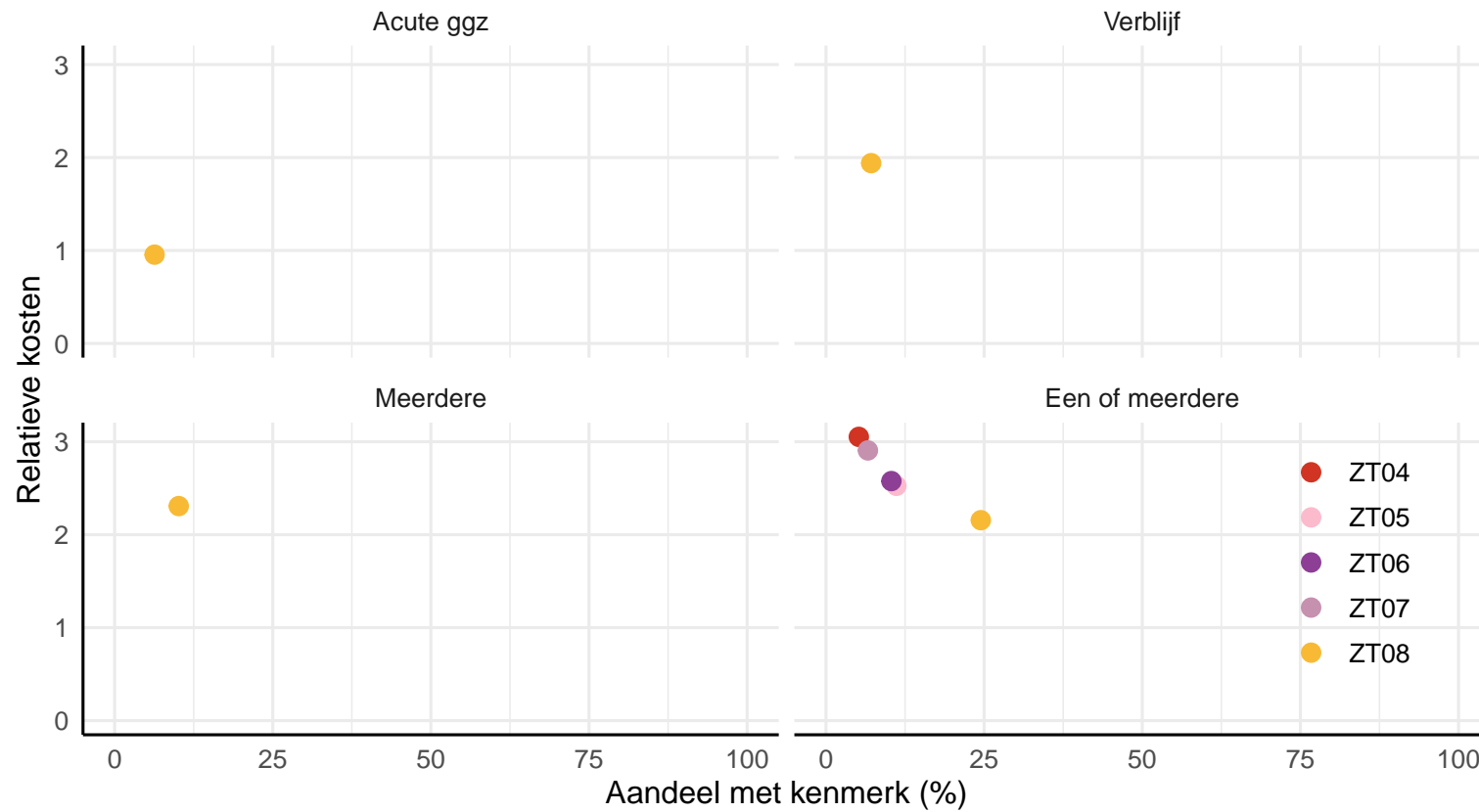
Voor de zichtbaarheid van de figuren en bruikbaarheid van een eventuele splitsing visualiseren we alleen de gevallen waarbij er minimaal 5% en minimaal 250 patiënten in een zorgvraagtype het kenmerk heeft.



**Relatieve kosten en aandeel - X** De kosten voor de gevisualiseerde zorgvraagtypes met kenmerk zijn vaak 2-3 keer zo hoog als de vergelijkbare groep zonder het kenmerk. Het valt op dat voor de meeste kenmerken en zorgvraagtypes het aandeel van het zorgvraagtype met het kenmerk rond of onder de 10% van de zorgvraagtyperingen ligt, behalve voor ZT08.

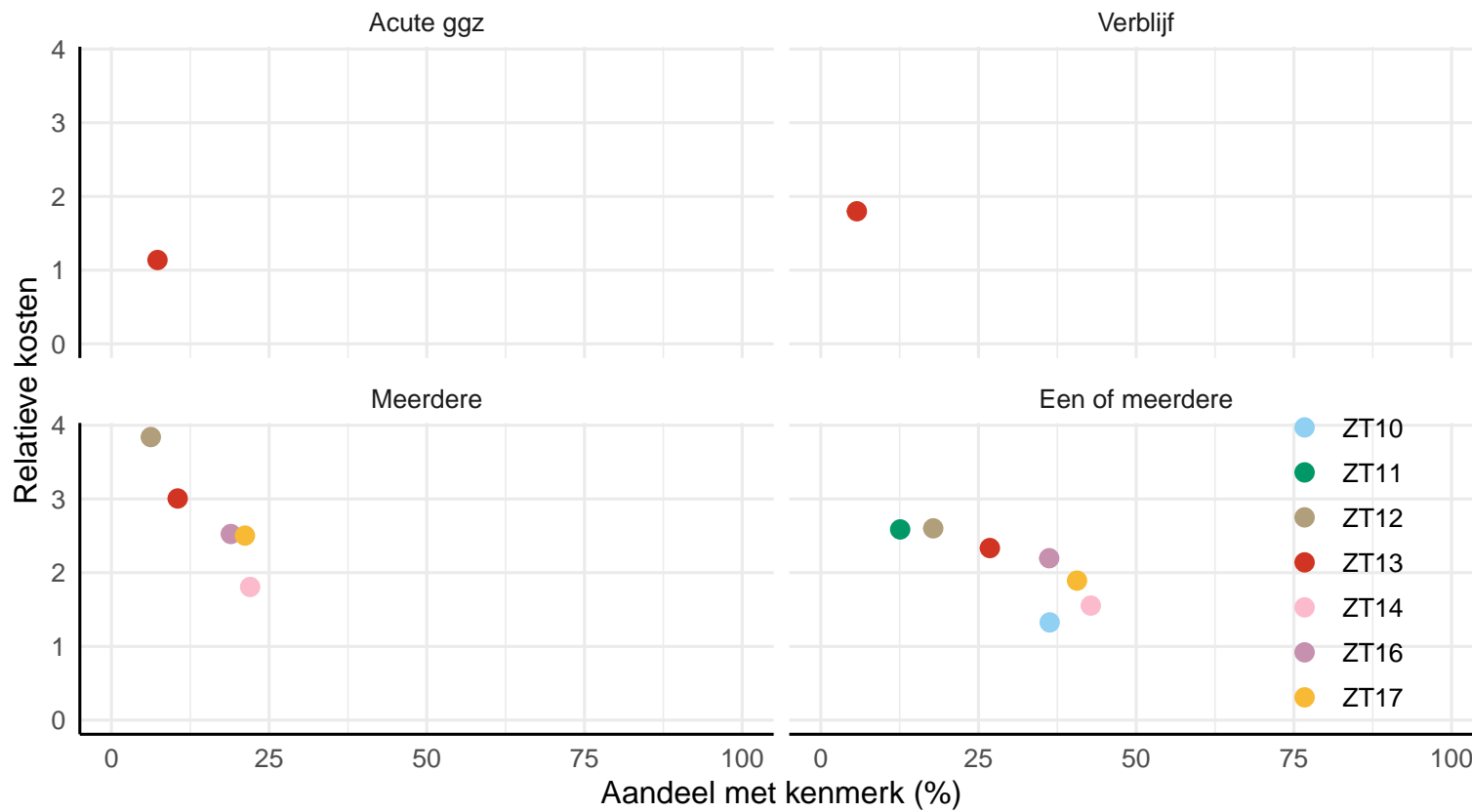
Hoofdgroep X bestaat voornamelijk uit zorgvraagtyperingen met grote aantallen patiënten en relatief kleine aantallen patiënten met één van de kenmerken. In veel gevallen heeft minder dan 5% van de zorgvraagtyperingen in een zorgvraagtype een kenmerk. Om deze reden zijn ZT01-ZT03 niet te zien in de figuren en zijn ZT04-ZT07 alleen te zien in het paneel met één of meerdere kenmerken.

Figuur 43



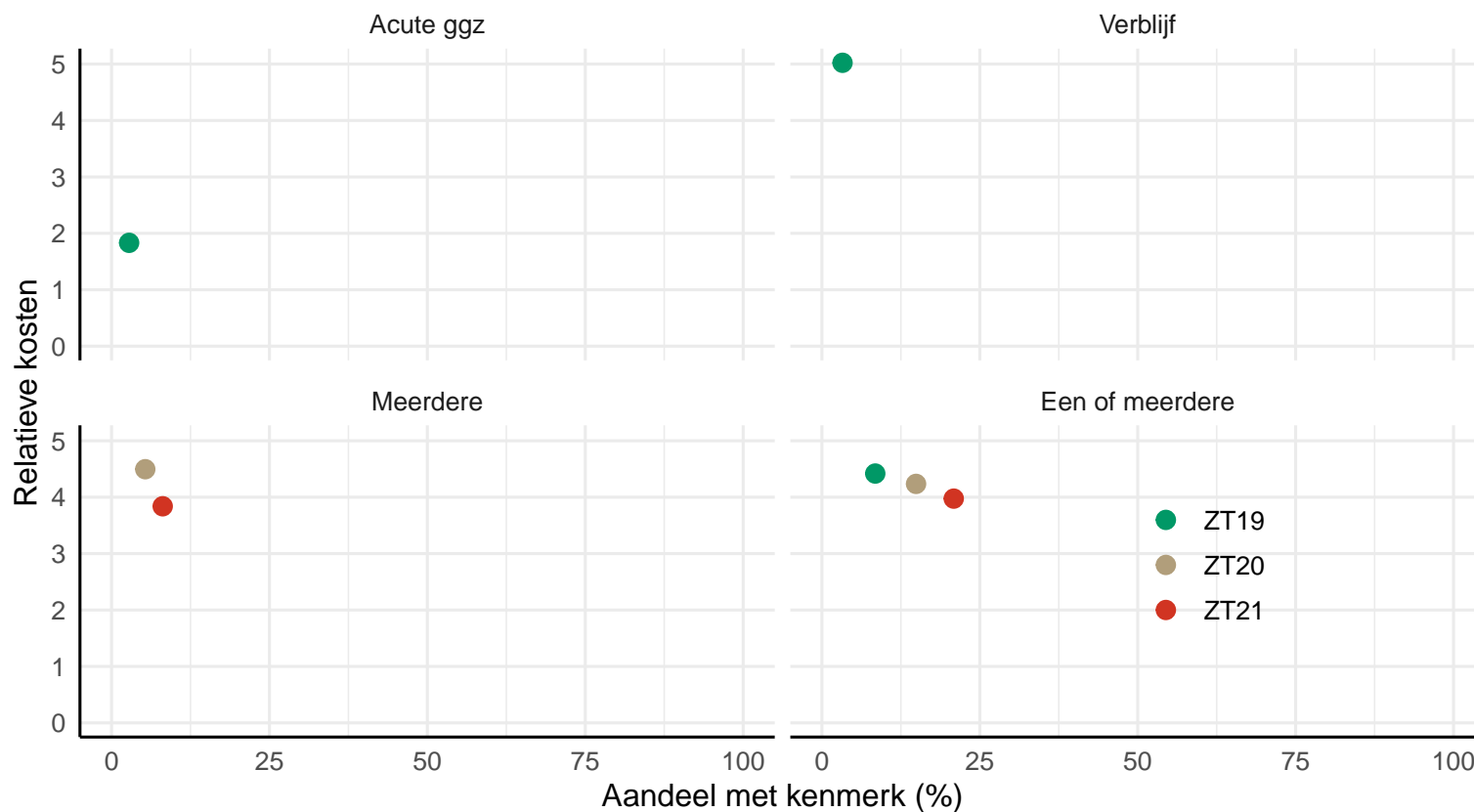
**Relatieve kosten en aandeel - Y** De kenmerken komen in hoofdgroep Y veel vaker voor dan in hoofdgroep X, in ZT10 en ZT14-17 komt zelfs in 35-45% van de zorgvraagtyperingen één of meerdere kenmerken voor. De gemiddelde kosten voor de groepen met het kenmerk liggen doorgaans 1,5-3 keer hoger dan de kosten voor de groep zonder het kenmerk.

Figuur 44



**Relatieve kosten en aandeel - Z** Hoofdgroep Z bevat slechts een klein aantal patiënten en de kenmerken komen slechts zelden voor. In geen van de gevallen zijn er meer dan 250 zorgvraagtyperingen met het kenmerk, zoals ook te zien is in tabel 9. Gezien de al kleine groepen is het de vraag of het zinvol is zorgvraagtyperingen in hoofdgroep Z te splitsen. Voor de volledigheid visualiseren we toch de groepen waar minimaal 50 zorgvraagtyperingen het kenmerk hadden en dit minimaal 2,5% van de trajecten met het zorgvraagtype was.

Figuur 45



Tabel 10: Aanvullende kenmerken die voldoen aan selectiecriteria

Zorgvraagtype	Verzwarend	Aantal met verzwarend	Aandeel van zorgvraagtype (%)	Kosten met kenmerk (€)	Kosten zonder kenmerk (€)
ZT04	Een of meerdere	3894	5	9954	3261
ZT05	Een of meerdere	2654	11	14876	5890
ZT06	Een of meerdere	1576	10	14430	5601
ZT07	Een of meerdere	5636	7	10600	3647
ZT08	Meerdere	491	10	23079	10002
ZT08	Een of meerdere	1186	24	19033	8830
ZT10	Een of meerdere	344	36	13771	10391
ZT11	Een of meerdere	1393	13	8270	3197
ZT12	Meerdere	731	6	18858	4912
ZT12	Een of meerdere	2084	18	11704	4499
ZT13	Meerdere	476	11	25231	8393
ZT13	Een of meerdere	1214	27	17471	7486
ZT14	Meerdere	503	22	28711	15895
ZT14	Een of meerdere	979	43	23502	15130
ZT16	Meerdere	642	19	27175	10763
ZT16	Een of meerdere	1228	36	21254	9678
ZT17	Meerdere	367	21	30469	12173
ZT17	Een of meerdere	704	41	22287	11779

### 2.7.6 Kenmerken en kwaliteitsparameters

Als we de kenmerken gebruiken om zorgvraagtypes op te splitsen heeft dit een impact op de kwaliteit van het model, groepen die zich onderscheiden op het gebied van kosten kosten zorgen immers voor een betere voorspelling. Dit zou te meten moeten zijn met de kwaliteitsparameters die we in hoofdstuk 4 beschrijven. In de werkgroep data en de adviescommissie is besproken dat we voor het splitsen van zorgvraagtypes met de aanvullende kenmerken twee varianten zullen doorrekenen. De eerste is dat we de trajecten met één of meerdere kenmerken afsplitsen van het zorgvraagtype, de tweede variant is dat we de trajecten afsplitsen waarvoor minimaal twee van de kenmerken gelden. In beide gevallen gebruiken we de afkappunten die we ook hierboven hebben beschreven, namelijk dat er minimaal 5% van het zorgvraagtype afgesplitst moet kunnen worden en dat dit minimaal 250 zorgtrajecten zijn. In tabel 10 is te zien bij welke zorgvraagtypes en kenmerken dat het geval is, hoe vaak ze voorkomen en wat de kosten zijn in de groep met of zonder kenmerk(en).

Tabel 11: Kwaliteitsparameters bij splitsen obv aanvullende kenmerken

Voorspeller	MAE (€)	Verklaarde basisfout ( )
Meerdere	1234	40,6
Een of meerdere	1138	45,1
Zvt	1270	39,1

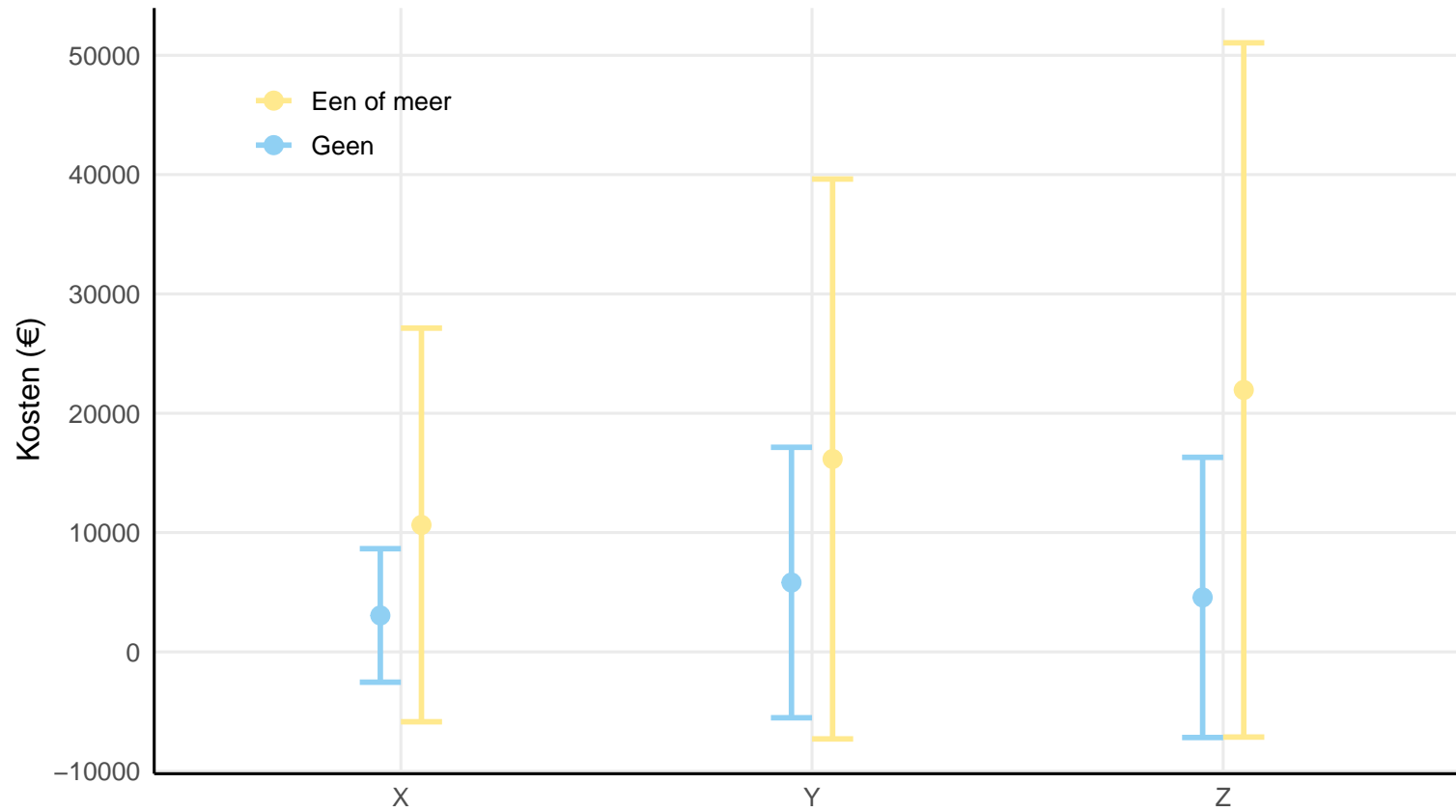
In tabel 11 is te zien hoe een splitsing van een aantal zorgvraagtypes op basis van ofwel meerdere aanvullende kenmerken ofwel één of meerdere aanvullende kenmerken de kwaliteitsparameters beïnvloeden. Bij het splitsen op basis van *meerdere* kenmerken zijn er 6 zorgvraagtypes opgesplitst. De gemiddelde absolute fout daalt hiermee iets en de verklaarde basisfout is iets hoger, het model wordt dus iets beter van de splitsingen. Bij het splitsen op basis van *één of meerdere kenmerken* worden 12 zorgvraagtypes gesplitst. Dit leidt tot een extra daling van de gemiddelde absolute fout en vooral een flinke stijging van de verklaarde basisfout. Dit is naar verwachting het gevolg van een betere voorspelling van de kosten bij de grote uitschieters.

### 2.7.7 Kenmerken en samengevoegde zorgvraagtypes

De adviescommissie heeft voorgesteld om te onderzoeken of er een goede manier is om zorgvraagtypes samen te voegen en vervolgens te splitsen of er voor het zorgtraject sprake is van één of meerder van de hierboven besproken kenmerken. Enkele varianten van splitsen en samenvoegen zijn denkbaar. Hieronder werken we enkele varianten uit.

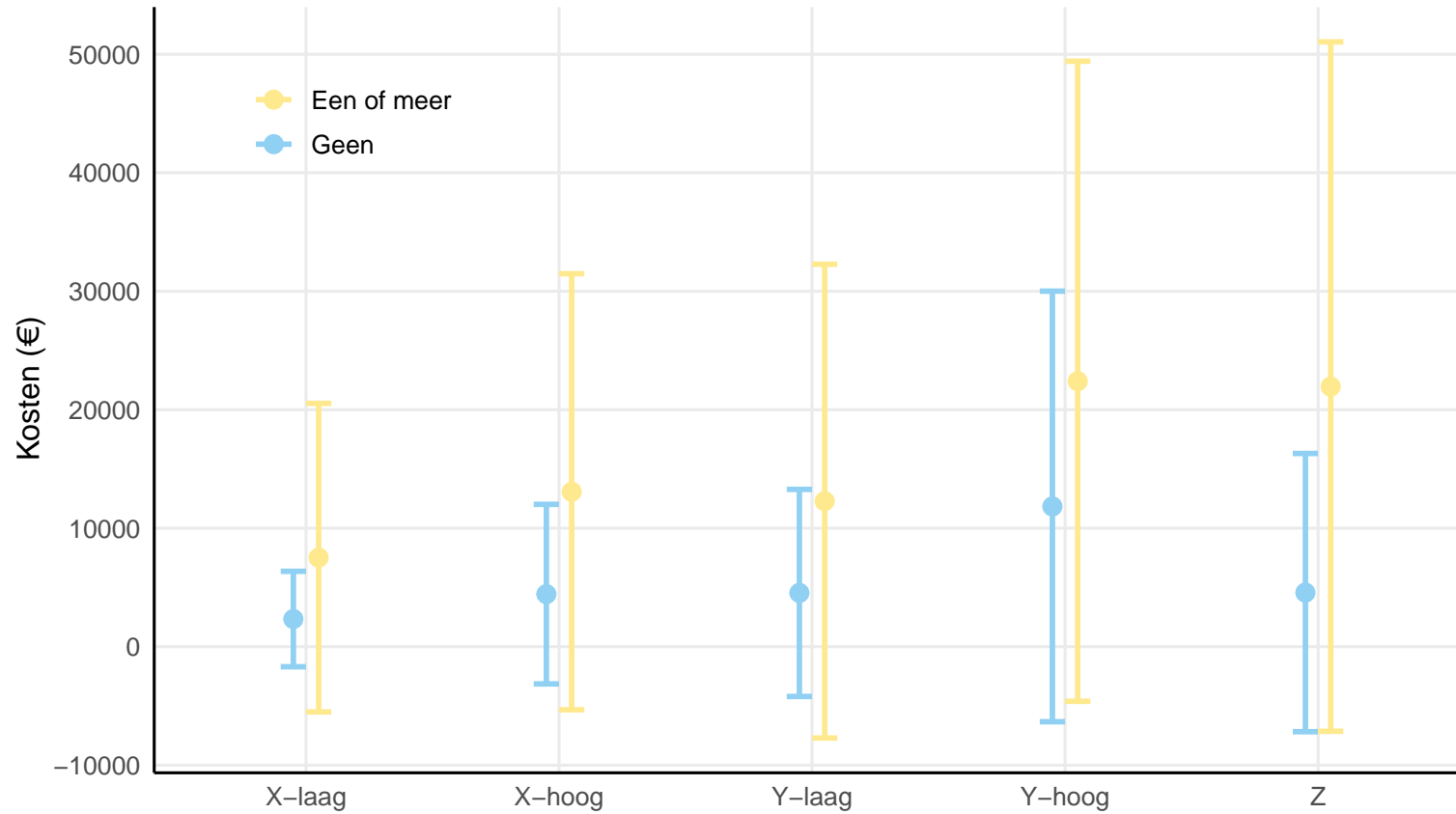
**Samenvoegen tot hoofdgroepen en splitsen** De eenvoudigste variant is simpelweg de hoofdgroepen samenvoegen tot één grote groep en deze vervolgens splitsen op een groep met en een groep zonder één of meerdere van de kenmerken. Hieronder visualiseren we de gemiddelde kosten en spreiding van de groepen die we op deze manier verkrijgen.

Figuur 46



**X en Y samevoegen tot hoog en laag** Alternatief kunnen we samenvoegen op het niveau van groepen binnen een hoofdgroep. Op verzoek van de adviescommissie hebben we gekozen voor de groepen zoals ze hierboven ook in afzonderlijke subplots zijn weergegeven. Specifiek voegen ZT01-ZT04 samen tot X-laag, ZT05-ZT08 tot X-hoog, ZT10-ZT13 tot Y-laag, ZT14-ZT17 tot Y-hoog en hoofdgroep Z.

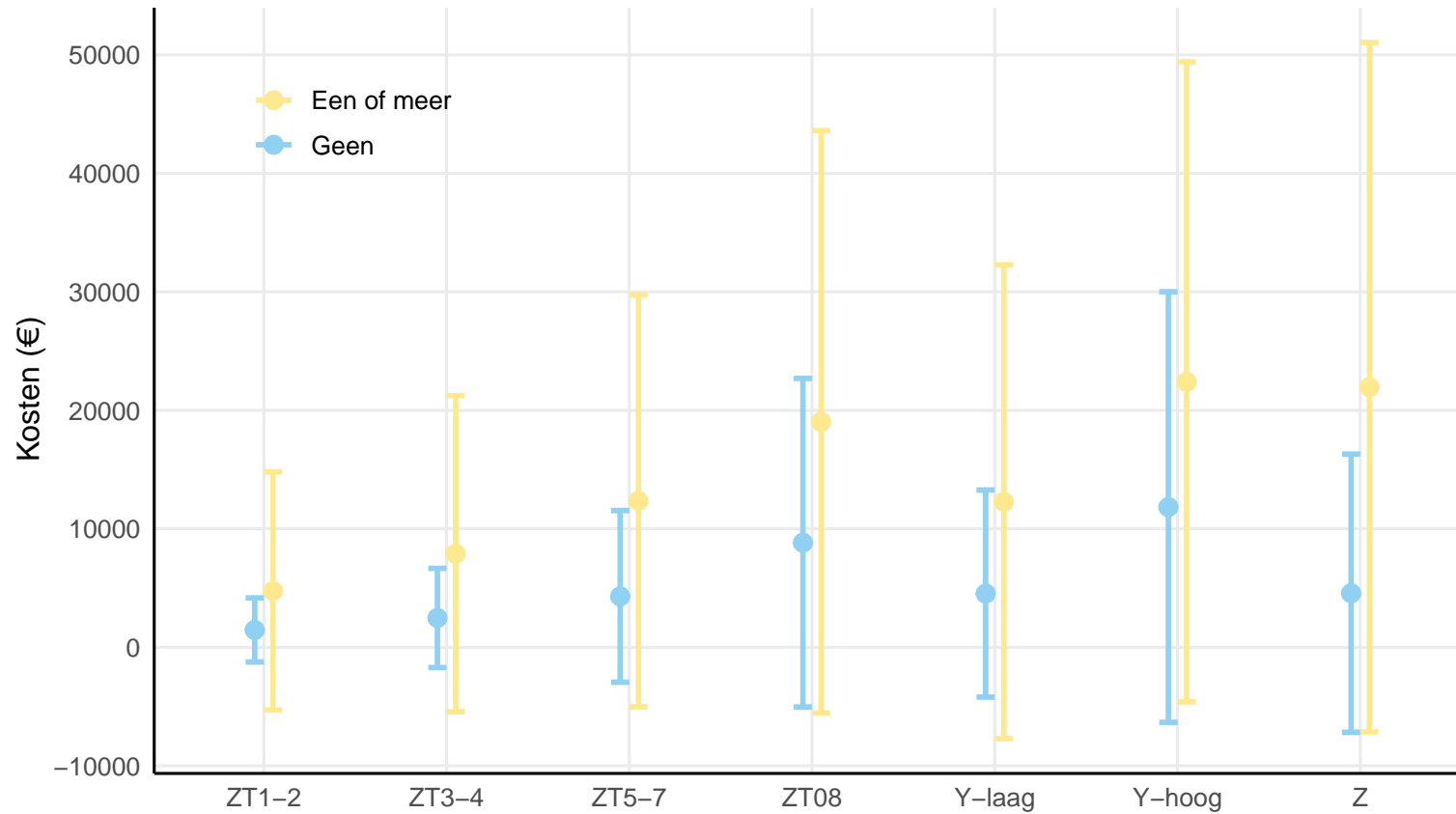
Figuur 47





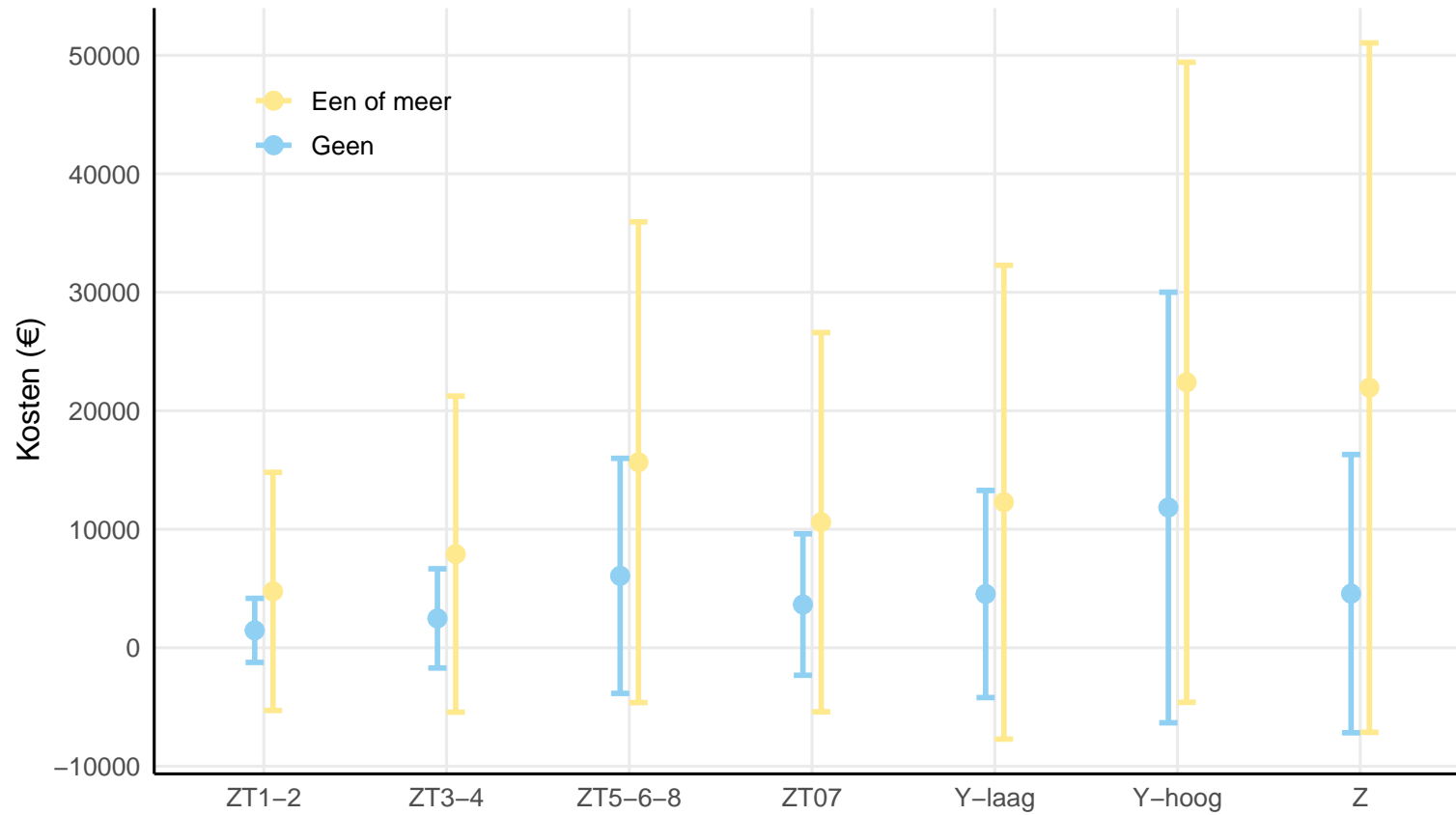
**Expert indeling 1** Daarnaast zijn er door zorgexperts uit de werkgroep data twee versies van samenvoegingen van de zorgvraagtypes in hoofdgroep X voorgesteld. De eerste variant geeft in hoofdgroep X de groepen ZT1-2, ZT3-4, ZT5-7 en ZT08. Dit is vanuit de gedachte dat in ZT08 de stoornis zich chaotischer uit en dat dit leidt tot aanzienlijk hogere kosten dan ZT5-7. Deze indeling combineren we met een samenvoeging tot Y-laag en Y-hoog zoals hierboven.

Figuur 48



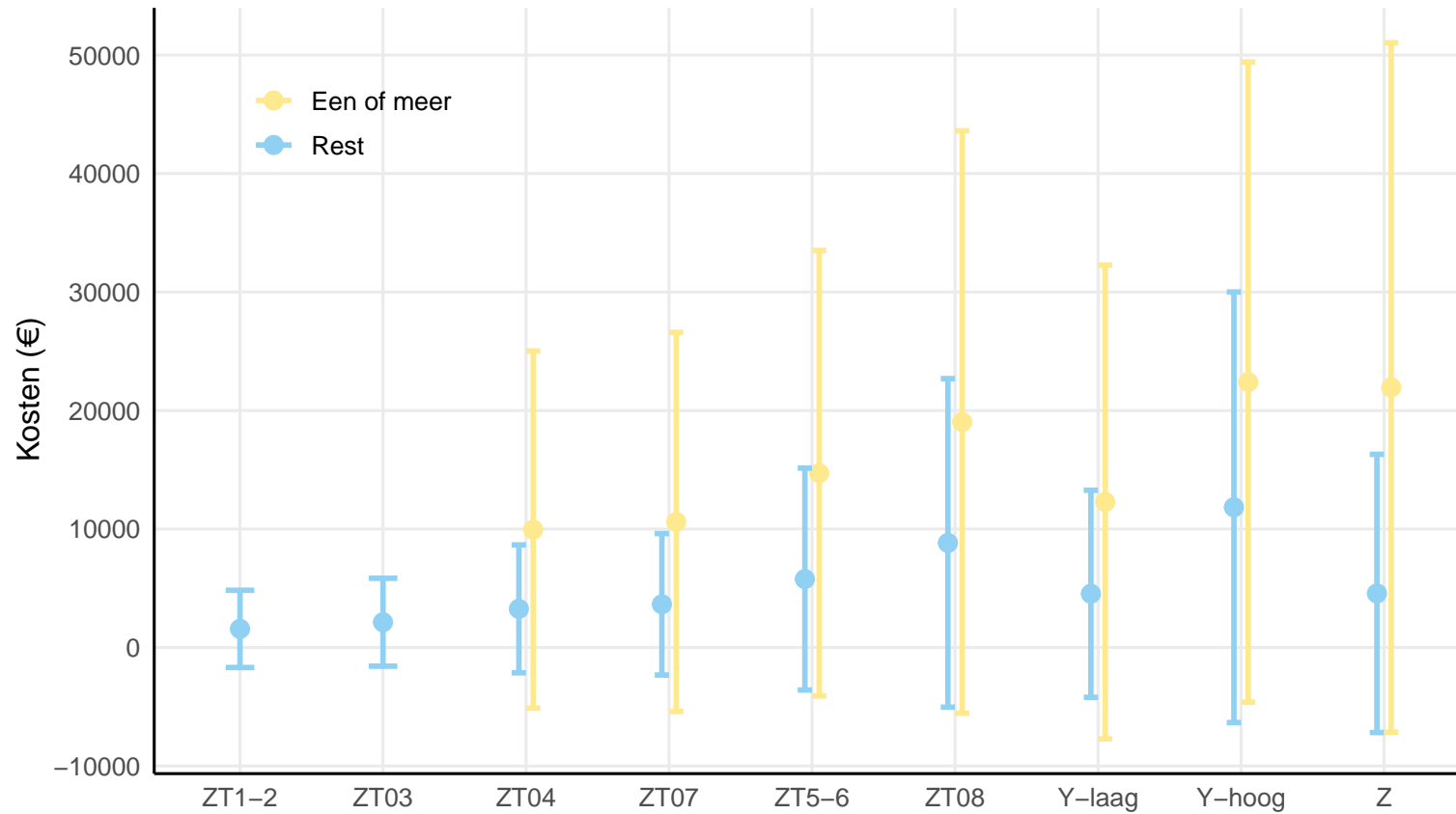
**Expert indeling 2** De tweede expertvariant geeft in hoofdgroep X de groepen ZT1-2, ZT3-4, ZT5-6-8 samen en juist ZT07 als losse groep. In groep ZT07 is grilligheid minder prominent dat de andere zorgvraagtypes tussen de complexe zorgvraagtypes in hoofdgroep X.

Figuur 49



**Expert indeling 3** De laatste expertvariant voegt in hoofdgroep X de groepen ZT01 en ZT02 samen en ZT05 met ZT06, de rest blijft gescheiden, omdat de groepen al groot zijn en/of verschillende kosten met zich mee brengen. Voor hoofdgroepen Y en Z wordt hetzelfde gedaan als bovenstaande groepen. Voor de groepen met lichte en matige problematiek wordt geen splitsing aangebracht op basis van de aanvullende kenmerken, voor de overige groepen wordt dit wel gedaan. Let op dat de volgorde van de zorgvraagtypes anders is dan in voorgaande figuren. Dit omdat de werkgroep deze indeling als alternatief wilt meegeven aan de adviescommissie en daarbij een alternatieve volgorde in gedachten heeft.

Figuur 50



Tabel 12: Kwaliteitsparameters bij samenvoegen gecombineerd met splitsen obv aanvullende kenmerken

Voorspeller	MAE (€)	Verklaarde basisfout ( )
Hoofdgroep	1301	33,7
X-Y hoog-laag	1163	42,9
Expertinput 1	1157	43,5
Expertinput 2	1147	44,6
Expertinput 3	1140	44,9
Zvt	1270	39,1
Zvt met kenmerken	1138	45,1

**Kwaliteitsparameters met samenvoegen en splitsen** Voor elk van de hierboven beschreven samenvoelingen hebben we de kwaliteitsparameters doorgerekend, voor elke variant zijn we uitgegaan van een combinatie met splitsing op het al dan niet aanwezig zijn van één of meerdere kenmerken. Ter vergelijking hebben we ook de zorgvraagtypering zoals hij nu op de factuur staat opgenomen (Zvt) en de splitsing van de zorgvraagtypes volgens de één of meerdere variant uit tabel 11 (Zvt met kenmerken). Het samenvoegen tot hoofdgroepen zorgt voor een aanzienlijk slechtere verklaarde basisfout en een iets hogere MAE. Samenvoelingen van X en Y tot een hoge en een lage groep in combinatie met voor elke groep een splitsing op de aanvullende kenmerken heeft een lagere MAE en een hogere verklaarde basisfout dan zorgvraagtypering op zichzelf. Voor de twee expertvarianten geldt dat deze op de kwaliteitsparameters steeds dichterbij de variant zoals beschreven in tabel 11.

### 2.7.8 Andere alternatieve indelingen

We hebben op de gegevens verschillende andere machine learning methodes toegepast om te onderzoeken of vanuit de data andere indelingen in zorgvraagtypes opgebouwd kan worden. Deze methodes hebben niet geleid tot indelingen die bruikbaar lijken in de praktijk. Enerzijds waren er indelingen op te bouwen met clusteringmethodes die zeer homogeen zijn in het zorggebruik, maar deze indeling is niet terug te leiden tot een stabiele indeling in het domein van de zorgvraag. Aan de andere kant is het mogelijk andere indelingen te vinden die herkenbaar lijken in het domein van de zorgvraag, maar deze groepen zijn niet beter ingedeeld in het domein van de zorginzet. De analyses zijn gecompliceerd en vragen veel figuren en uitleg. Dit leidt af van de analyses die wel relevant zijn. De NZa laat deze analyses daarom achterwege.

### 2.7.9 Algoritme bij expertindeling 3

Het is nodig de algoritmes opnieuw te schatten wanneer de indeling in zorgvraagtypes verandert, om de nieuwe grenzen tussen de zorgvraagtypes goed te bepalen. Dit hebben we gedaan en afgezet tegen de voorspelling op basis van het oude algoritme. Het is hierbij nodig enkele rode regels uit te schakelen, zodat niet onterecht bepaalde 'nieuwe' zorgvraagtypes worden uitgesloten, wanneer er verschillende rode regels bestaan voor de 'oude' zorgvraagtypes die in de nieuwe groep vallen. Het nieuwe algoritme lijkt iets lagere correcte voorspellingen te halen dan het oude algoritme, maar heeft niet de mogelijkheid gehad de behandelaar te beïnvloeden. Verder valt op dat het minder vaak voorkomt dat er geen voorspelling gedaan wordt door het algoritme, dit is vermoedelijk het gevolg van het uitsluiten van enkele rode regels.

Tabel 13: Prestatie voorspelling algoritme expertindeling 3 of oud

keuze	kenmerk	Goed voorspeld nieuw ( )	Goed voorspeld oud ( )	Geen voorspelling nieuw ( )	Geen voorspelling oud ( )
ZT1-2	Nvt	78.1	78.1	15.6	15.8
ZT03	Nvt	87.9	90.9	1.2	1.2
ZT04	Ja	76.9	79.1	3.1	5.9
ZT04	Nee	85.2	87.8	0.7	1.0
ZT07	Ja	75.1	79.5	2.1	3.0
ZT07	Nee	81.3	82.7	0.7	1.0
ZT5-6	Ja	79.3	85.8	2.3	6.7
ZT5-6	Nee	79.9	86.6	1.4	2.8
ZT08	Ja	88.4	87.4	2.9	6.4
ZT08	Nee	88.5	87.9	3.1	6.3
Y-laag	Ja	88.3	92.4	0.0	0.0
Y-laag	Nee	94.0	95.9	0.8	0.0
Y-hoog	Ja	85.0	88.6	0.0	0.0
Y-hoog	Nee	81.8	87.7	0.0	0.0
Z	Ja	100.0	97.7	0.0	2.3
Z	Nee	100.0	96.1	0.0	3.9

## 2.8 Individuele items

We bekijken in welk aandeel van de scorelijsten scores per item voorkomen bij de verschillende zorgvraagtypes. Het blijkt dat sommige scores op bepaalde items bij sommige zorgvraagtypes vaker voorkomen dan bij andere zorgvraagtypes. Deels is dit het gevolg van de rode regels (zie ook figuren 24 - 26). Bij de dynamische zorgvraagtypering komen niet alle items voor, om niet te vertekenen nemen we de dynamische zorgvraagtyperingen niet mee in deze figuren.

De figuren bevatten zeer veel informatie, die niet goed in tekst kan worden gevat. We laten tekstuele beschrijving bij deze figuren dan ook achterwege.

2.8.1 Scores per item - hoofdgroep X

Figuur 51





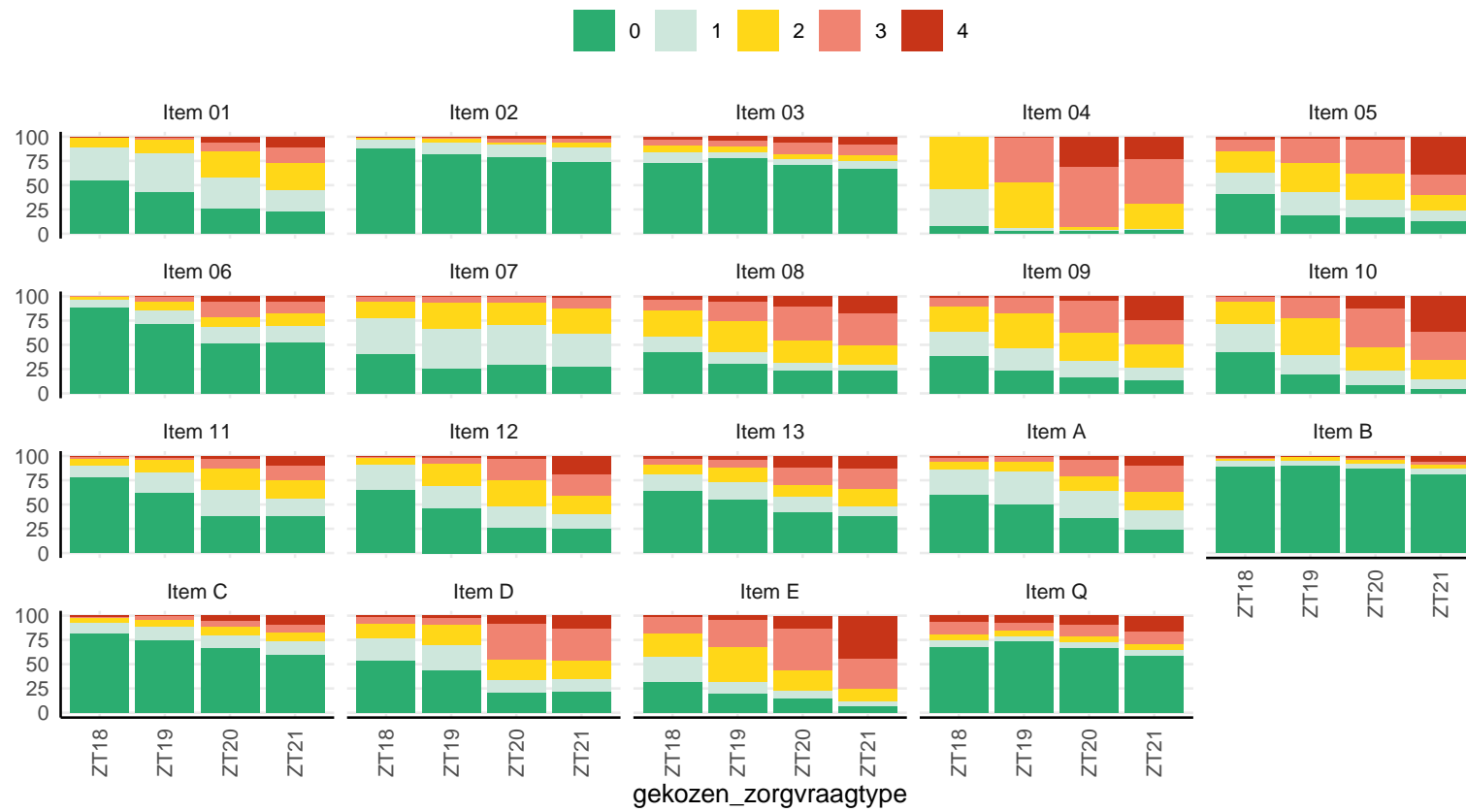
2.8.2 Scores per item - hoofdgroep Y

Figuur 52



2.8.3 Scores per item - hoofdgroep Z

Figuur 53



## 2.9 Vragen en kosten binnen zorgvraagtypering

Het is mogelijk dat hogere scores op specifieke items binnen een zorgvraagtype duiden op hogere kosten, bijvoorbeeld omdat een bepaald symptoom vraagt om andere zorginzet dan afwezigheid van het symptoom. Als dit zo is dan zouden specifieke items mogelijk gebruikt kunnen worden om zorgvraagtypes te splitsen in twee kostenhomogenere groepen. We onderzoeken dit door per zorgvraagtype een lineair model te schatten waarin de kosten voorspeld worden op basis van de score op één van de items. Dit doen we voor alle items en alle zorgvraagtypes. Een hoge coëfficiënt duidt op een sterke positieve relatie tussen de score op het item en de kosten binnen het zorgvraagtype. Deze analyse volgt uit vragen uit de werkgroep data of bepaalde zorgvraagtypes (specifiek ZT03, ZT04 en ZT07) mogelijk gesplitst zouden kunnen worden op basis van de scores op één item.

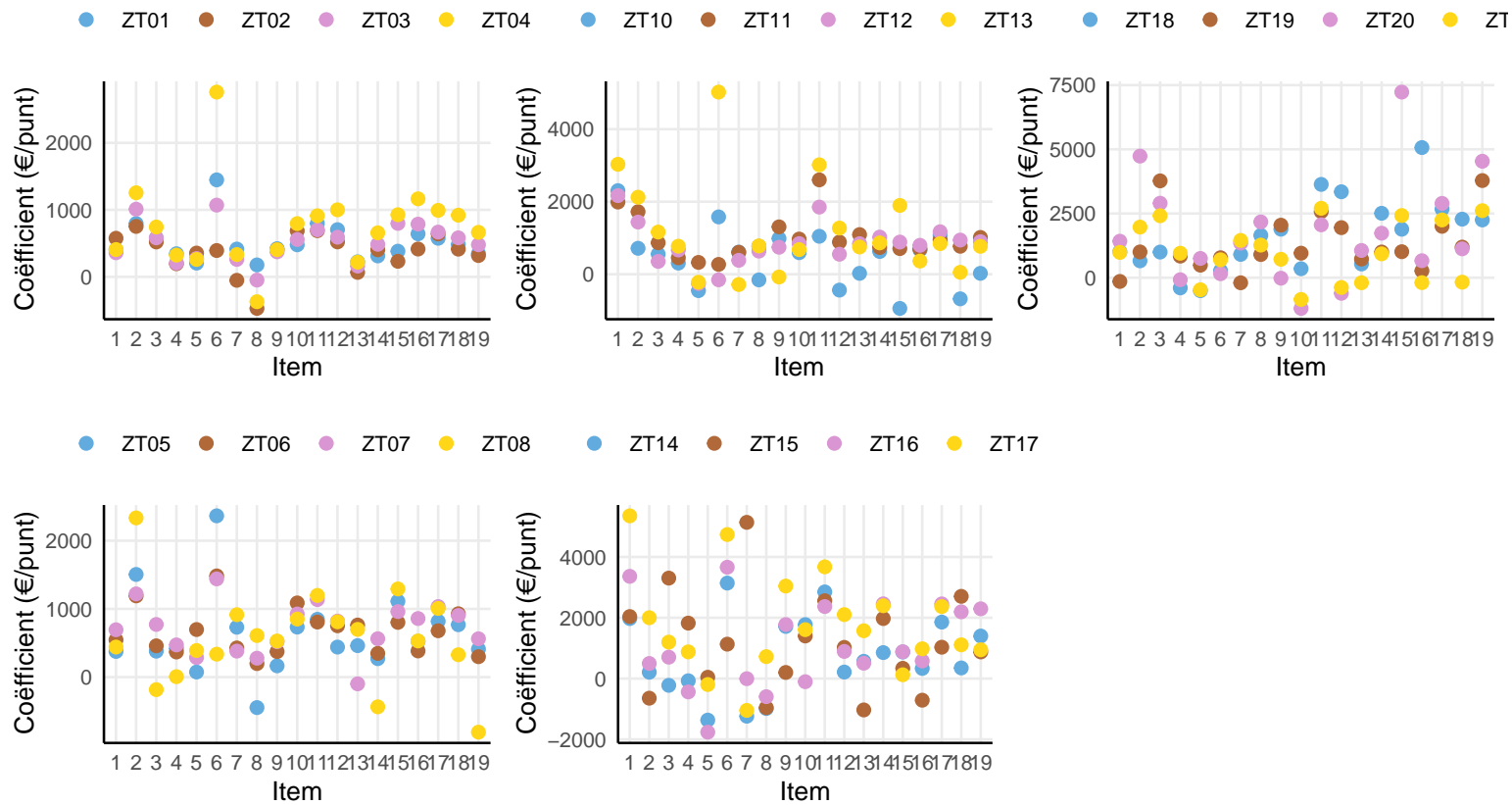
Houd bij deze analyses rekening dat er meer nodig is voor het splitsen van een zorgvraagtypering dan een (sterke) positieve relatie tussen score en kosten. Zo is het bijvoorbeeld ook nodig dat een voldoende groot aandeel zorgvraagtyperingen met een hogere score bestaat binnen een zorgvraagtype en dat de splitsing de proportionaliteitsafweging rond de privacy doorstaat.

Splitsing op basis van items of subschalen lijkt minder op te kunnen leveren dan het gebruik van de aanvullende kenmerken. De adviescommissie ziet ook meer in de route van splitsing met aanvullende kenmerken. Analyses uit dit deel van het document zijn daarom niet verder uitgediept.

### 2.9.1 Correlatie tussen score en kosten per zorgvraagtype

Voor de interpreteerbaarheid beschouwen we voor deze analyse de score op de items als numerieke waarden, in plaats van de categorische variabele die het eigenlijk is. Daarnaast gebruiken we alleen volledige zorgvraagtyperingen, zodat de schattingen onderling goed te vergelijken zijn.

Figuur 54



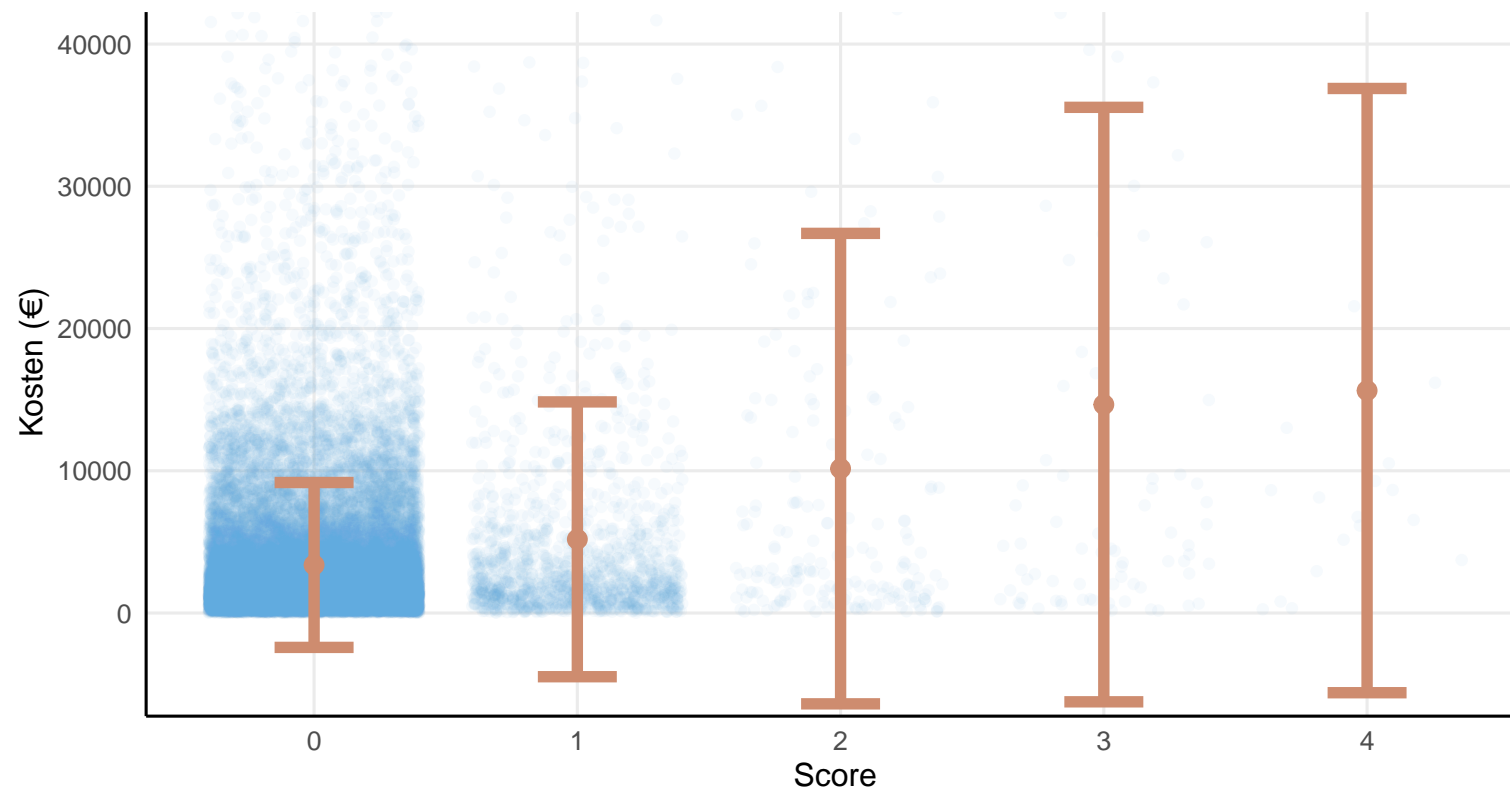
### 2.9.2 Voorbeeld klein aantal hoge scores

Voor ZT04 lijkt er een sterke relatie te zijn tussen de score op item 6 en de kosten. Wanneer we dit in meer detail bekijken valt op dat er in ZT04 maar heel weinig zorgvraagtyperingen zijn met andere scores dan 0 op item 6. Dit maakt item 6 vermoedelijk ongeschikt om een eventuele splitsing van ZT04 op te baseren. Het aantal zorgvraagtyperingen met een score is hierin bijna net zo relevant als de relatie tussen kosten en score.

Vanwege privacy zijn de buitenste percentielen van de kosten per score niet weergegeven.

Figuur 55

Relatie kosten en score voor ZT04 en item 6

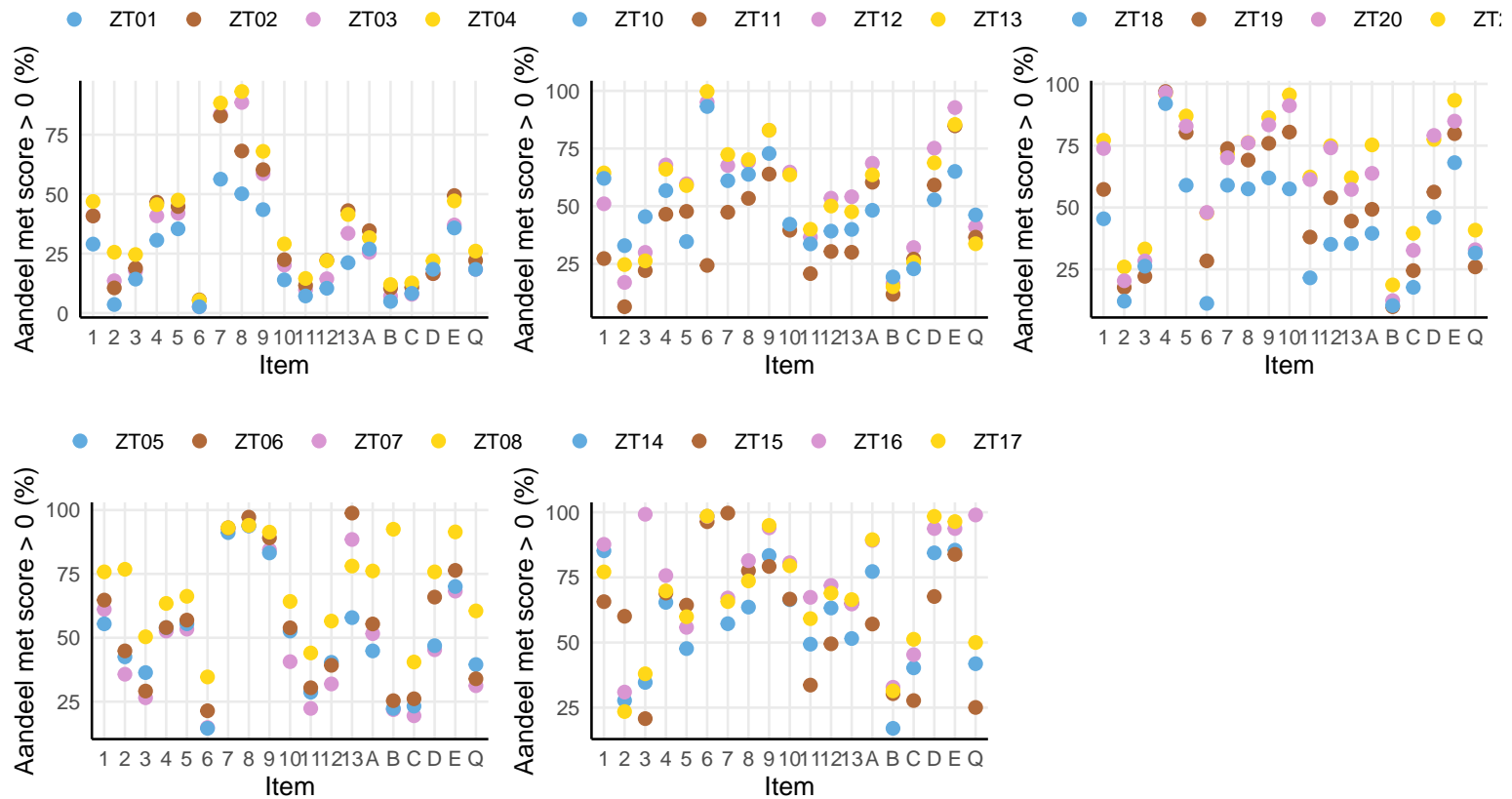


### 2.9.3 Aandeel scorend op item

Uit voorgaande figuur blijkt dat het belangrijk is ook oog te houden voor de verdeling van de scores binnen het zorgvraagtype. Dit gaat het overzichtelijkst door het zorgvraagtype te splitsen in twee groepen, één met en één zonder een score op een item. Dit operationaliseren we op twee verschillende manieren, 1) een patiënt scoort op een item als de score  $> 0$  of 2) een patiënt scoort op een item als de score  $> 2$ . Dit zijn arbitraire afkappunten, dus het is goed hierover een discussie te voeren in de werkgroep en adviescommissie.

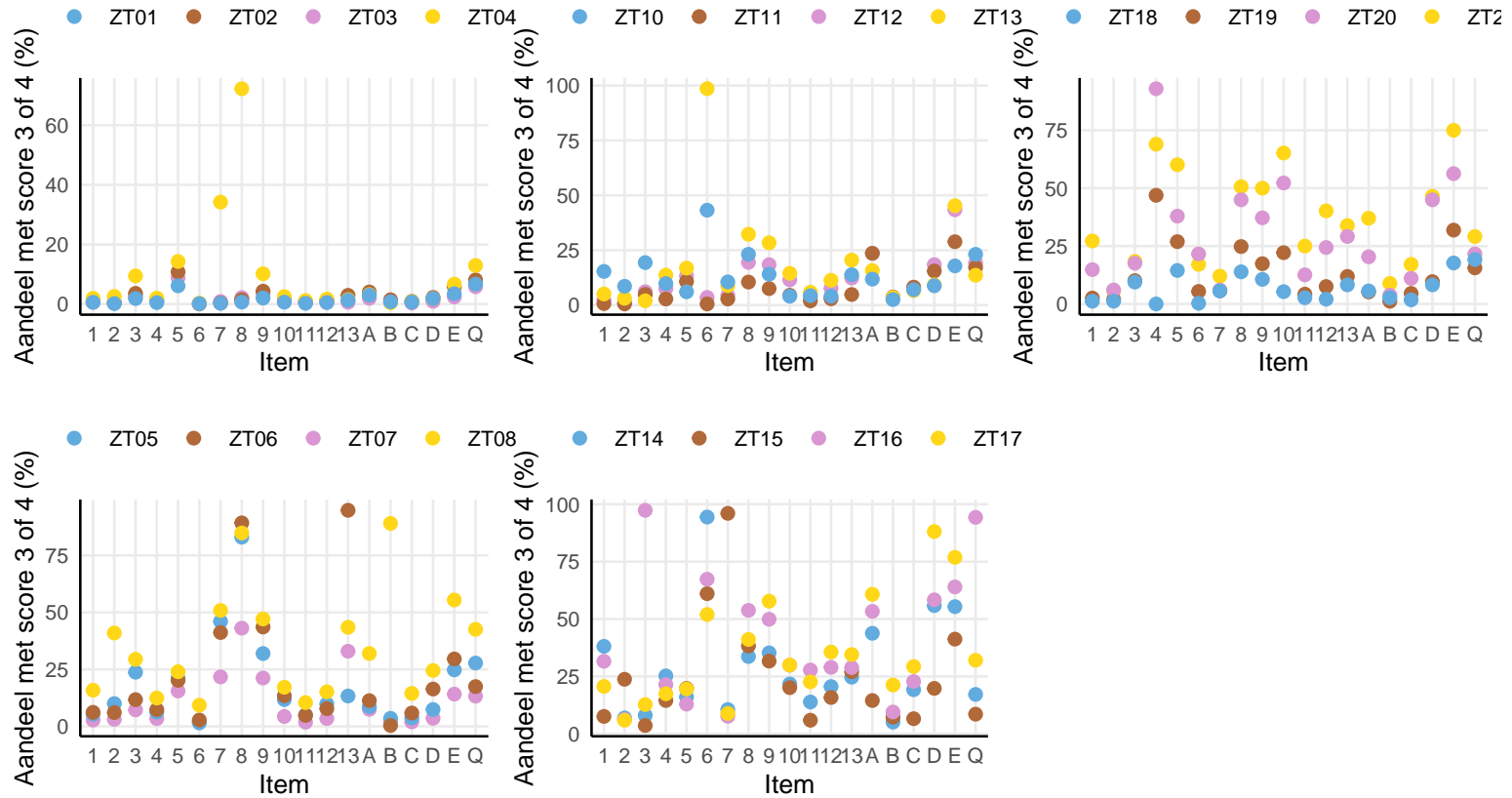
**Aandeel score > 0** Wanneer enige klachten op een item leiden tot een hoger zorgvraagtype kan dit een reden zijn een score > 0 op het item te overwegen om een zorgvraagtype op te splitsen. In dit geval is het dan ook belangrijk te weten in welk aandeel van de zorgvraagtypering een score > 0 voor komt. Het is logisch dat enige klachten (score > 0) op de items met enige regelmaat voorkomen en dit blijkt ook uit de figuren. Uit het voorkomen in isolatie is nog niet veel op te maken over of het al dan niet logisch is te splitsen op basis van de score op het item. Er is ook informatie nodig over de relatieve kosten van de twee groepen. Dit wordt verdiept vanaf figuur 58.

Figuur 56



**Aandeel score 3 of 4** Het valt op dat slechts zelden een significant aandeel van zorgvraagtypes ZT01-ZT04 een score van 3 of 4 hebben op enig item. Hogere scores komen relatief vaker voor bij ZT10-13, maar nog steeds niet vaak in hogere percentages. Bij ZT05-ZT08, ZT14-ZT17 en ZT18-ZT21 komen scores van 3 of 4 regelmatig voor bij een significant aandeel van de scorelijsten.

Figuur 57





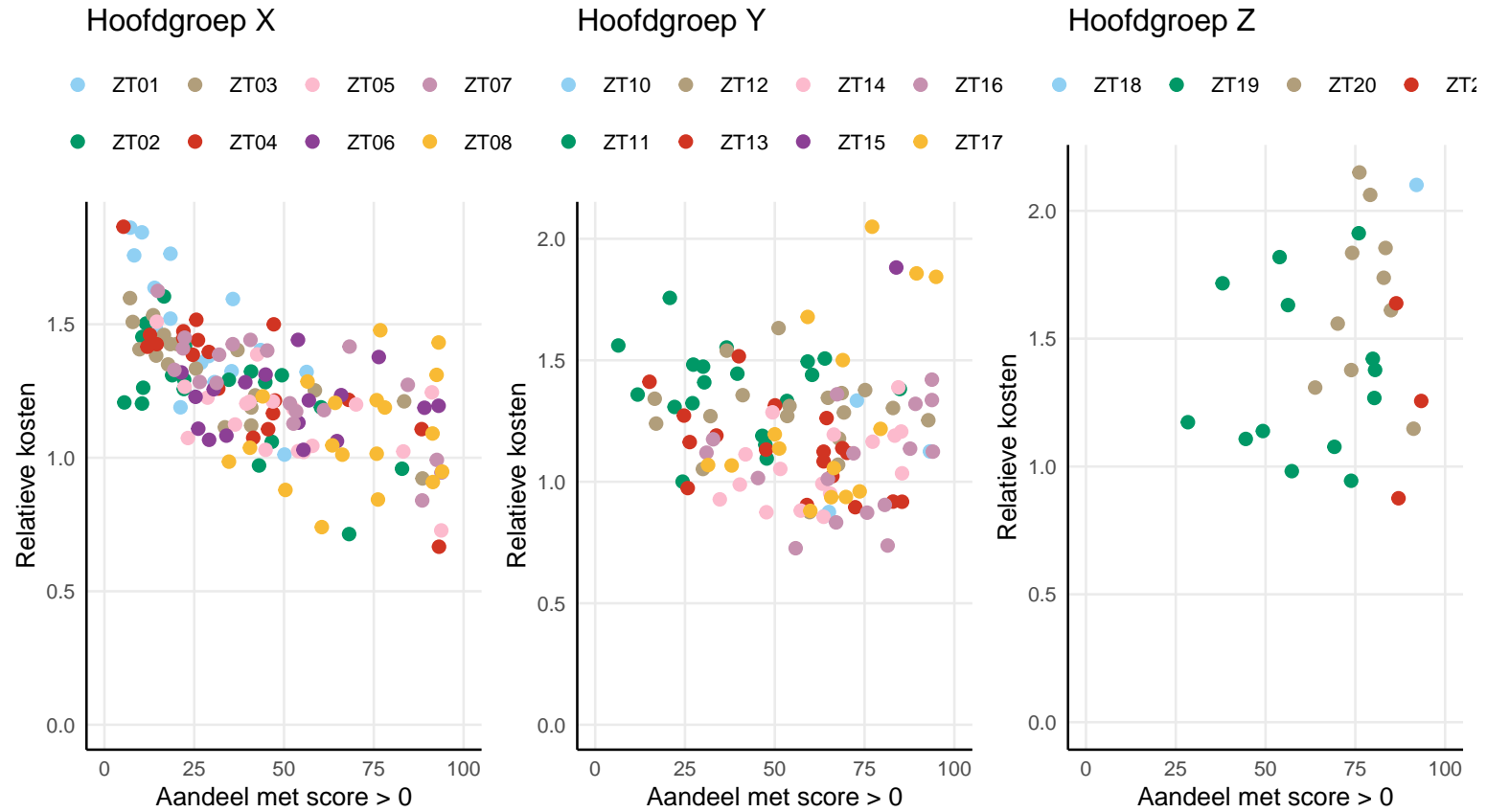
#### 2.9.4 Kosten en aandeel bij splitsing - score > 0

Net zoals bij de kenmerken mbt verblijf, acute ggz of Wvggz/Wzd in het afgelopen jaar is de haalbaarheid van een eventuele splitsing van een zorgvraagtype op basis van de items afhankelijk van een aantal elementen. In ieder geval moeten de kosten significant verschillen tussen de groep die wel en de groep die niet scoort op het item en beide groepen moeten een significant aandeel van het zorgvraagtype uitmaken. Voor elk zorgvraagtype en item hebben we deze twee elementen berekend.

NB in onderstaande figuren zijn wederom de groepen met minder dan 250 scorelijsten of minder dan 5% of meer dan 95% van de zorgvraagtyperingen weggelaten, om de leesbaarheid van de figuren iets te vergroten.

**Score > 0** Het is te zien dat de kosten van de groep die wel scoort op een item iets hoger liggen, maar dat dit in veel gevallen geen groot verschil geeft tussen de groep met en zonder score op het item. Uit het figuur wordt niet direct duidelijk bij welke items het al dan niet scoren een relatie lijkt te hebben met de kosten, zie daarvoor tabel 14 hieronder.

Figuur 58



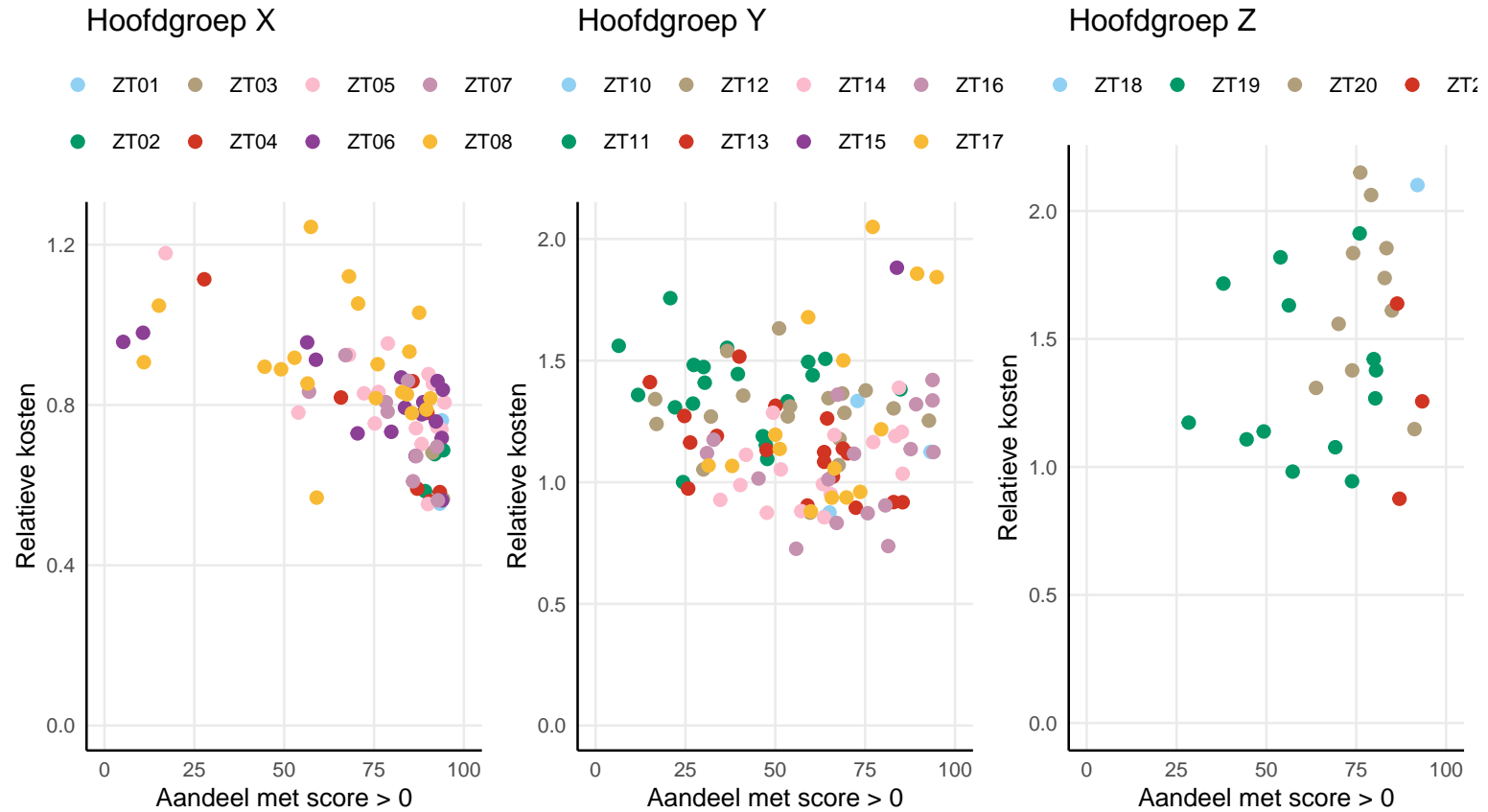
**Score > 0 - tabel** Hieronder laten we de zorgvraagtypes en items zien waarvoor de relatieve kosten minstens 1,5 keer zo hoog zijn voor de groep waarin de score groter is dan 0 ten opzichte van de groep met score 0 én waarvan minstens 250 scorelijsten bestaan én die minstens 10% en maximaal 90% van het zorgvraagtype uitmaakt.

Tabel 14: Selectie van zorgvraagtypes en items met hoge relatieve kosten wanneer de score > 0

Zvt	Item	Beschrijving	Aantal	Aandeel (%)	Relatieve kosten
ZT12	Item 01	Hyperactief, agressief, destructief of geagiteerd gedrag	2218	51.0	1.6
ZT17	Item 01	Hyperactief, agressief, destructief of geagiteerd gedrag	623	77.0	2.0
ZT03	Item 02	Opzettelijke zelfverwonding (suïcidepogingen, gedachten over suïcide, automutilatie)	7820	14.0	1.5
ZT04	Item 02	Opzettelijke zelfverwonding (suïcidepogingen, gedachten over suïcide, automutilatie)	8440	26.0	1.5
ZT20	Item 05	Lichamelijke problemen of handicaps	330	83.0	1.7
ZT05	Item 06	Problemen als gevolg van hallucinaties en waanvoorstellingen	1265	15.0	1.5
ZT07	Item 06	Problemen als gevolg van hallucinaties en waanvoorstellingen	5651	15.0	1.6
ZT20	Item 07	Problemen met depressieve stemmingen	279	70.0	1.6
ZT20	Item 08	Overige psychische en gedragsproblemen	303	76.0	2.1
ZT11	Item 09	Problemen met relaties	2861	64.0	1.5
ZT19	Item 09	Problemen met relaties	717	76.0	1.9
ZT20	Item 09	Problemen met relaties	332	83.0	1.8
ZT21	Item 09	Problemen met relaties	273	86.0	1.6
ZT01	Item 10	Problemen met ADL	1612	14.0	1.6
ZT02	Item 11	Problemen met woonomstandigheden	667	12.0	1.5
ZT11	Item 11	Problemen met woonomstandigheden	930	21.0	1.8
ZT12	Item 11	Problemen met woonomstandigheden	1591	37.0	1.5
ZT13	Item 11	Problemen met woonomstandigheden	722	40.0	1.5
ZT17	Item 11	Problemen met woonomstandigheden	478	59.0	1.7
ZT19	Item 11	Problemen met woonomstandigheden	359	38.0	1.7
ZT01	Item 12	Mogelijkheden voor het gebruiken en verbeteren van vaardigheden: beroepsmatig en vrije tijd	1205	10.0	1.8
ZT17	Item 12	Mogelijkheden voor het gebruiken en verbeteren van vaardigheden: beroepsmatig en vrije tijd	557	69.0	1.5
ZT19	Item 12	Mogelijkheden voor het gebruiken en verbeteren van vaardigheden: beroepsmatig en vrije tijd	509	54.0	1.8
ZT20	Item 12	Mogelijkheden voor het gebruiken en verbeteren van vaardigheden: beroepsmatig en vrije tijd	295	74.0	1.8
ZT17	Item A	Historisch: Geagiteerd gedrag / expansieve stemmingen	723	89.0	1.9
ZT01	Item D	Historisch: Betrokkenheid/motivatie	2124	18.0	1.8
ZT02	Item D	Historisch: Betrokkenheid/motivatie	952	17.0	1.6
ZT19	Item D	Historisch: Betrokkenheid/motivatie	531	56.0	1.6
ZT20	Item D	Historisch: Betrokkenheid/motivatie	315	79.0	2.1
ZT01	Item E	Historisch: Kwetsbaarheid	4132	36.0	1.6
ZT15	Item E	Historisch: Kwetsbaarheid	254	84.0	1.9
ZT20	Item E	Historisch: Kwetsbaarheid	338	85.0	1.6
ZT01	Item Q	Historisch: Problematisch drank- of drugsgebruik	2122	18.0	1.5
ZT11	Item Q	Historisch: Problematisch drank- of drugsgebruik	1637	37.0	1.6

**Score 3 of 4** Het is te zien dat de kosten van de groep die hoog scoort op een item vaak vergelijkbaar zijn met de groep die niet of laag scoort. Vooral in hoofdgroep X valt op dat er weinig hoge relatieve kosten zijn bij hoge scores.

Figuur 59



**Score 3 of 4 - tabel** Alleen in hoofdgroep Z komt het voor dat een item en zorgvraagtype uitkomen op relatieve kosten > 1,5 keer zo hoog voor de groep met hoge score (3 of 4) versus de groep met lage score én waarvan minstens 250 scorelijsten bestaan én die minstens 10% en maximaal 90% van het zorgvraagtype uitmaakt.

Tabel 15: Selectie van zorgvraagtypes en items met hoge relatieve kosten wanneer de score > 0

Zvt	Item	Beschrijving	Aantal	Aandeel (%)	Relatieve kosten
ZT18	Item 05	Lichamelijke problemen of handicaps	290	86.0	1.7
ZT21	Item 07	Problemen met depressieve stemmingen	278	88.0	1.9
ZT20	Item 12	Mogelijkheden voor het gebruiken en verbeteren van vaardigheden: beroepsmatig en vrije tijd	301	76.0	1.5

## 2.10 Subschalen

De items van de zorgvraagtyperingsscorelijst vallen in verschillende subsschalen. Er kan een relatie bestaan tussen de score op de subschalen en de kosten. Daarom bekijken we de subschalen. De subschalen sluiten goed aan bij de theorie van de HoNOS+ en daarom zijn de subschalen mogelijk geschikter dan de individuele items om zorgvraagtypes op te splitsen.

De score op de subschaal berekenen we door de scores op de individuele items binnen de subschaal op te tellen. Items 1-3 vallen in subschaal gedrag, items 4 en 5 zijn subschaal beperkingen, items 6-8 en 13 vallen in subschaal symptomen, items 9-12 zijn subschaal sociaal en de items A-E en Q vallen in subschaal historisch.

De analyses worden niet verder uitgediept door de voorkeur van de adviescommissie voor het gebruik van aanvullende kenmerken voor het al dan niet splitsen van de zorgvraagtypes.

### 2.10.1 Voorkomen subschaalscores bij zorgvraagtype

Welke subschalen scoren hoog bij een zorgvraagtype? En welke subschaalscores komen voor? Een hoge waarde staat gelijk aan grote problemen op de subschaal.

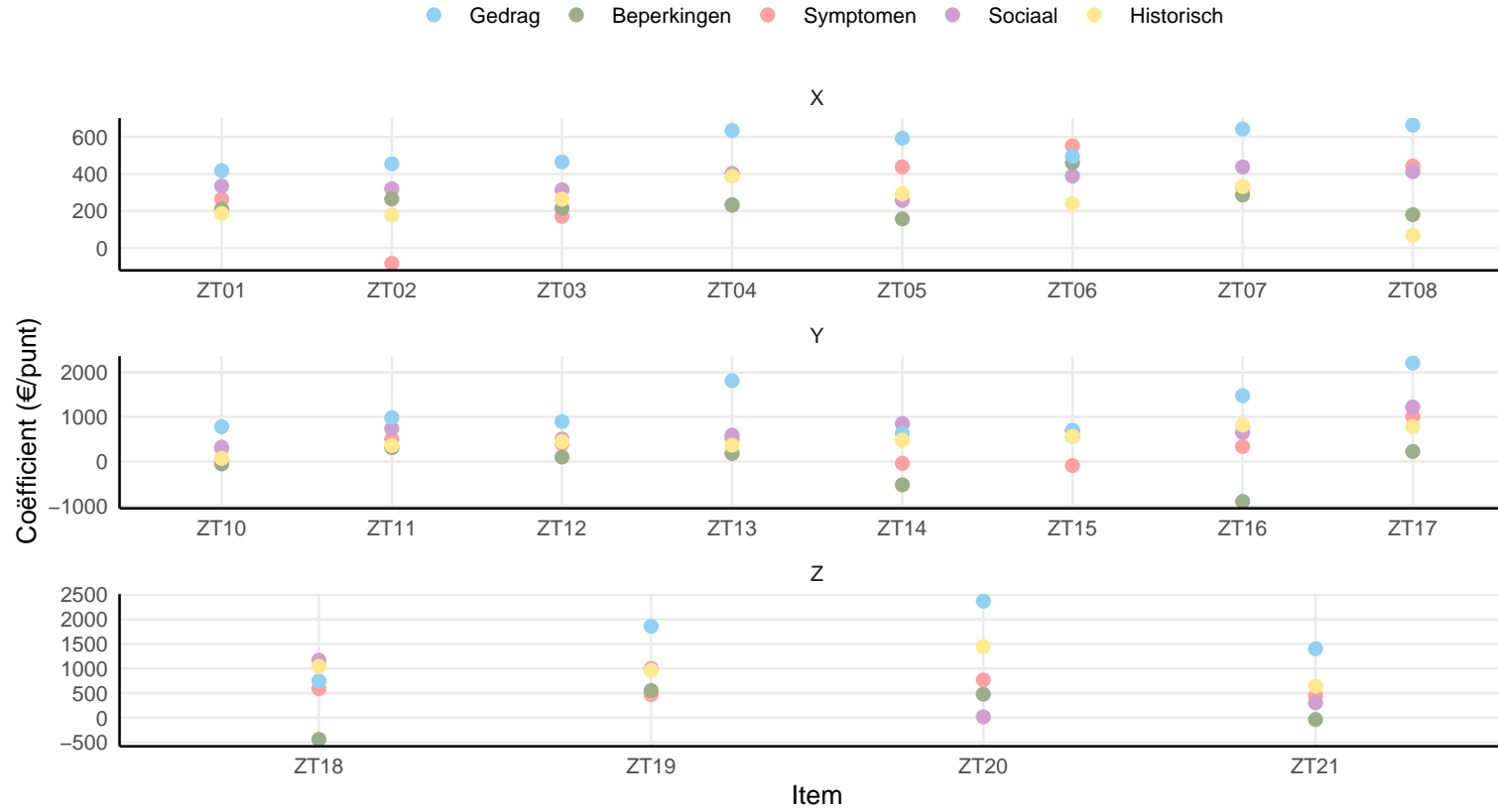
Figuur 60



### 2.10.2 Subschaal en kosten

We modelleren de kosten als functie van de scores op de individuele subschalen. Vooral de subschalen gedrag en sociaal vallen op met relatief hoge correlaties tussen kosten en score op de subschaal. Daarom bekijken we deze in meer detail.

Figuur 61





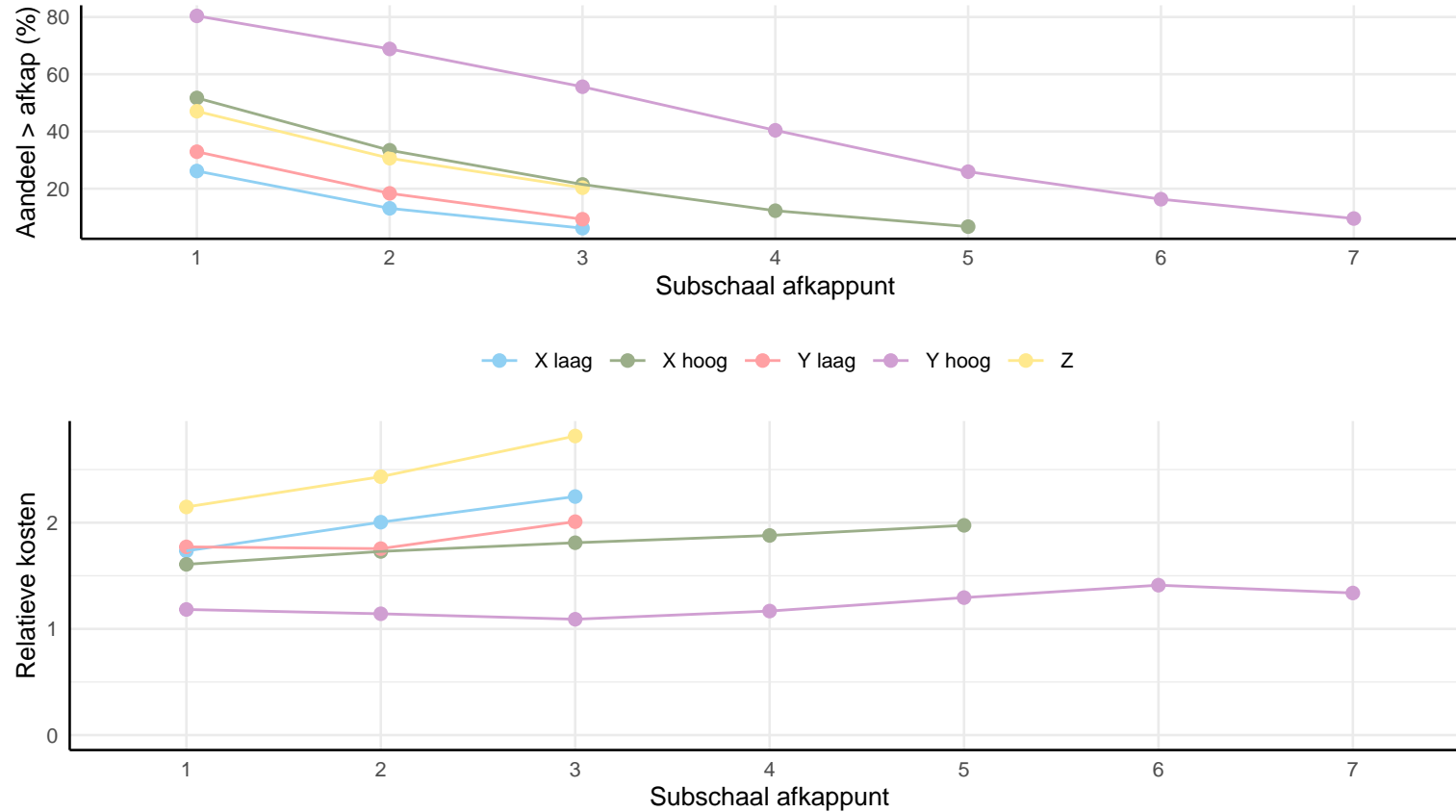
### 2.10.3 Splitsingspunt subschalen bepalen

Omdat de verwachting is dat de subschaal een vergelijkbare rol speelt bij vergelijkbare zorgvraagtypering en om de figuren transparant te houden brengen we de zorgvraagtypering terug tot enkele groepen. Specifiek groeperen we ZT01-ZT04 samen in 'X laag', ZT05-ZT08 in 'X hoog', ZT10-ZT13 in 'Y laag', ZT14-ZT17 in 'Y hoog' en ZT18-21 uiteraard als hoofdgroep Z.

De subschalen hebben een groot aantal waarden die ze kunnen aannemen, het is immers de som van de scores in de subschaal. Het is daarom niet direct duidelijk wat een logisch afkappunt zou zijn om een subschaal te gebruiken als potentieel punt om een zorgvraagtype te splitsen. Door te kijken naar het voorkomen en de relatieve kosten van patiënten met een subschaalscore boven of onder een afkappunt. We visualiseren alleen als de groep boven het afkappunt minimaal 250 observaties bevat en meer dan 5% en minder dan 95% van de zorgvraagtypering in de groep heeft.

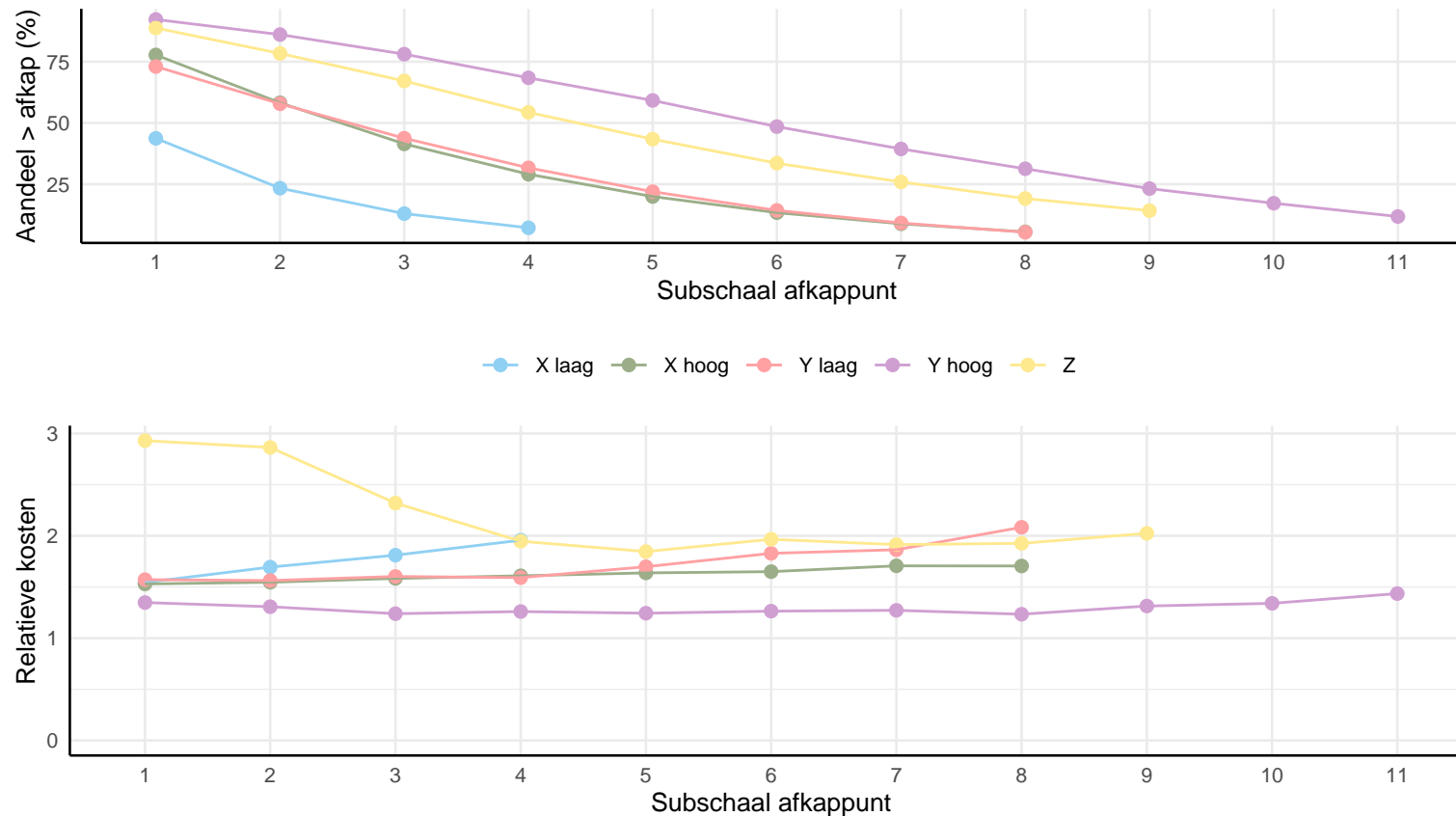
**Afkappunt gedrag** Het valt op dat het afkappunt niet veel lijkt uit te maken voor de relatieve kosten bij Y hoog, deze fluctueert tussen 1 en 1,5 keer kosten boven vergeleken met onder het afkappunt en ook voor Y laag lijkt het afkappunt in de subschaal gedrag weinig invloed te hebben op de relatieve kosten. Voor X laag, en Z nemen de relatieve kosten toe met het verschuiven van het afkappunt, al neemt het aandeel scorelijsten in de hoge groep snel af na de subschaal score van 3. Voor X hoog is de toename van de relatieve kosten en de afname van het aandeel scorelijsten boven het afkappunt wat geleidelijker.

Figuur 62



**Afkappunt sociaal** Net als bij de subschaal gedrag valt op dat het afkappunt bij subschaal sociaal niet veel lijkt uit te maken voor de relatieve kosten bij Y hoog, deze fluctueert tussen 1 en 1,5 keer kosten boven vergeleken met onder het afkappunt. Voor Y laag lijkt het afkappunt in de subschaal gedrag langzaam te stijgen met hogere scores als afkappunt. Voor X laag nemen de relatieve kosten toe met het verschuiven van het afkappunt, al neemt het aandeel scorelijsten in de hoge groep snel af na de subschaal score van 3. Voor X hoog is de toename van de relatieve kosten en de afname van het aandeel scorelijsten boven het afkappunt wat geleidelijker. Interessant genoeg nemen de relatieve kosten in hoofdgroep Z af met de toename van de subschaal.

Figuur 63



### 3 Declaratiegegevens

De declaratiegegevens volgen uit het declaratieverkeer tussen zorgaanbieder en zorgverzekeraar. Dit zijn de gegevens die vermeld staan op de factuur. De facturen worden door de individuele verzekeraars gecontroleerd en uitbetaald. Vervolgens worden de gegevens over de facturatie en betaling doorgeleverd aan Vektis. Vektis levert uiteindelijk gegevens aan bij de NZa. De declaratiegegevens bevatten veel informatie over de behandelinzet en bevatten over de zorgvraag van de patiënt informatie over het zorgvraagtype, de dsm-hoofdgroep en/of het basis-ggz-profiel.

NB Het is bekend dat er enkele problemen zijn met de declaratiegegevens. Deze problemen zijn bekend en worden opgelost. Specifiek gaat het om problemen het koppelen van diagnose- of zorgvraagtyperingsgegevens aan de gegevens over de geleverde zorg en om dubbelingen in de gegevens door problemen in een controlesysteem van Vektis. Het betreft een kleine minderheid van de gegevens, naar schatting 2-3% van de zorgprestaties. De fouten lijken niet bij specifieke groepen patiënten of zorgaanbieders voor te komen, dus het is niet aannemelijk dat de fouten een grote invloed zullen hebben op de resultaten van de analyses hieronder. Voor de volledigheid zal de NZa dit document nogmaals maken wanneer de fouten zijn hersteld. Volgens schatting van Vektis is dit eind september 2024. De NZa zal monitoren of er grote veranderingen in de resultaten ontstaan met de nieuwe verbeterde gegevens.

## 3.1 Selectie declaratiegegevens en labelling

### 3.1.1 Selectie gegevens

Om goede conclusies te kunnen trekken over overeenkomsten en verschillen tussen zorgvraagtypes is het belangrijk vergelijkbare gegevens te gebruiken. Patiënten met bepaalde zorgvraagtypes zijn veelal na enkele maanden klaar met de behandeling, terwijl voor andere zorgvraagtypes de behandeling een jaar of langer kan duren. Het is met de declaratiegegevens nog niet mogelijk trajecten over vele jaren te volgen, omdat het Zorgprestatie-model en de daarbij horende registratie van de zorgvraagtypering pas sinds januari 2022 een feit is. We selecteren daarom zorgtrajecten waarvoor we minimaal een jaar aan gegevens hebben. Dit operationaliseren we door zorgtrajecten te nemen waarvoor er declaraties zijn die minstens 365 na de startdatum van het zorgtraject vallen. Sommige trajecten zijn eerder afgelopen dan 365 dagen na de start, maar dit is niet op te maken uit de gegevens. Als er geen declaraties bestaan na 365 dagen kan dit komen omdat het traject is afgelopen, of omdat de zorgaanbieder deze zorg nog niet heeft gedeclareerd. Daarom bekijken we voor elk traject ook wat de follow-up is bij de aanbieder. Dit is het laatste moment waarop de zorgaanbieder nog enige zorg gedeclareerd heeft bij de verzekeraar van de patiënt. Als de follow-up minstens 365 dagen is nemen we het zorgtraject mee. Van alle zorgtrajecten nemen we alleen de zorg mee die is geleverd in de eerste 365 dagen van het zorgtraject, of de eerste 365 dagen na invoering van het Zorgprestatie-model.

De kosten van de zorgtrajecten bepalen we op basis van de NZa-tarieven, niet op het uiteindelijk uitbetaalde bedrag. Voor de vergelijkbaarheid nemen we de tarieven uit één jaar, in dit geval 2022. Voor prestaties die nog niet bestonden in 2022 of prestaties met een vrij tarief nemen we het gemiddelde van de gedeclareerde bedragen voor de prestatie.

### 3.1.2 Labelling trajecten

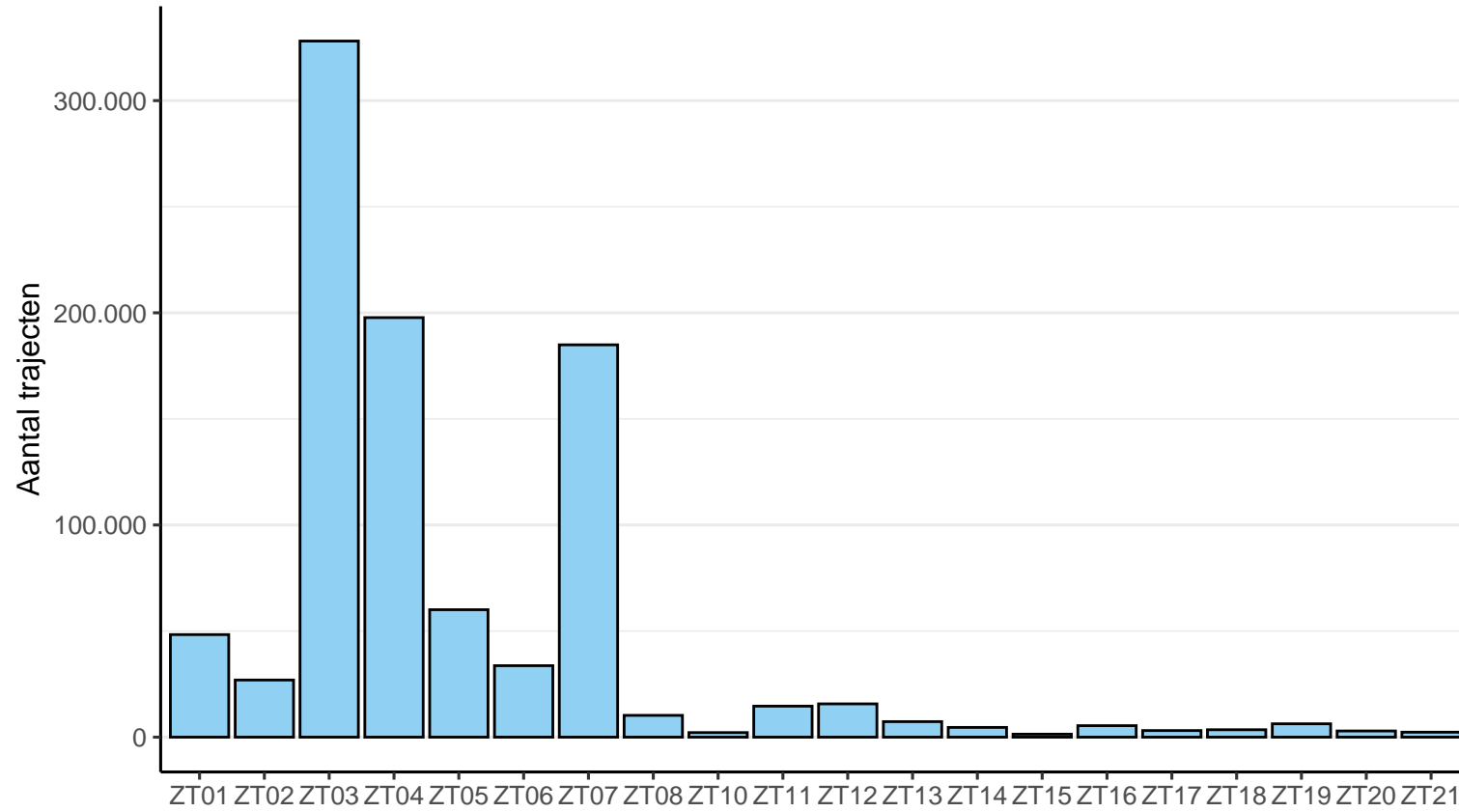
Voor de verschillende zorgtrajecten kan het zo zijn dat er meerdere zorgvraagtypering, diagnosehoofdgroepen of basis-ggz-profielen voor één traject zijn geregistreerd. Dit kan komen doordat er evaluatie heeft plaatsgevonden en na deze evaluatie een ander zorgvraagtype of hoofdgroep is gekozen, of omdat er registratiefouten zijn opgetreden bij de zorgaanbieder. Het kan ook bestaan dat één zorgtraject bij meerdere zorgaanbieders loopt, denk aan onderaannemerschap, of bij meerdere verzekeraars in het geval van overstappen van verzekeraar of meerdere settingen bij op- of afschalen van de zorg. In alle bovenstaande gevallen is voor het zorgtraject maximaal één zorgvraagtype, hoofdgroep, gb-ggz-profiel, zorgaanbieder of zorgverzekeraar bepaald. Dit is, na advies van de werkgroep data van de adviescommissie, gedaan door voor elk van de voorkomende varianten de kosten te nemen en de variant met de hoogste kosten als bepalende waarde te nemen. Het kan dus zo zijn dat een zorgtraject 11 van de 12 maanden het label ZT11 had (stabiele psychotische stoornis), maar tijdens één maand in crisis is beland en veel kosten heeft gemaakt onder ZT14 (psychotische episode), dat deze als ZT14 in de dataset aanwezig is.

In overeenstemming met de werkgroep data is ook een definitie van een afgerond zorgtraject vastgesteld. Hiervan is sprake als in een zorgtraject drie maanden (90 dagen) geen declaraties hebben plaatsgevonden, terwijl er wel minimaal 90 dagen follow-up is na de laatste declaratie van het zorgtraject.

### 3.2 Verkennen declaraties

#### 3.2.1 Aantal trajecten per zorgvraagtype

Figuur 64: Aantal trajecten per zorgvraagtype

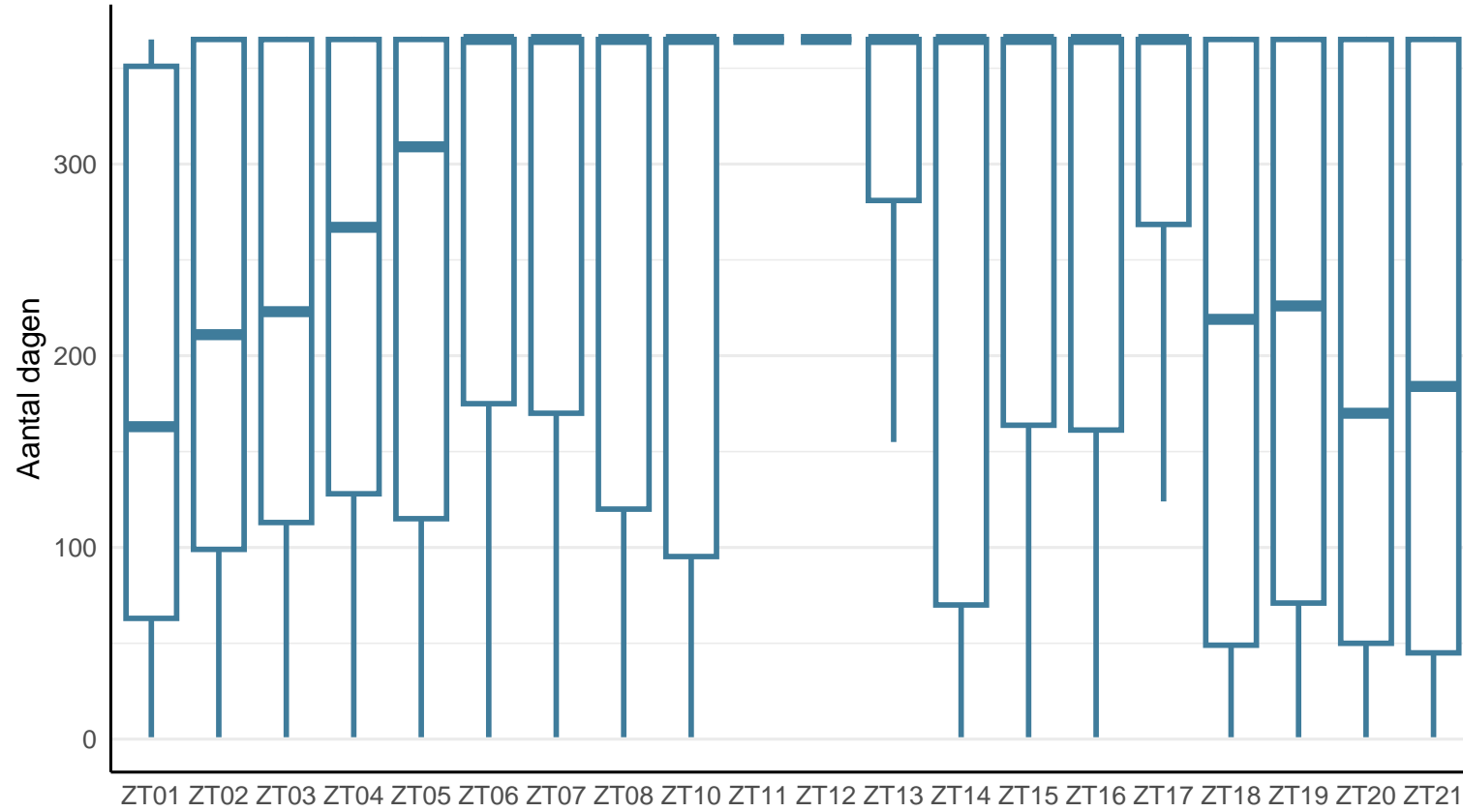


### 3.2.2 Looptijd per zorgvraagtype

Niet alle zorgtrajecten duren even lang. We vermoeden dat er verschil is in de verdeling van de duur van trajecten tussen de verschillende zorgvraagtypering. Dit kunnen we inzichtelijk maken door te kijken naar de tijd tussen de startdatum van het zorgtraject en de laatste bekende prestatie. Dit noemen we de looptijd van het traject. We hebben alleen de eerste 365 dagen van het traject geselecteerd, dus wanneer de laatste prestatie meer dan 365 dagen na de startdatum ligt stellen we de looptijd in op 365 dagen.

Zorgtrajecten voor ZT01-ZT05 zijn in meer dan de helft van de trajecten afgelopen binnen een jaar, voor ZT01 is dit zelfs meer dan 75%. ZT06-ZT17 lopen in meer dan 50% van de gevallen een jaar of langer. Voor ZT11 en ZT12 (respectievelijk stabiele psychotische stoornis en chronische en beperkende psychotische stoornis) is dit zelfs meer dan 75% van de trajecten. Voor hoofdgroep Z geldt dat meer dan de helft van de trajecten binnen een jaar is afgerond.

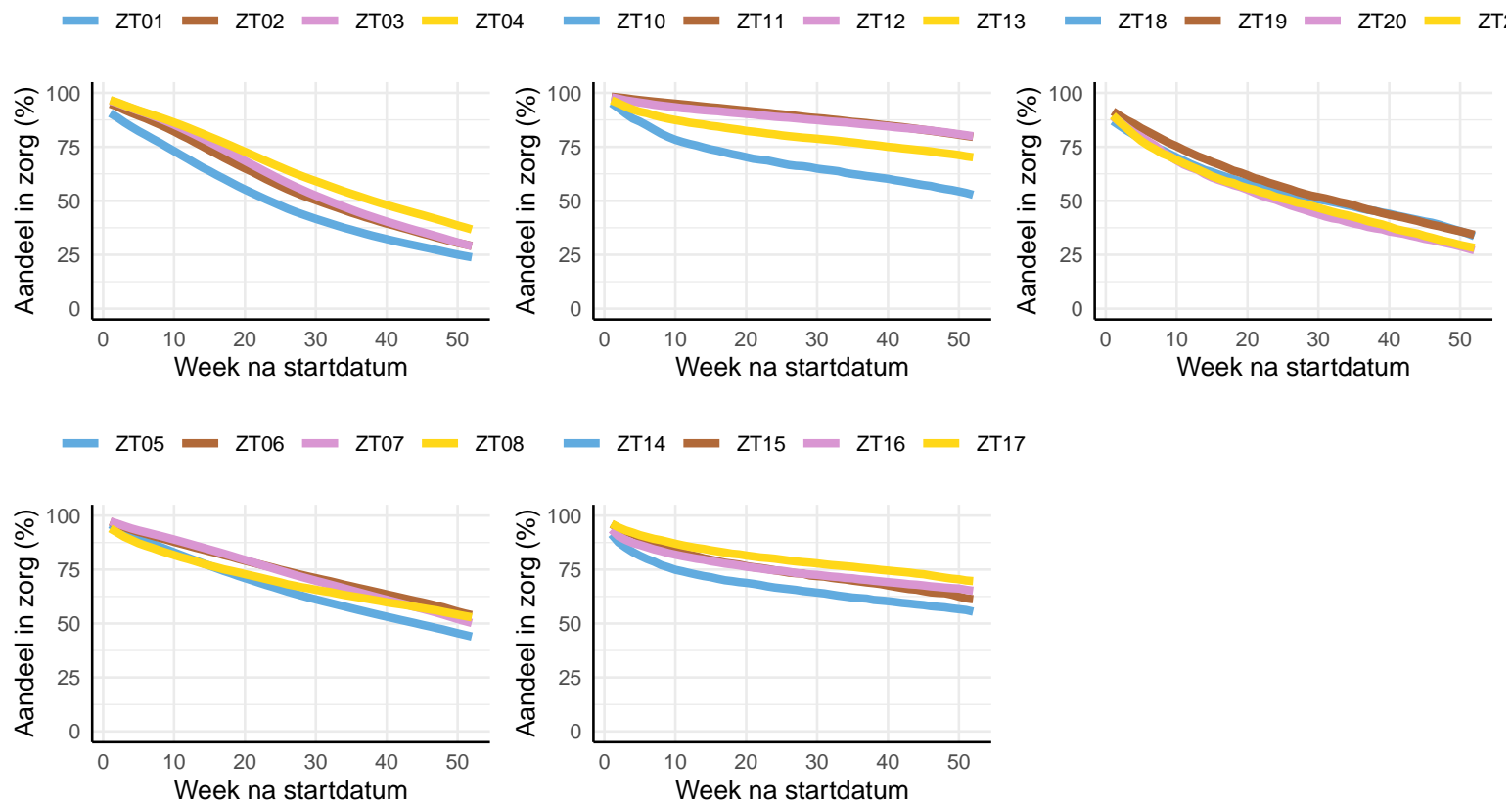
Figuur 65: Looptijd per zorgvraagtype





**Uitstroom over tijd** In onderstaande figuur is te zien welk percentage van de patiënten in een zorgvraagtype nog in zorg is (y-as) na verloop van tijd (x-as).

Figuur 66: Uitstroom per zorgvraagtype

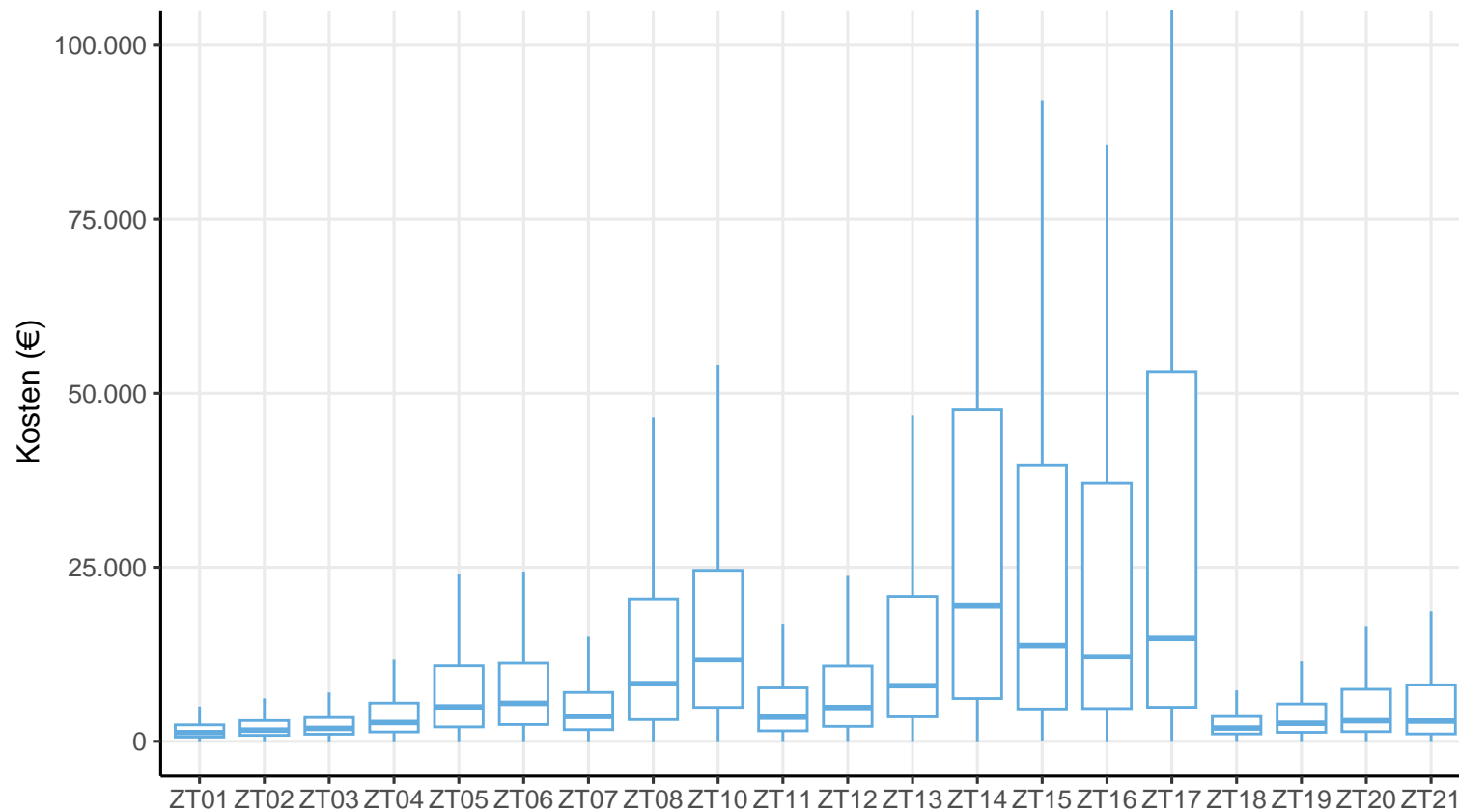


### 3.2.3 Kosten per zorgvraagtype

In elk zorgtraject worden kosten gemaakt. Hoe deze kosten samenhangen met het zorgvraagtype maken we hieronder inzichtelijk. We nemen hiervoor de kosten die in het eerste jaar na het begin van het zorgtraject zijn gemaakt, of na overgang van het zorgtraject naar het ZPM in 2022.

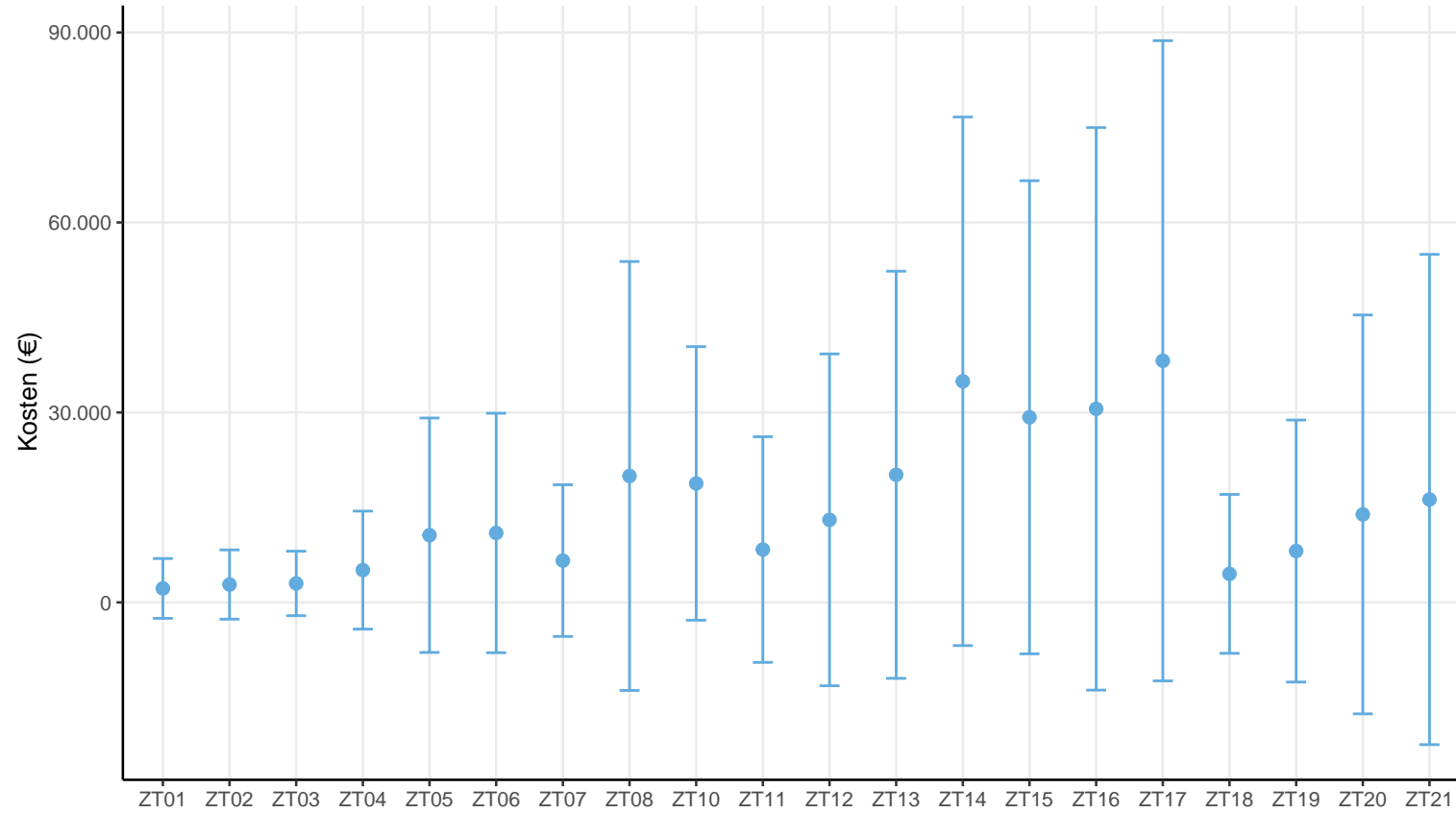
**Verdeling kosten - boxplot** Er lijkt verschil te zijn tussen de verdeling van kosten tussen een aantal zorgvraagtypes, al is de spreiding binnen en de overlap tussen de zorgvraagtypes groot. Uit de verdeling is ook op te maken dat kosten binnen een zorgvraagtype in veel gevallen niet normaal verdeeld is, de box (50-75% van de trajecten) boven de mediaan is in veel gevallen een stuk groter dan de box onder de mediaan (25-50%).

Figuur 67: Verdeling van kosten per zorgvraagtype



**Gemiddelde kosten per zvt** Er zijn grote verschillen in de gemiddelde kosten tussen sommige zorgvraagtypes, maar door de grote spreiding bij elk zorgvraagtype is er ook veel overlap in de mogelijke kosten van de verschillende zorgvraagtypes.

Figuur 68: Gemiddelde en spreiding van kosten per zorgvraagtype



Tabel 16: Gemiddelde kosten en spreiding per zorgvraagtype

ZVT	Gemiddelde kosten (€)	Standaardafwijking	Aantal
ZT01	2.203	4.722	48.341
ZT02	2.819	5.471	26.919
ZT03	2.998	5.094	328.063
ZT04	5.100	9.317	197.725
ZT05	10.609	18.507	60.069
ZT06	10.964	18.907	33.685
ZT07	6.601	11.976	184.897
ZT08	19.963	33.871	10.267
ZT10	18.787	21.601	2.146
ZT11	8.343	17.817	14.610
ZT12	13.037	26.201	15.681
ZT13	20.154	32.150	7.322
ZT14	34.914	41.734	4.590
ZT15	29.237	37.357	1.404
ZT16	30.568	44.419	5.410
ZT17	38.156	50.550	3.111
ZT18	4.506	12.548	3.495
ZT19	8.113	20.689	6.305
ZT20	13.896	31.498	2.908
ZT21	16.252	38.713	2.348

**Tabel gemiddelde kosten en spreiding**

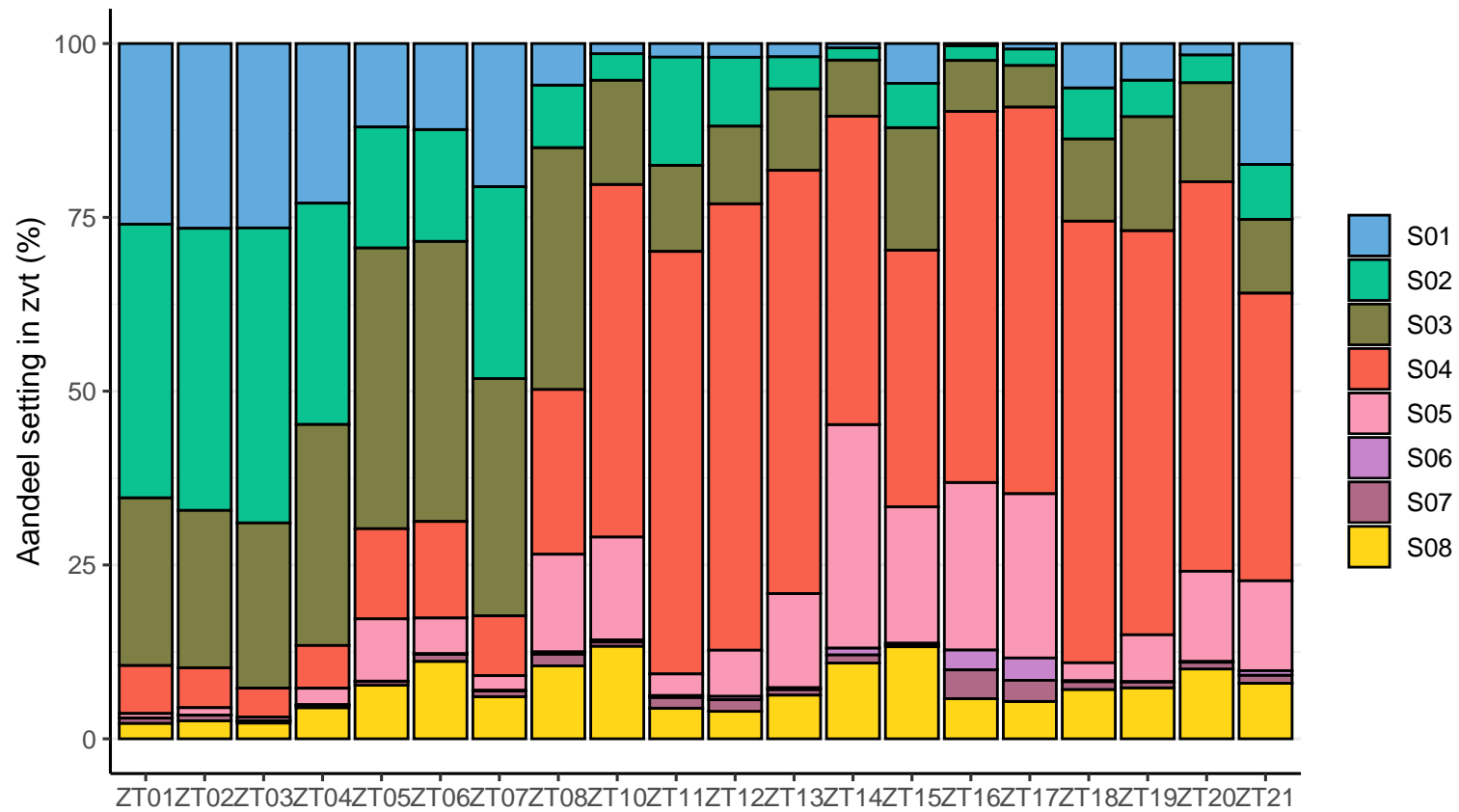
### 3.3 Verdeling over settings

Het is de verwachting dat er een relatie is tussen het zorgvraagtype en de setting waarin een patiënt behandeld wordt. Hieronder bekijken we dat op twee manieren. Ten eerste bekijken we of zorgvraagtypes meestal in één of enkele settings terecht komt. Dit kan antwoord geven op de vraag of er een 'logische' setting is voor een zorgvraagtype. Ten tweede bekijken we welke zorgvraagtypering veel voorkomen binnen een setting. Dit kan laten zien of een setting zich lijkt te richten op een specifieke patiëntenpopulatie. Let op dat dit tweede vertekend kan worden doordat niet alle zorgvraagtypes in even grote aantallen bestaan.

### 3.3.1 Settings bij zorgvraagtype

In welke settings komen de zorgvraagtypes voor? Zorgvraagtypes ZT01-ZT04 komen voornamelijk voor in de ambulante settings (S01-S03). Zorgtrajecten met zorgvraagtypes in hoofdgroepen Y en Z hebben relatief vaak outreachende zorg (S04). De psychotische episodes en andere complexere zorgvraagtyperingen hebben een relatief groot aandeel van de trajecten in de klinische (S05) en hoogspecialistische (S08) setting. De beveiligde settings (S06 en S07) komen bij geen van de zorgvraagtyperingen in meer dan 4% van de trajecten voor en zijn voor de meeste zorgvraagtypes verwaarloosbaar.

Figuur 69: In welke settings komen de zorgvraagtypes voor?

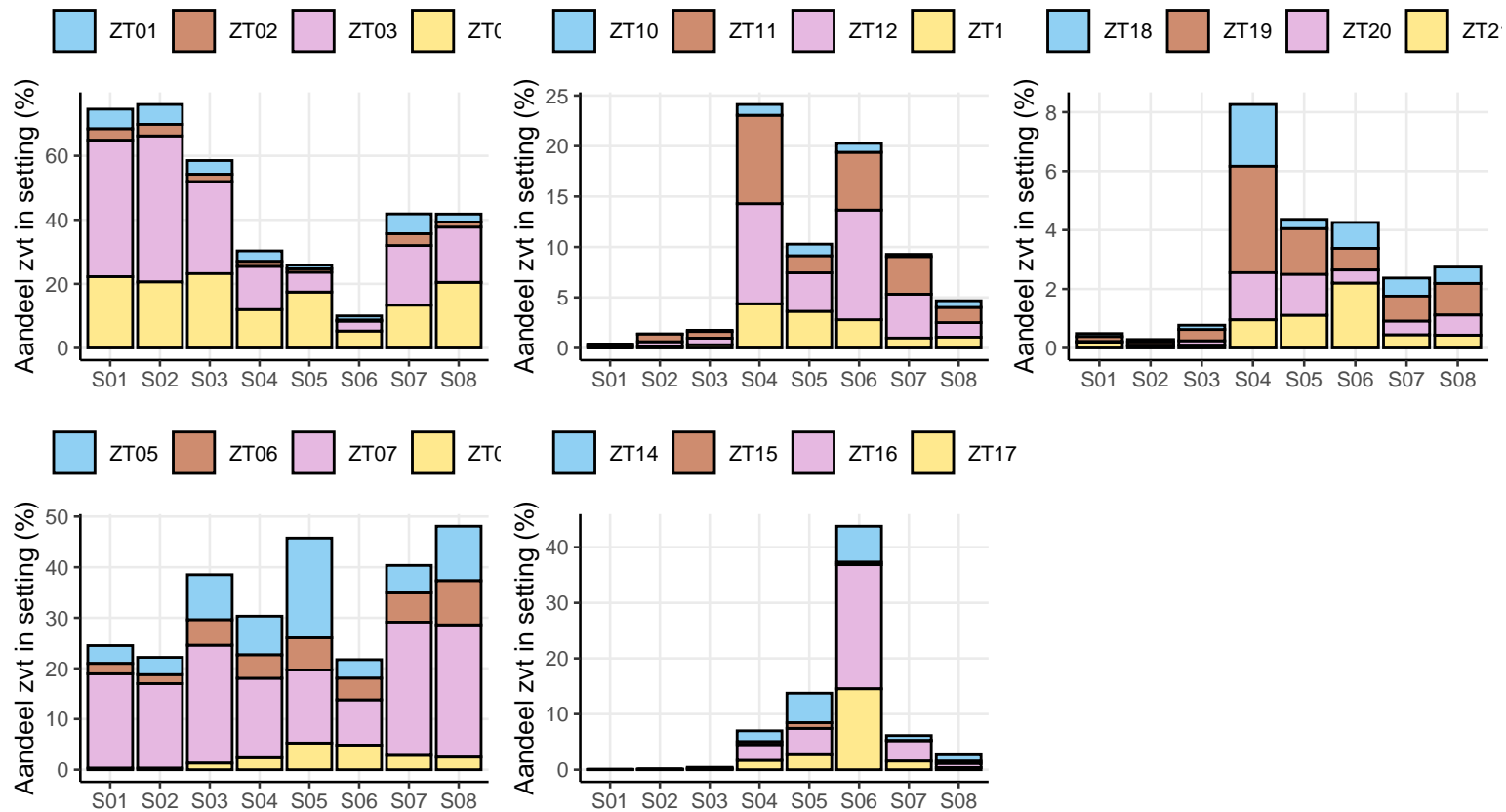


### 3.3.2 Zorgvraagtype in setting

Welke zorgvraagtypes komen veel voor in een setting? NB de y-as verschilt tussen de panelen van het figuur.

Het valt op dat zo'n 75% van S01 en S02 gevuld is met ZT01-ZT04 en 60% van S03, terwijl dat ongeveer 25% is voor S04 en S05. Voor alle settings geldt dat 20-50% van de trajecten in de setting bestaat uit ZT05-ZT08. S04 en S06 zijn voor 20-25% gevuld met ZT10-ZT13 en S06 is voor bijna de helft gevuld met ZT14-ZT17. Voor alle settings is hoofdgroep Z een kleine minderheid, met maximaal 8 % van de trajecten (S04) bestaande uit deze hoofdgroep.

Figuur 70: Zorgvraagtypes per setting



### 3.3.3 Kosten zvt en setting

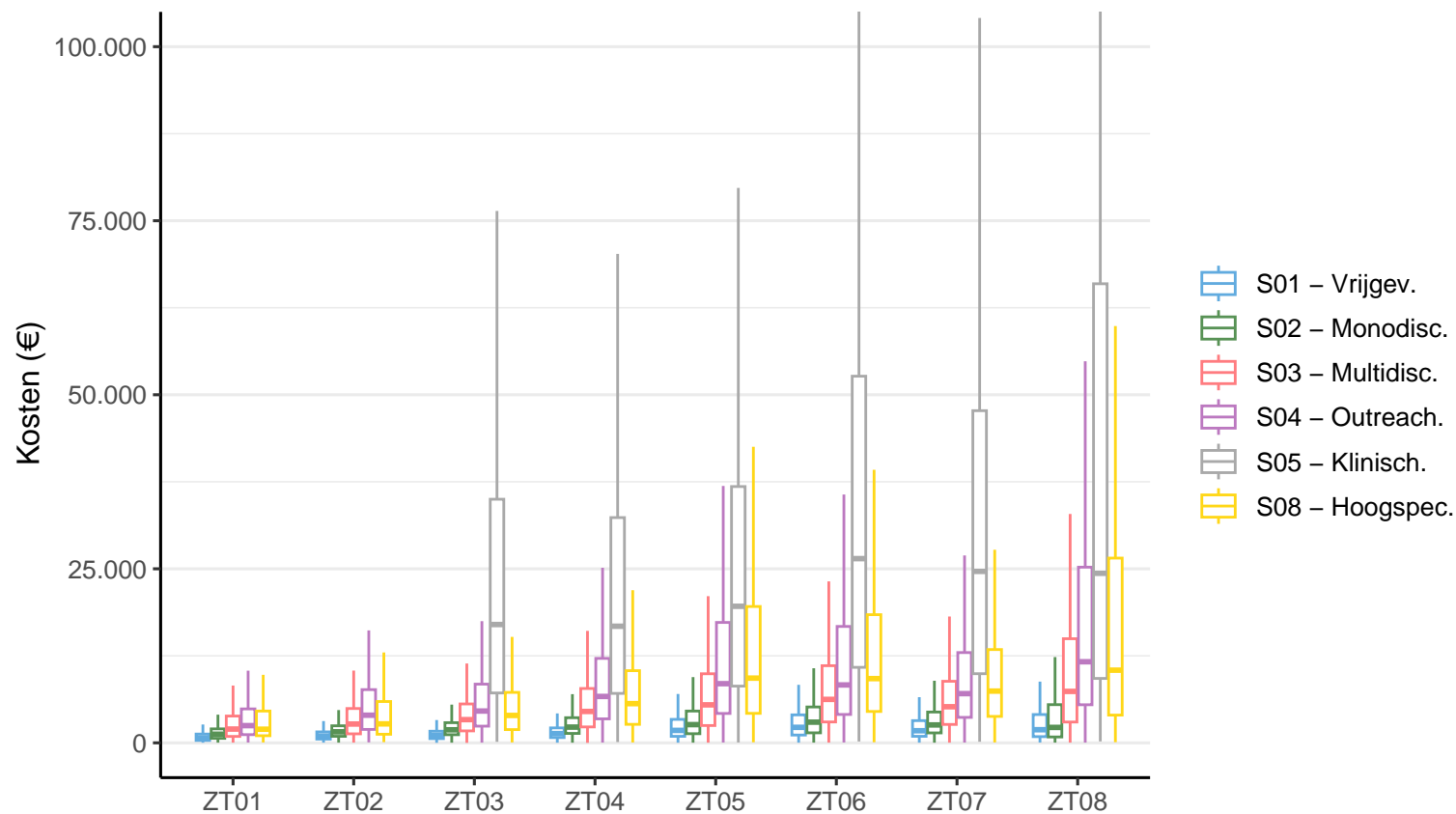
Het is waarschijnlijk dat de kosten van de behandeling van een patiënt afhankelijk is van meer dan alleen het zorgvraagtype. Daarom bekijken we hoe de verdeling van kosten per zorgvraagtypering afhangt van de belangrijkste setting waarin de behandeling plaats heeft gevonden.

We sluiten de settings uit waarin slechts weinig patiënten met een bepaald zorgvraagtype worden behandeld. Dit is de reden dat de beveiligde settings (S06 en S07) in geen van de figuren wordt weergegeven en in sommige doorsnedes ook sommige van de andere settings niet worden gevisualiseerd.



**Kosten en setting - X** In hoofdgroep X lopen de kosten stapsgewijs op van settingen S01-S05. Bij alle zorgvraagtypes met voldoende S05-trajecten zijn de kosten het hoogst voor in S05. Dit is vrijwel zeker het gevolg van de kosten aan verblijfsdagen. De verdeling van kosten in S08 lijkt weer iets lager dan in S05 en lijkt vergelijkbaar met de verdeling in de outreachende setting (S04).

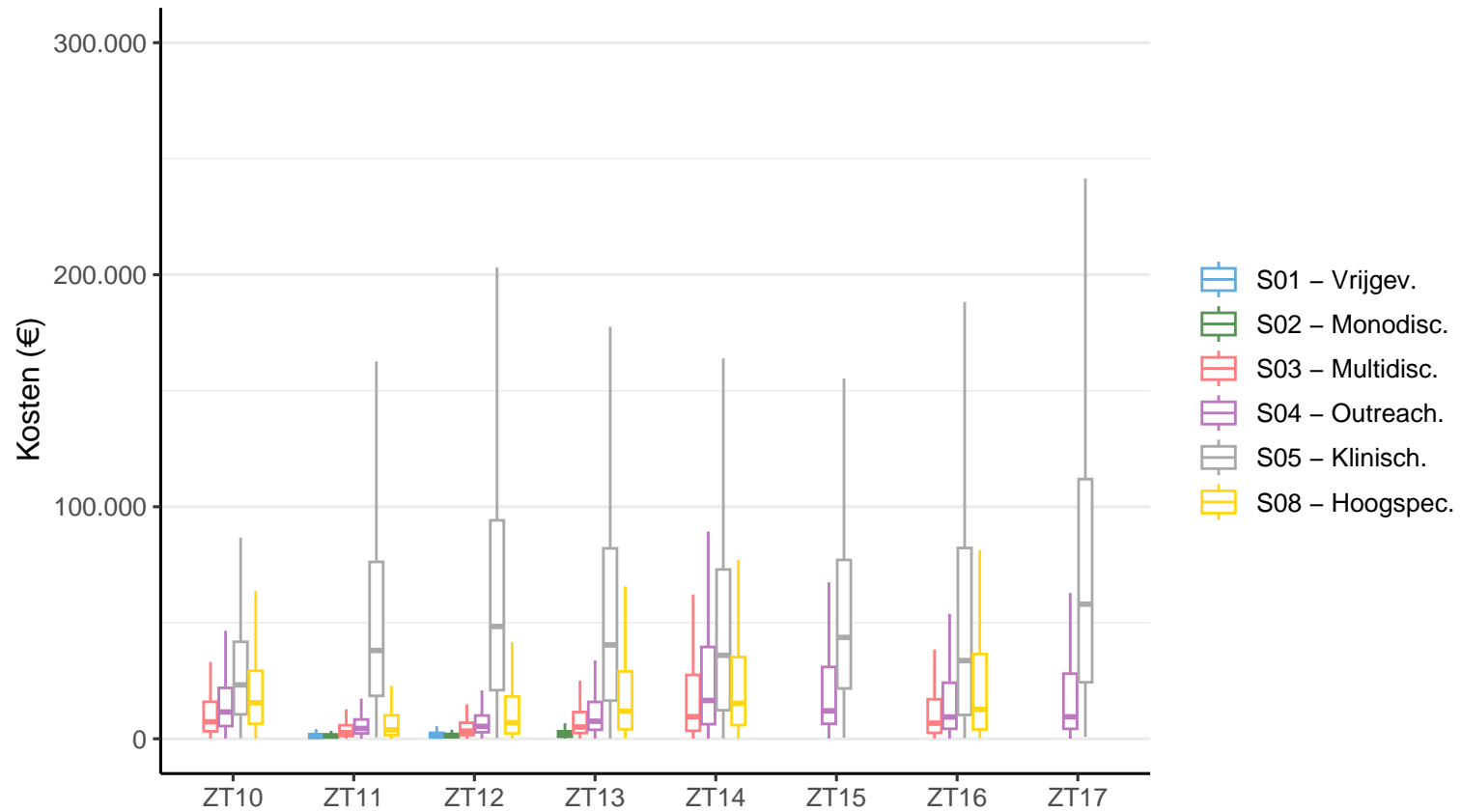
Figuur 71: Verdeling kosten tussen settingen in hoofdgroep X



**Kosten en setting - Y** In hoofdgroep Y zijn de kosten van de trajecten in klinische setting (S05) hoger dan de kosten voor hetzelfde zorgvraagtype in een andere setting.

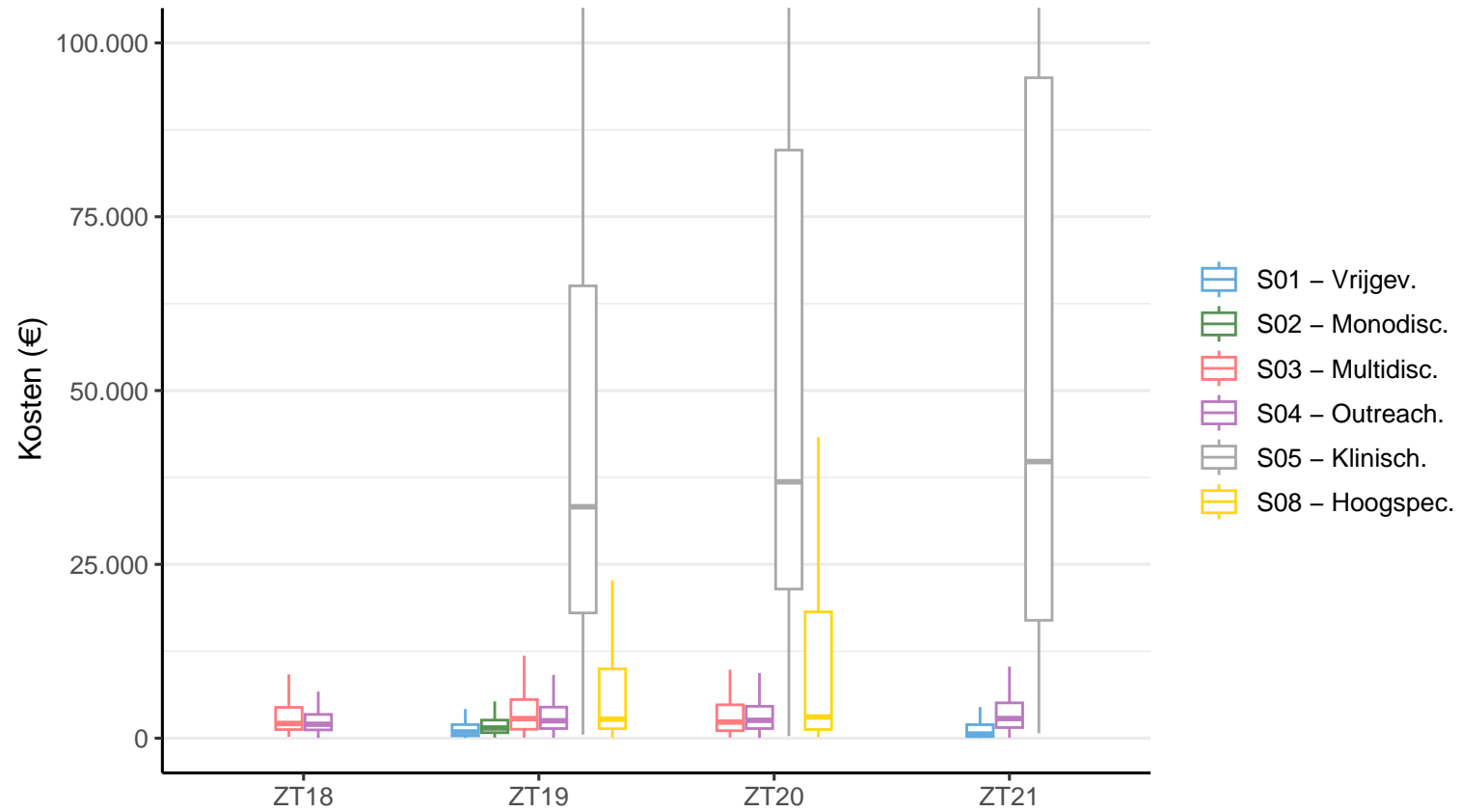
In hoofdgroep Y komt behandeling in de vrije vestiging logischerwijs niet veel voor, behalve voor de stabiele zorgvraagtypes. Voor de overige zorgvaragtypes worden dan ook geen kosten gevisualiseerd.

Figuur 72: Verdeling kosten tussen settings in hoofdgroep Y



**Kosten en setting - Z** In hoofdgroep Z springen vooral ZT19, ZT20 en ZT21 in setting S05 er bovenuit.

Figuur 73: Verdeling kosten tussen settingen in hoofdgroep Z



### 3.4 Zorgvraagtype en diagnosehoofdgroep

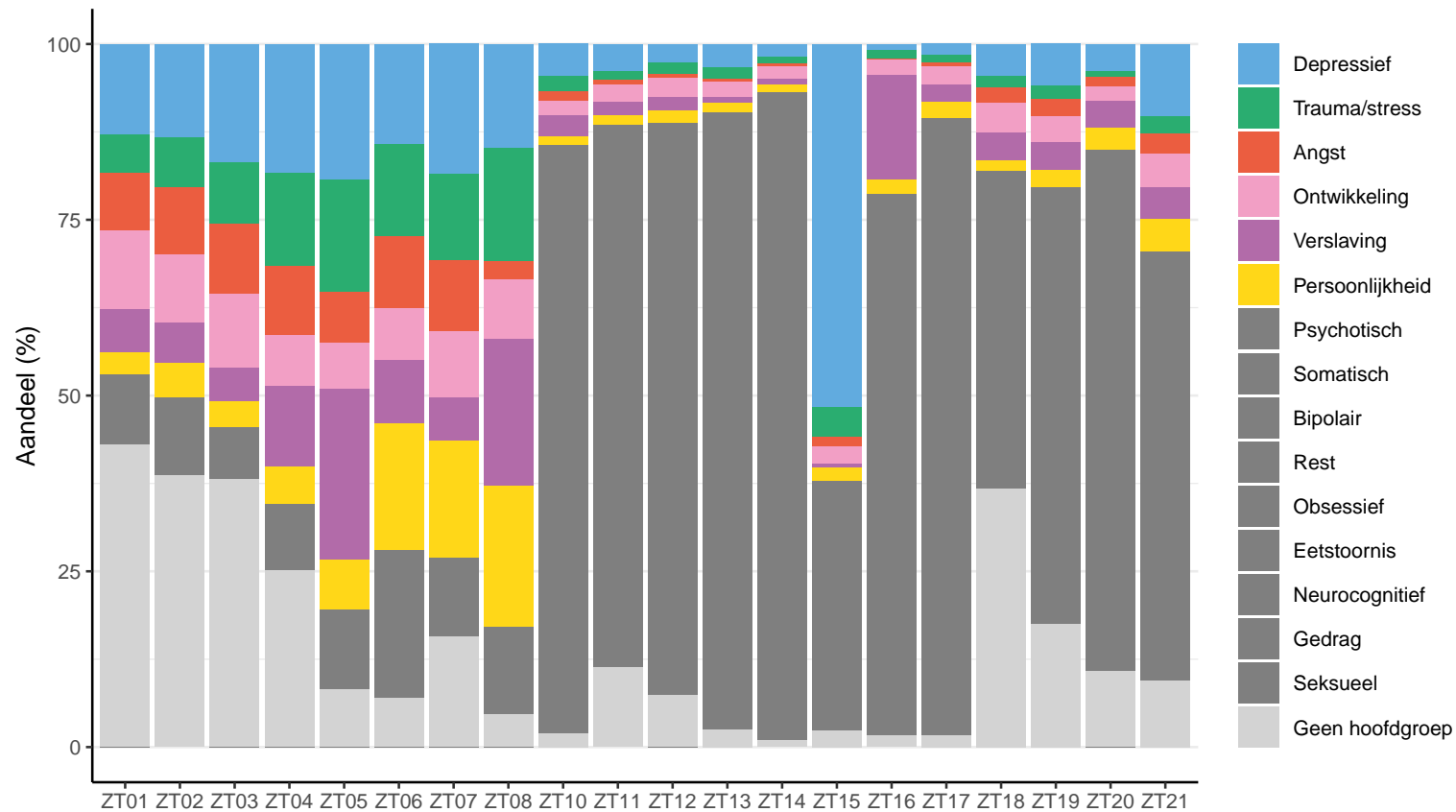
Het is te verwachten dat er een correlatie is van diagnosehoofdgroep (voor overzichtelijkheid DSM) met zorgvraagtyperingshoofdgroep (voor overzichtelijkheid hoofdgroep). Tegelijkertijd is het ook te verwachten dat één DSM in meerdere zorgvraagtypes voor kan komen, afhankelijk van de ernst en de complexiteit van de zorgvraag. We maken dat hier inzichtelijk. Voor het overzicht splitsen we dit in twee figuren, met enerzijds de zes meest voorkomende DSM (op basis van het aantal trajecten in de huidige dataset) en anderzijds de overige DSM.

NB er is niet altijd een DSM aanwezig op de factuur. In de meeste gevallen is dit omdat er een basis-ggz-profiel is geregistreerd.

### 3.4.1 Meest voorkomende diagnosehoofdgroepen

We visualiseren de DSM's in twee afzonderlijke figuren. In de eerste selecteren we de zes DSM's die het meest voorkomen in de dataset. Deze lijken grotendeels voor te komen in hoofdgroep X. De verdeling binnen deze hoofdgroep verschilt op enkele gebieden van elkaar. Specifiek komen verslavingsstoornissen meer voor in ZT05 en ZT08, terwijl persoonlijkheidsstoornissen sterker vertegenwoordigd zijn in ZT06-ZT08. Opvallend is het grote aandeel depressieve stoornissen in ZT15, al was een groter aandeel ook mogelijk geweest, gezien de naam van het zorgvraagtype: depressie met psychotische kenmerken. Bij de andere zorgvraagtypes komen de meest voorkomende stoornissen niet vaak voor, al valt het aandeel verslavingsproblemen in ZT16 op. Ook dit is logisch te verklaren op basis van de naam: psychotische stoornis - chronisch, instabiel met ernstig middelenmisbruik.

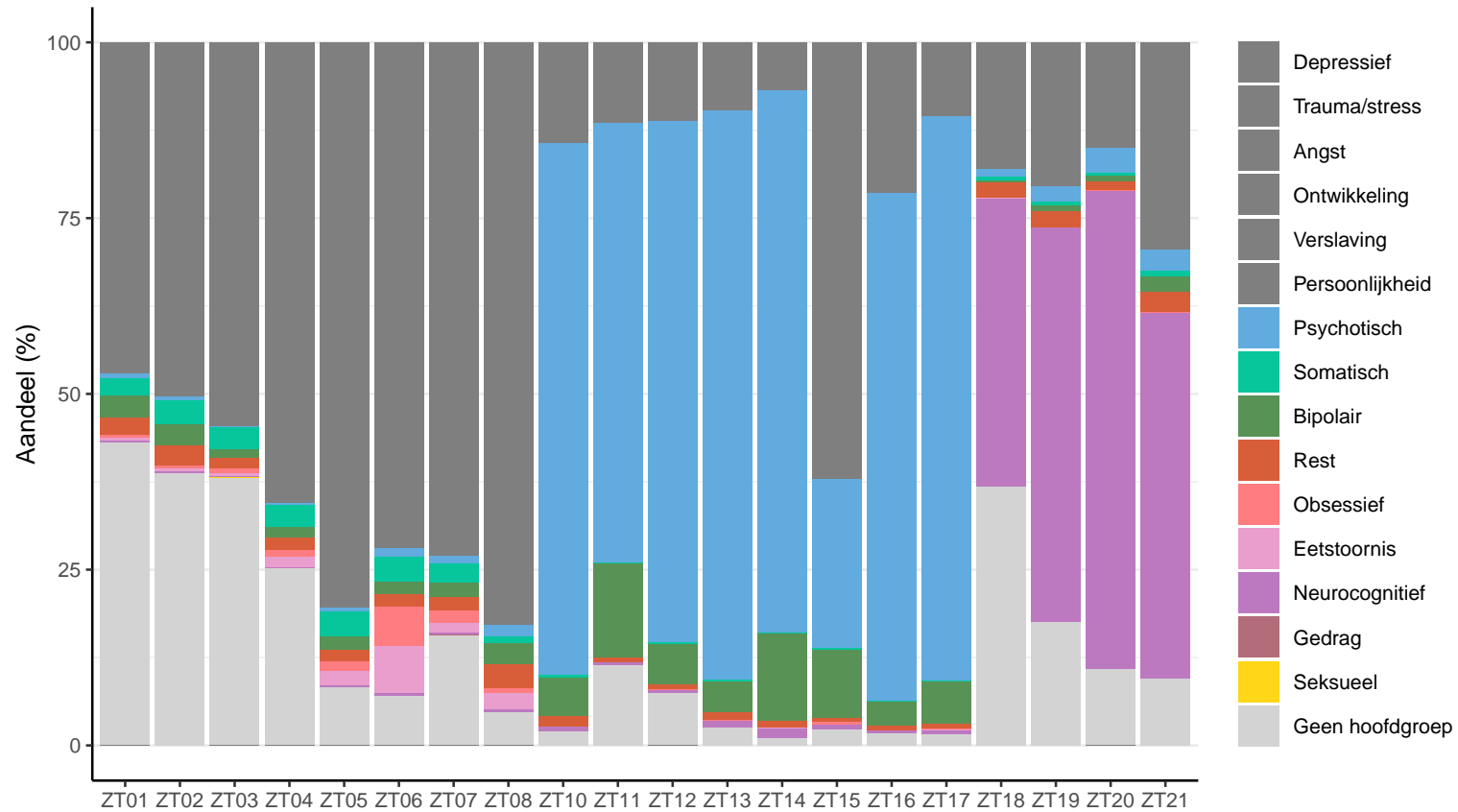
Figuur 74: Verdeling diagnosehoofdgroepen per zorgvraagtype - top 6



### 3.4.2 Overige diagnosehoofdgroepen

Voor de andere DSM dan die het meest voorkomen is te zien dat deze niet veel voorkomen in hoofdgroep X. In ZT06 valt alleen op dat een naar verhouding groot aandeel van de patiënten een obsessieve stoornis of een eetstoornis heeft. In hoofdgroep Y is logischerwijs te zien dat een groot aandeel van de patiënten een psychotische stoornis heeft. Ook bipolaire stoornissen komen vaker voor dan in de overige hoofdgroepen. In hoofdgroep Z valt logischerwijs het grote aandeel neurocognitieve stoornissen op.

Figuur 75: Verdeling diagnosehoofdgroepen per zorgvraagtype - overig

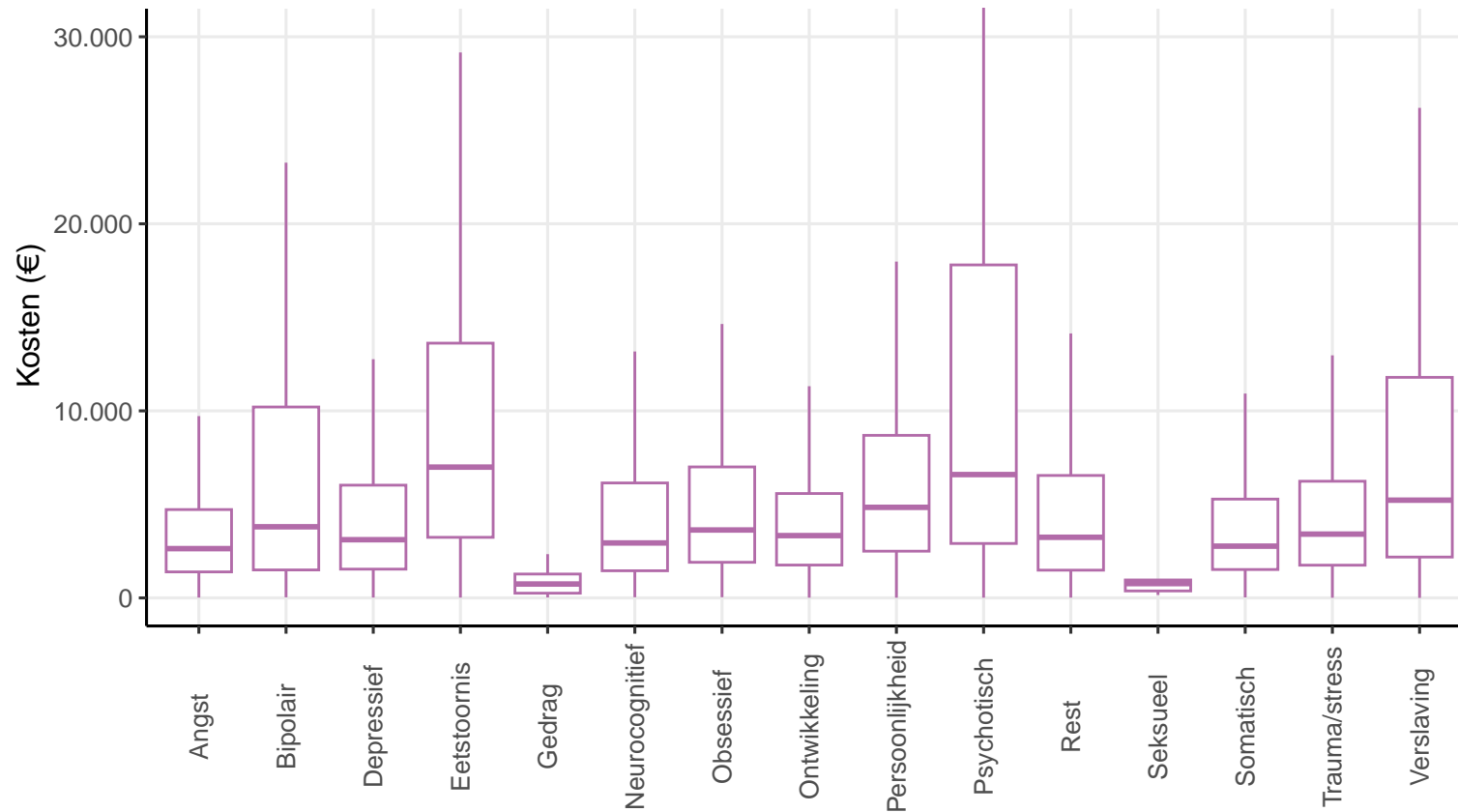


### 3.4.3 Kosten per diagnosehoofdgroep

Om de vergelijking te maken met de sturingsinformatie die historisch gezien onderdeel was van de declaraties in de specialistische ggz bekijken we de verdeling van kosten per diagnosehoofdgroep. Bij dit figuur zijn de kosten op dezelfde manier berekend als in de figuren met kosten per zorgvraagtype op basis van declaratiegegevens hierboven. Het zijn dus de kosten uit het eerste jaar na het eerste zorgcontact na invoering van het zorgprestatie model per 01 januari 2022.

NB de schaal van de y-as is in dit figuur sterk verschillend van het vergelijkbare figuur 67 met gegevens per zorgvraagtype.

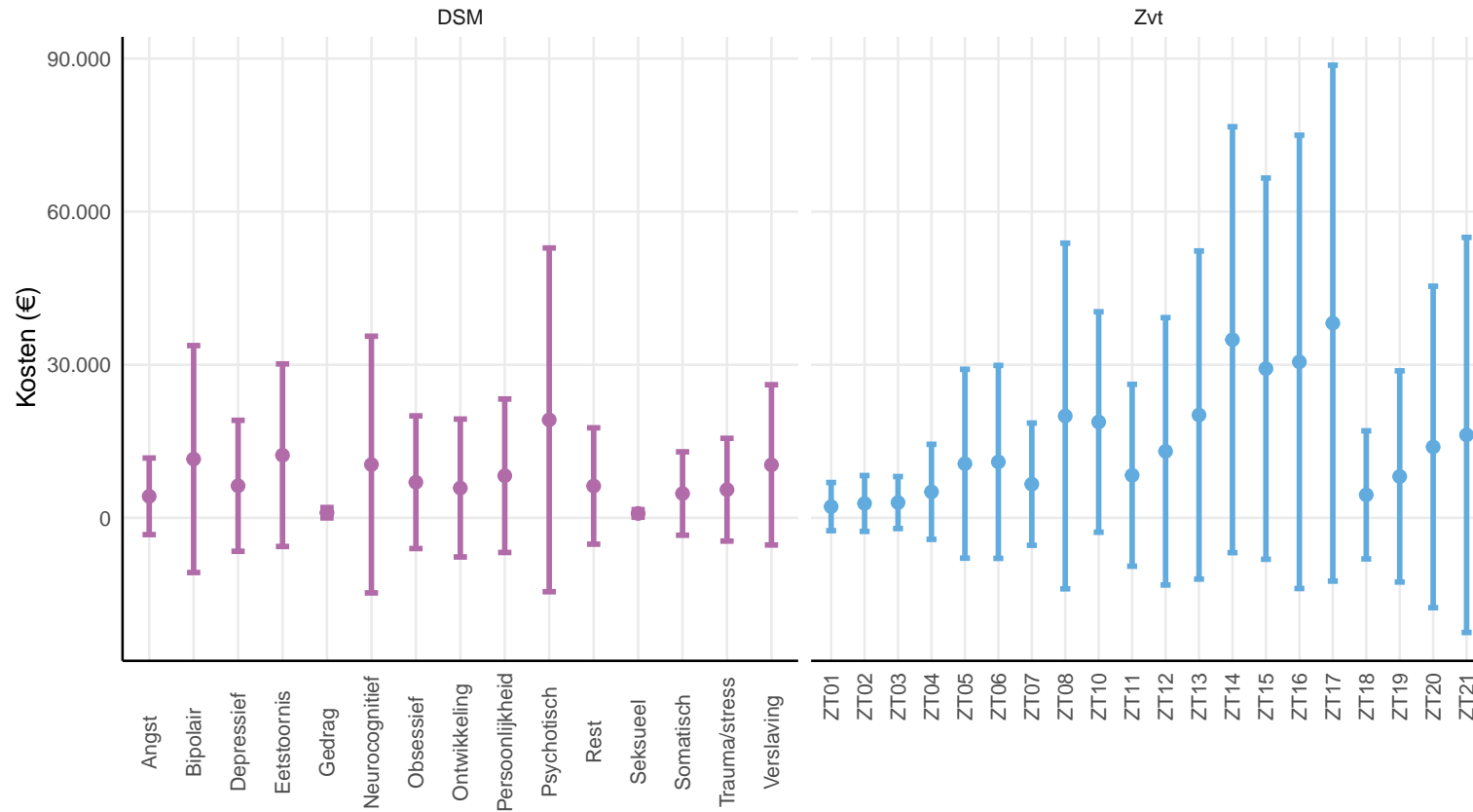
Figuur 76: Verdeling kosten per diagnosehoofdgroep



### 3.4.4 Gemiddelde kosten

Zowel voor de diagnosehoofdgroep, als voor de zorgvraagtypering is in alle groepen de spreiding groot rond het gemiddelde. De gemiddelde kosten van de groepen met hoge kosten zijn voor de zorgvraagtypering een stuk hoger. Daarmee lijkt de zorgvraagtypering beter in staat groepen te onderscheiden met gemiddeld hoge kosten.

Figuur 77: Gemiddelde en spreiding van kosten per diagnosehoofdgroep of zorgvraagtype





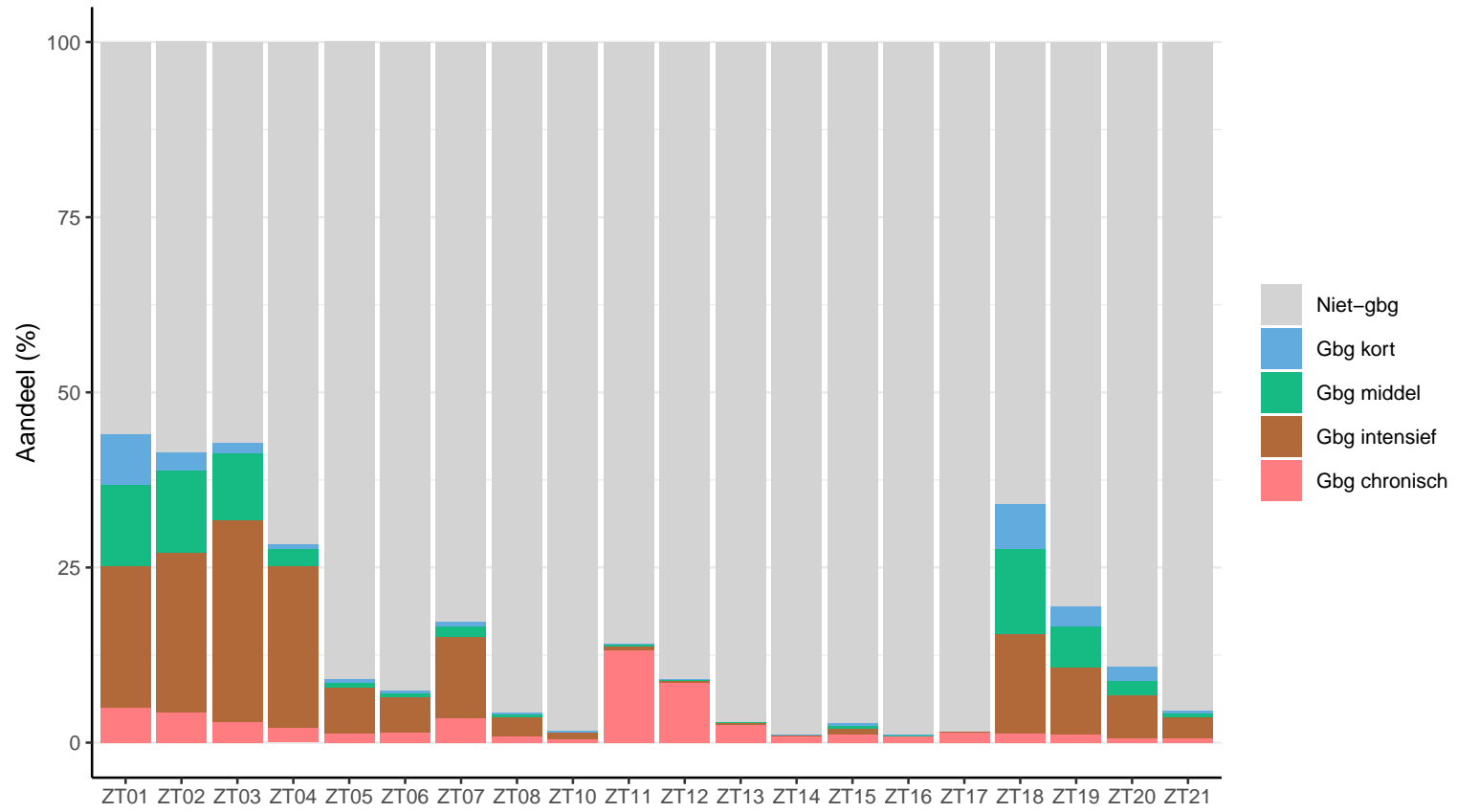
### 3.5 Generalistische basis-ggz

De gb-ggz-profielen komen in hoofdgroep X voornamelijk voor bij ZT01-ZT03 en in mindere mate bij ZT04 en ZT07. In hoofdgroep Y komt alleen bij ZT11 en ZT12, de relatief stabiele psychotische zorgvraagtypes, een noemenswaardig aandeel basis-ggz voor. In hoofdgroep Y is dit voornamelijk het chronische profiel. In hoofdgroep Z valt op dat er een relatief groot aandeel ZT18 een basis-ggz-profiel heeft (30%) en dat dit aandeel afneemt met de ernst van de cognitieve problemen.

NB het aandeel trajecten met een basis-ggz-profiel komt niet één-op-één overeen met het aandeel trajecten zonder DSM. Dit komt omdat voor een klein aandeel van de zorgtrajecten zowel een basis-ggz-profiel als DSM bekend is en voor een vergelijkbaar aandeel geen van beide bekend is.

### 3.5.1 Aandeel per gbg-profiel

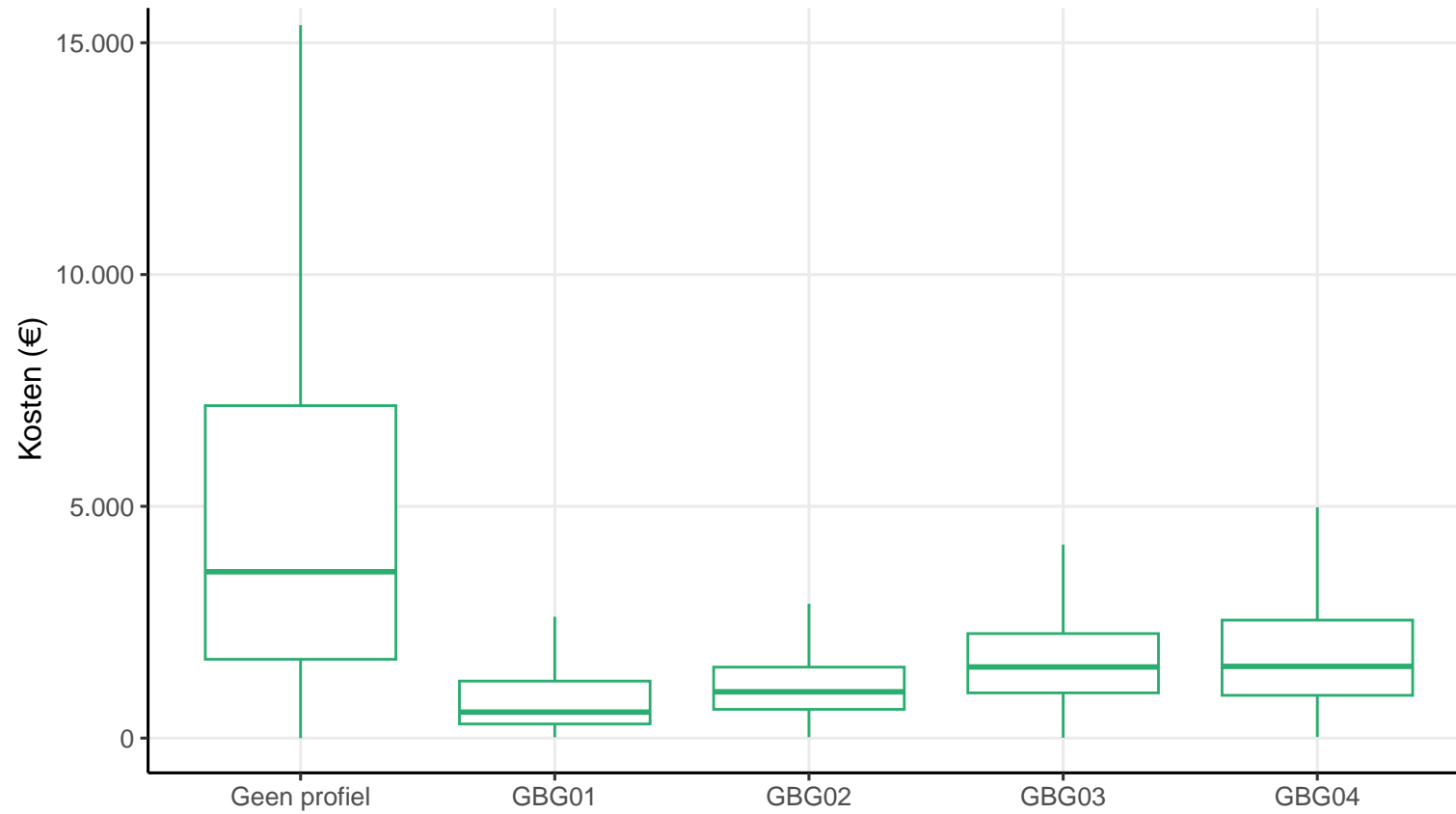
Figuur 78: Voorkomen basis-ggz-profielen per zorgvraagtype



### 3.5.2 Kosten per gbg-profiel

De verdeling van kosten loopt geleidelijk op van profiel GBG01 tot GBG04. De verdeling van alle profielen lijkt lager en de spreiding kleiner dan van de trajecten zonder gbg-profiel.

Figuur 79: Verdeling kosten per gbg-profiel



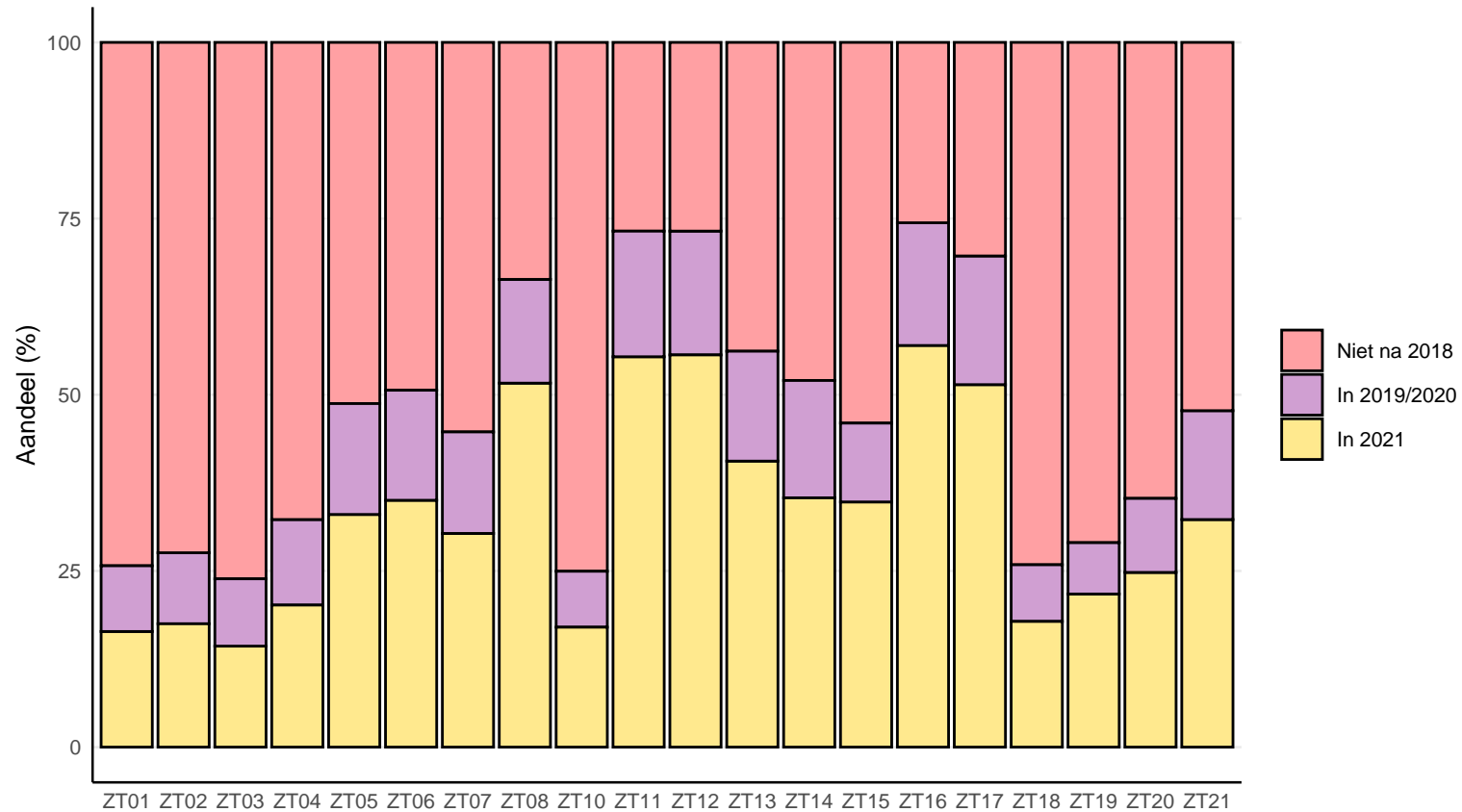
### **3.6 Eerdere s-ggz behandeling**

Tijdens bijeenkomsten van de werkgroep data-analyse werd genoemd dat het mogelijk is dat doorlopende of eerdere behandeling van invloed kan zijn op de kosten en dus bekeken moet worden in relatie tot de de indeling in zorgvraagtypes en de kosten. Daarom bekijken we bij de geselecteerde trajecten of de patiënt ook in 2021, of 2019-2020 is behandeld in de specialistische ggz. We bekijken dus niet of er behandeling is geweest bij POH-ggz of basis-ggz. Voor vergelijkbaarheid nemen we bij deze analyse alleen trajecten mee die in 2022, maar na januari zijn begonnen en die niet door zijn gelopen vanuit de tweede helft van 2021. Behandeling van langer dan drie jaar voor invoering van het zorgprestatie model en zorgvraagtypering (2018 en eerder) nemen we niet mee in deze analyse.

### 3.6.1 Aandeel eerder s-ggz

Voor de 'lichtere' zorgvraagtypes in hoofdgroep X is er in het grootste deel van de gevallen geen ggz-zorg geweest na 2018. Voor ZT05-ZT07 is het aandeel met zorg in de specialistische ggz ongeveer de helft, terwijl in ZT08 dit ruim 2/3 van de trajecten is. In hoofdgroep Y heeft een veel groter aandeel van de trajecten eerder specialistische ggz ontvangen, tot 75% voor ZT11, ZT12, ZT 16 en ZT17. In hoofdgroep Y heeft alleen in ZT10 (eerste psychotische episode) minder dan 50% van de trajecten eerder s-ggz ontvangen, namelijk ongeveer 25%. In hoofdgroep Z is het aandeel weer lager dan in hoofdroep Y en het loopt gradueel op met de ernst van het zorgvraagtype.

Figuur 80: Voorkomen eerdere ggz-behandeling



### 3.6.2 Kosten en eerder s-ggz-gebruik

Er lijkt geen grote afhankelijkheid van het al dan niet eerder gebruiken van specialistische ggz in hoofdgroep X en Z. In hoofdgroep Y lijken de kosten iets lager te liggen wanneer er ook in 2021 behandeld is. Het is mogelijk dat we voor deze gevallen niet de intensieve beginfase van de behandeling in de kosten zien, omdat deze fase in 2021 lag.

Figuur 81: Verdeling van kosten afhankelijk van eerder gebruik van s-ggz



### 3.7 Hertypering

Werkgroep data heeft gevraagd om een analyse naar het evalueren of hertyperen van patiënten. In de dataset met scorelijsten zijn opeenvolgende zorgvraagtypering van dezelfde patiënt niet bij elkaar te zoeken, daarom gebruiken we voor deze analyse de declaratiegegevens.

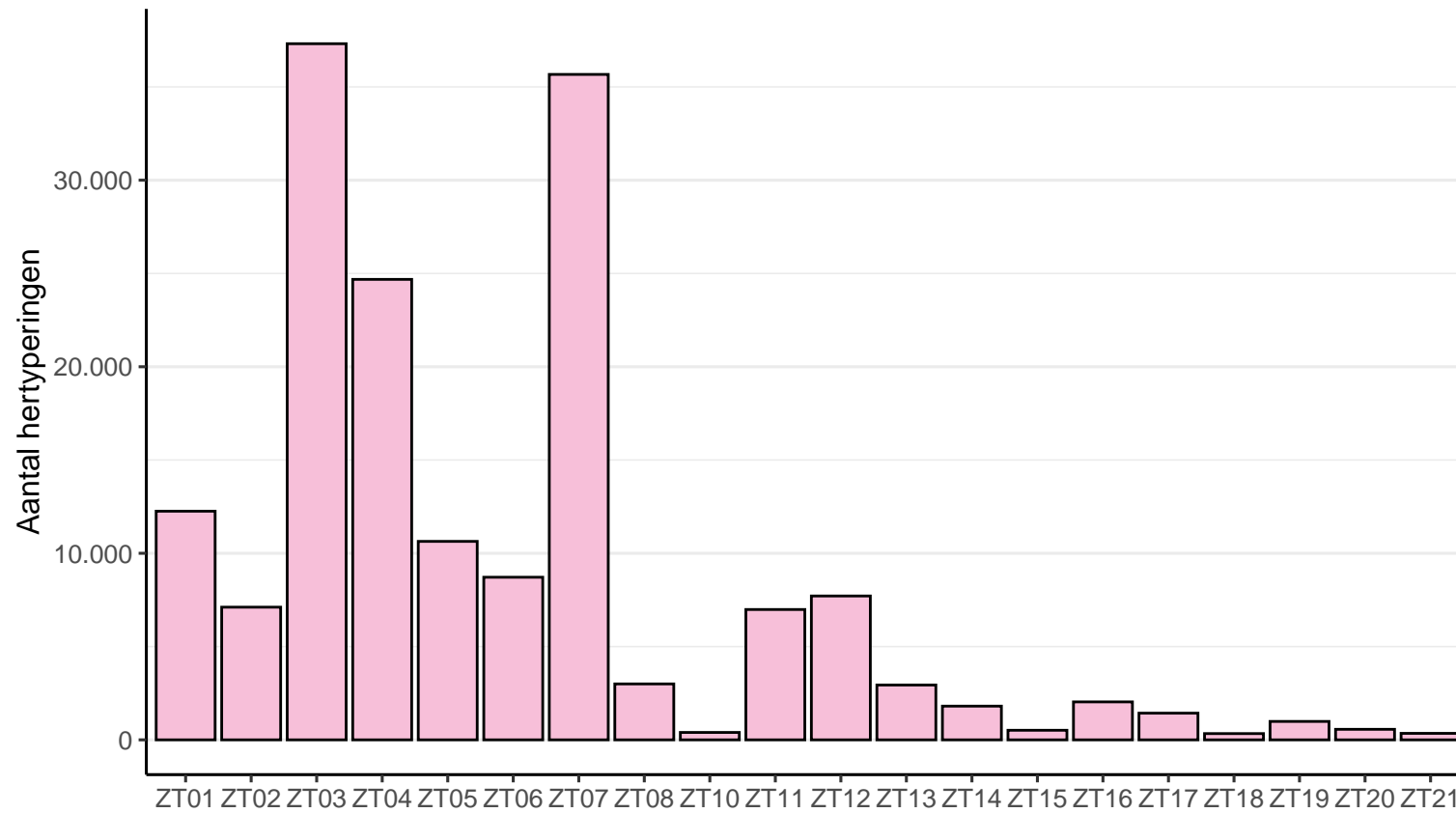
Omdat we weten dat (zelden) de koppeling met het zorgvraagtype niet goed is en er dus kleine aantallen verspringende zorgvraagtypering voor kunnen komen, selecteren we alleen trajecten als er drie of meer prestaties met het oude zorgvraagtype zijn én drie of meer prestaties met het nieuwe zorgvraagtype.

NB. Wanneer opnieuw zorgvraagtypering wordt uitgevoerd en hierbij hetzelfde zorgvraagtype wordt gekozen zien we dit niet terug in de gegevens, op de factuur staat immers hetzelfde zorgvraagtype. Daarnaast kan het voorkomen dat er meermaals wordt geëvalueerd, we nemen veranderingen van zorgvraagtype mee, mits het voldoet aan de drie-prestatie-regel van hierboven. Het kan voorkomen dat er meerdere identieke wisselingen in één traject zijn, zoals meerdere wissels van ZT01 naar ZT02 en terug, in dat geval nemen we elke uniek type wisseling één keer mee. Door bovenstaande filters betreffen de analyses naar de hertyperingsgegevens een beperktere set dan de eerdere analyses van de declaratiegegevens.

### 3.7.1 Aantal hertyperingen

Niet alle zorgvraagtyperingen hebben dezelfde voorgestelde evaluatietermijn, het is dus te verwachten dat er verschil is tussen de aantallen hertyperingen. Het aantal hertyperingen laat ook zien hoe betrouwbaar de hieropvolgende analyses zijn. Hieronder maken we het aantal hertyperingen per (oorspronkelijk) zorgvraagtype inzichtelijk.

Figuur 82: Aantal hertyperingen



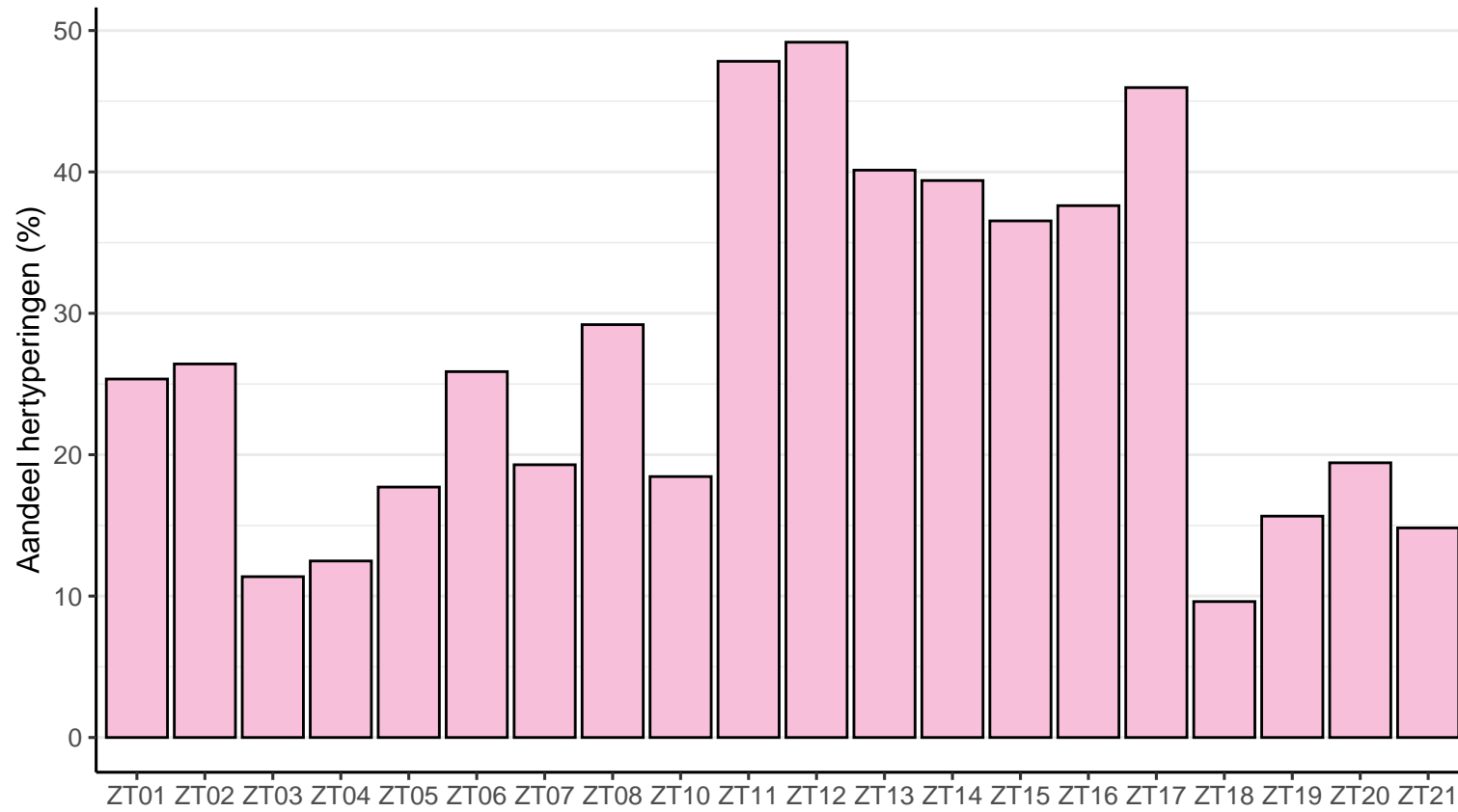


### 3.7.2 Aandeel van trajecten

Er zijn grote verschillen tussen de zorgvraagtypes op het gebied van het aantal trajecten. Daarom is voor sommige zorgvraagtyperingen het aandeel van de trajecten waarvoor een ander zorgvraagtype bestaat informatiever dan het absolute aantal. Hieronder beschrijven we het aantal hertyperingen als percentage van het aantal trajecten, zoals we hierboven steeds hebben gebruikt, op basis van het zorgvraagtype vóór de hertypering. Het kan voorkomen dat er voor één traject meerdere hertyperingen zijn uitgevoerd en het kan zo zijn dat de meeste kosten van een traject gemaakt worden na de typering en dus dat het gehele traject niet het label heeft van voor de hertypering.

Het valt op dat van de trajecten in hoofdgroep X hertyperingen bestaan ter grootte van zo'n 10-30% van het aantal trajecten. In hoofdgroep Y is het aandeel hertyperingen iets groter, veelal 35% tot 50% ten opzichte van het aantal trajecten, behalve 18% voor ZT10. In hoofdgroep Z wordt minder vaak hertypering uitgevoerd (10-20%), dit aandeel is het laagst voor ZT18.

Figuur 83: Aandeel trajecten met hertyping per zorgvraagtype



### 3.7.3 Hertypering - zvt voor en na

In onderstaande tabel laten we zien in welk zorgvraagtype de patiënt na hertypering vaak voor komt. De rijen geven het zorgvraagtype vóór hertypering, de kolommen het zorgvraagtype erna. De getallen geven het aandeel ten opzichte van het zorgvraagtype voor hertypering, de rijen tellen dus (bij benadering, gezien afrondingen) op tot 1.

Het valt op dat het grootste deel van de zorgvraagtypering binnen de hoofdgroep opnieuw wordt getypeerd, maar dat het ook regelmatig voor komt dat een andere hoofdgroep wordt geselecteerd, dit lijkt vooral te gelden voor patiënten die vanuit hoofdgroepen Y en Z na hertypering in de 'grote' zorgvraagtypes ZT03, ZT04 en ZT07 terecht komen.

Tabel 17: Hertypering

ZVT	ZT01	ZT02	ZT03	ZT04	ZT05	ZT06	ZT07	ZT08	ZT10	ZT11	ZT12	ZT13	ZT14	ZT15	ZT16	ZT17	ZT18	ZT19	ZT20	ZT21
ZT01		0.08	0.48	0.16	0.04	0.02	0.17	0.01		0.02										
ZT02	0.11		0.44	0.15	0.04	0.02	0.21	0.01		0.01										
ZT03	0.09	0.06		0.38	0.08	0.04	0.31	0.01		0.01	0.01									
ZT04	0.03	0.02	0.34		0.17	0.06	0.32	0.02		0.01	0.01									
ZT05	0.02	0.01	0.15	0.3		0.13	0.29	0.05		0.01	0.01				0.01					
ZT06	0.01	0.01	0.11	0.16	0.19		0.43	0.04		0.01	0.01	0.01			0.01					
ZT07	0.03	0.02	0.27	0.29	0.13	0.15		0.04		0.01	0.02	0.01								
ZT08	0.02	0.01	0.09	0.16	0.19	0.12	0.34			0.01	0.02	0.01	0.01	0.01	0.01	0.01				
ZT10	0.02	0.01	0.09	0.07	0.03	0.04	0.11	0.02		0.1	0.09	0.14	0.12	0.02	0.11	0.03				
ZT11	0.03	0.01	0.05	0.03	0.02	0.01	0.07	0.01	0.05		0.4	0.11	0.09	0.01	0.06	0.04				
ZT12	0.01	0.01	0.03	0.02	0.01	0.02	0.08	0.01	0.02	0.29		0.22	0.1	0.01	0.09	0.08				
ZT13	0.01	0.01	0.02	0.02	0.01	0.02	0.08	0.01	0.03	0.13	0.4		0.11	0.02	0.06	0.07				
ZT14	0.02	0.01	0.03	0.03	0.02	0.02	0.06	0.01	0.03	0.22	0.22	0.15		0.01	0.07	0.09				
ZT15	0.02		0.04	0.09	0.08	0.06	0.14	0.03	0.03	0.1	0.13	0.13	0.06		0.04	0.04				0.01
ZT16	0.01		0.02	0.03	0.03	0.03	0.08	0.02	0.03	0.15	0.27	0.11	0.09	0.01		0.12				
ZT17		0.01	0.02	0.01	0.02	0.02	0.05	0.02	0.01	0.12	0.25	0.12	0.18	0.02	0.15					
ZT18	0.02	0.02	0.15	0.12	0.02	0.02	0.1	0.01	0.01	0.02	0.02	0.01						0.36	0.07	0.04
ZT19	0.02	0.02	0.12	0.09	0.04	0.03	0.11	0.01		0.02	0.02	0.02	0.01		0.01		0.25		0.17	0.06
ZT20	0.01	0.01	0.04	0.04	0.04	0.02	0.08	0.01		0.01	0.04	0.03	0.02		0.01	0.01	0.07	0.43		0.15
ZT21	0.01	0.01	0.05	0.05	0.04	0.03	0.1	0.01		0.01	0.04	0.03	0.03	0.02	0.01		0.04	0.25	0.24	

### 3.7.4 Op- of afschalen

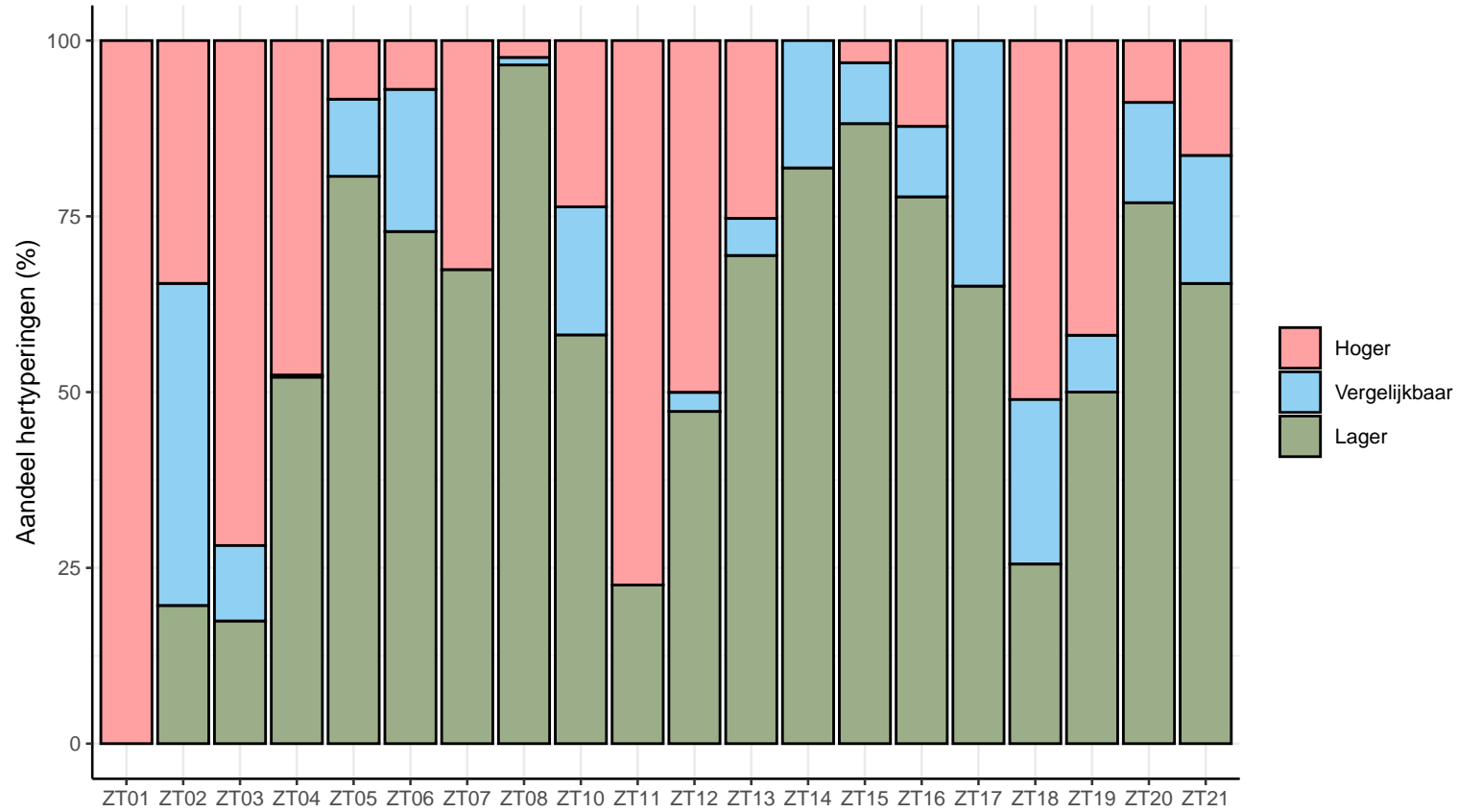
Wanneer er na hertypering een ander zorgvraagtype wordt gekozen kan dit betekenen dat de zorgvraag na behandeling is afgenomen, dit betekent mogelijk dat de zorg kan worden afgeschaald, of dat de zorgvraag juist is toegenomen en de zorg mogelijk moet worden opgeschaald. Niet voor alle combinaties van zorgvraagtyperingen is er een logisch verband van op- of afschalen. Om dit verband toch te benaderen gebruiken we de gemiddelde zorgkosten van de zorgvraagtypes, zoals berekend voor figuur 68.

We definiëren vijf groepen, op basis van het verschil in kosten tussen de zorgvraagtypes voor en na hertypering. Wanneer het zorgvraagtype na hertypering kosten heeft van 20% lager vergeleken met het zorgvraagtype voor hertypering behoort de hertypering tot de groep 'Lager' (afschalen). Tussen 20% lager en 20% hogere kosten behoort de hertypering tot de groep 'Vergelijkbaar' (beperkt of niet op- of afschalen). Bij een toename van 20% van kosten of meer is het een hertypering in de groep 'Hoger' (opschalen). We gebruiken hier de feiten over het hoger of lager zijn van de *gemiddelde kosten* van het zorgvraagtype, omdat het niet op voorhand bekend is of de zorg ook daadwerkelijk is op- of afgeschaald na hertypering.

Het kan zijn dat er meerdere keren hertypering wordt uitgevoerd. Om in de volgende analyses vergelijkbare groepen te hebben maken we twee selecties van de hertyperingen. In de eerstvolgende figuren bekijken we de hertyperingen gegroepeerd op basis van het zorgvraagtype dat gold vóór hertypering én degene was met de meeste kosten in het geselecteerde jaar van het traject. Daarna bekijken we de hertyperingen waarvoor het zorgvraagtype ná hertypering degene is met de meeste kosten in het geselecteerde jaar van het traject.

**Aandeel op-/afschaalgroepen per zvt - vóór hertyping** Zoals verwacht wordt er in de lagere zorgvraagtypes voornamelijk opgeschaald, het is bijvoorbeeld niet mogelijk af te schalen vanaf ZT01. In de complexere/duurdere zorgvraagtyperingen wordt meer afgeschaald dan opgeschaald. Vooral vanuit ZT08 wordt bijna alleen bewogen naar zorgvraagtypes met lagere kosten.

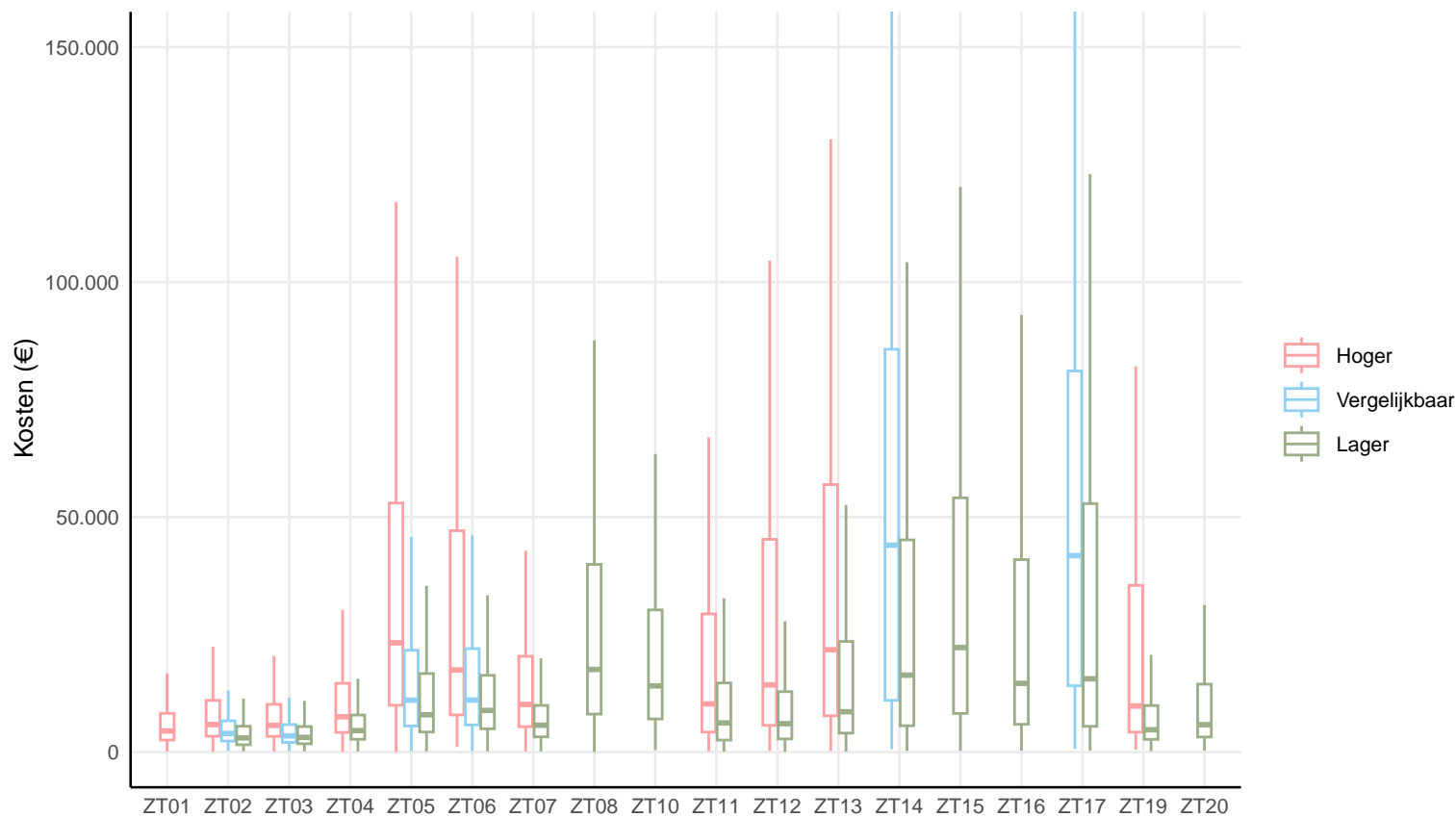
Figuur 84: Aandeel op- en afschaalgroepen



**Kosten op-/afschaalgroepen per zvt - vóór hertypering** We kunnen kijken hoe de kosten zich verhouden voor de zorgvraagtyperingen uit de verschillende groepen per zorgvraagtype. We visualiseren hier alleen kosten als er minimaal 50 trajecten in de groep zitten. Doordat er weinig hertyperingen zijn, het zorgvraagtype vóór hertypering niet altijd overeen komt met het zorgvraagtype met de meeste kosten en door de splitsing in drie groepen komt het voor veel zorgvraagtypering-groep-combinaties voor dat deze minder dan 50 trajecten bevatten en dus niet worden gevisualiseerd.

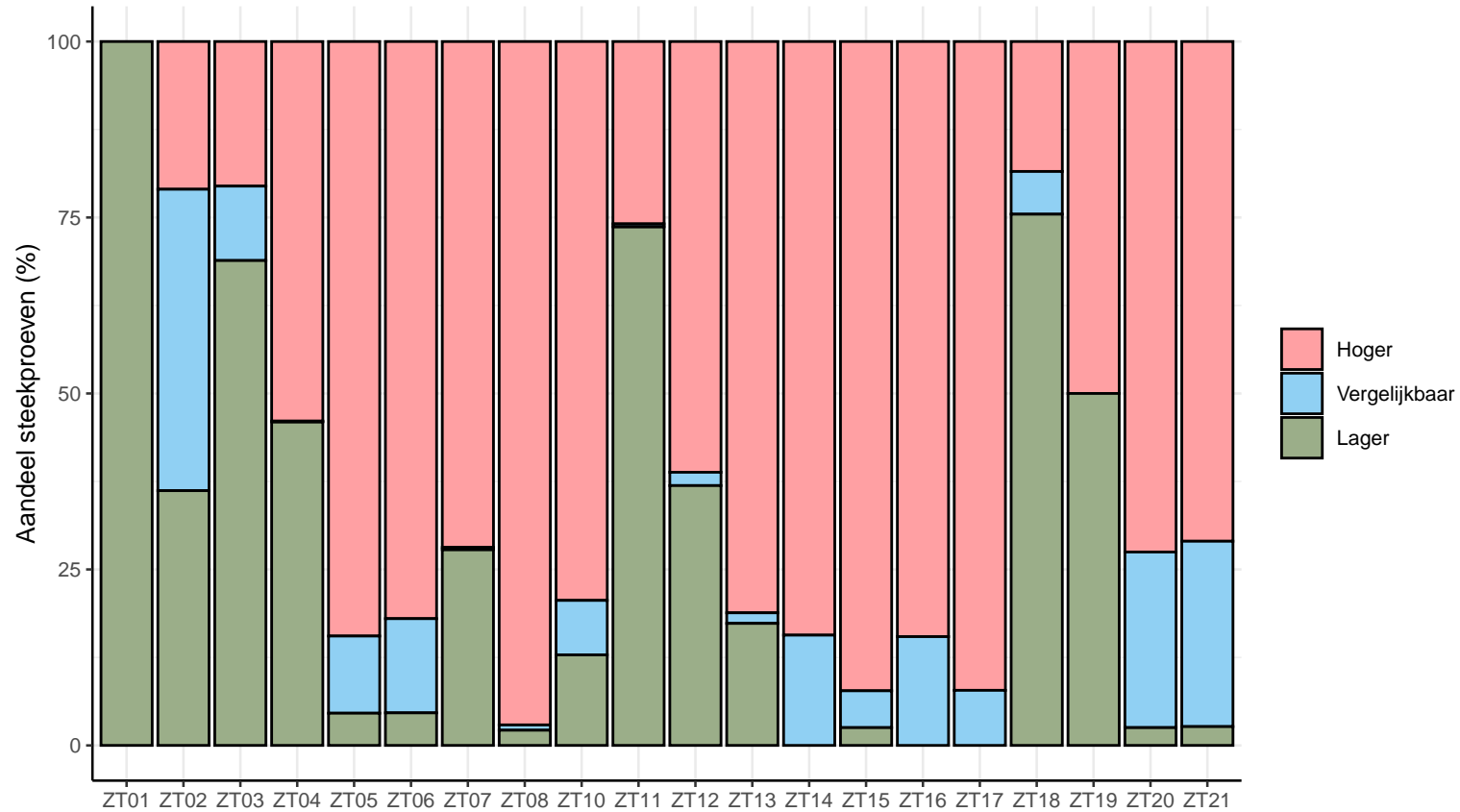
Zoals te verwachten lijkt de verdeling van kosten van trajecten die naar een hoger zorgvraagtype gaan hoger te liggen dan hertypering naar een vergelijkbaar zorgvraagtype, terwijl de verdeling van kosten van de groep die naar een lager zorgvraagtype bewegen lager lijkt te liggen.

Figuur 85: Verdeling van kosten per op- afschaalgroep



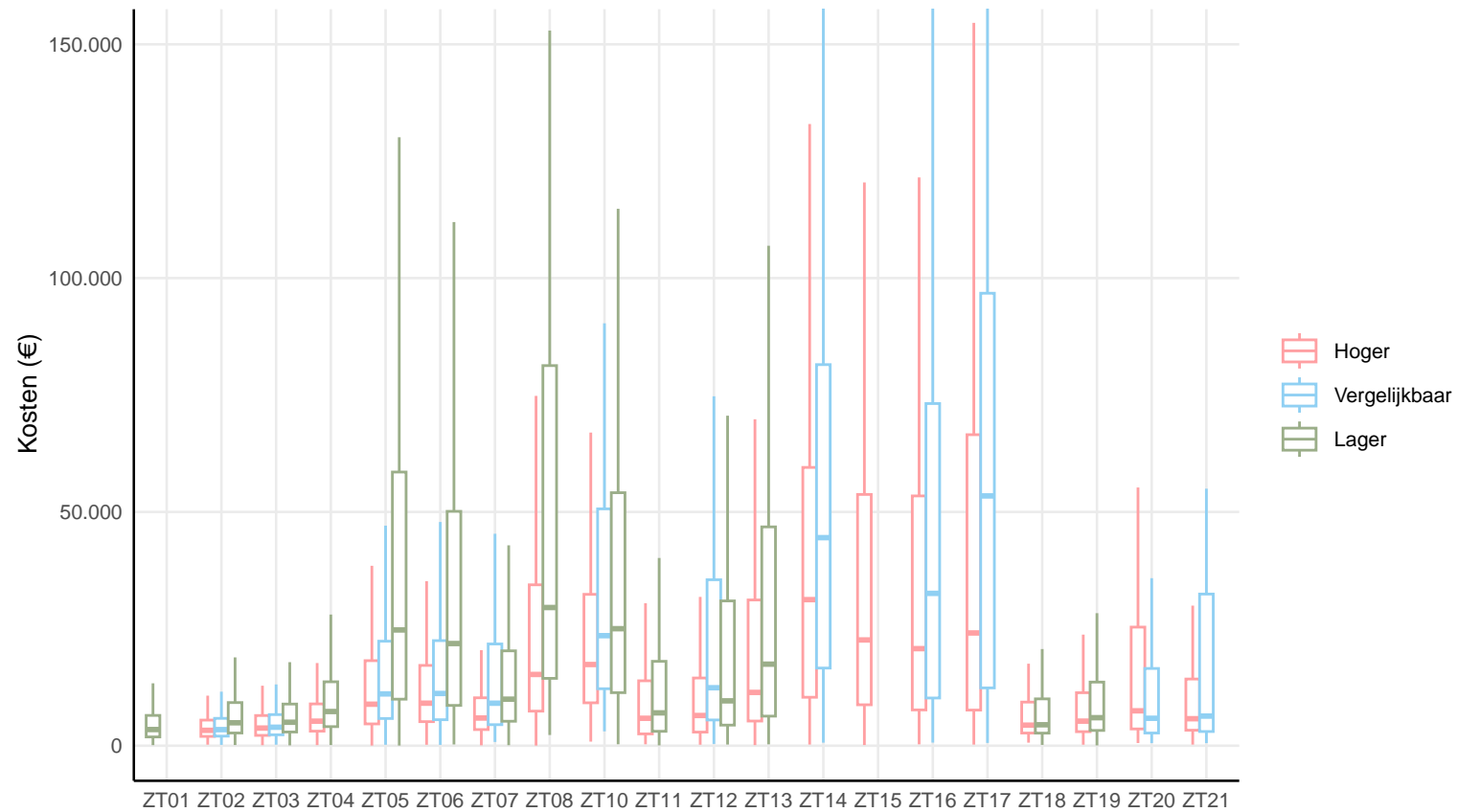
**Aandeel op-/afschaalgroepen per zvt - na hertypering** Wanneer we de hertyperingen filteren zodat het zorgvraagtype na hertypering de meeste kosten met zich mee brengt zien we een omgekeerde verdeling van de kostengroepen. In de lichtere zorgvraagtypes komt het logischerwijs vaker voor dat het nieuwe zorgvraagtype lagere kosten heeft dan het zorgvraagtype voor hertypering. Bij de complexere zorgvraagtypes komt het vaker voor dat het nieuwe zorgvraagtype hogere kosten heeft dan het zorgvraagtype voor hertypering.

Figuur 86: Aandeel op- en afschaalgroepen



**Kosten op-/afschaalgroepen per zvt - na hertypering** Ook bij de verdeling van kosten is een vergelijkbaar, maar omgekeerd patroon te zien met groepering voor hertypering. De verdeling van kosten is lager voor de groepen die vanaf een goedkoper zorgvraagtype zijn opgeschaald, terwijl de verdeling van kosten hoger is wanneer is afgeschaald vanaf een duurder/complexer zorgvraagtype.

Figuur 87: Verdeling van kosten op- en afschaalgroepen



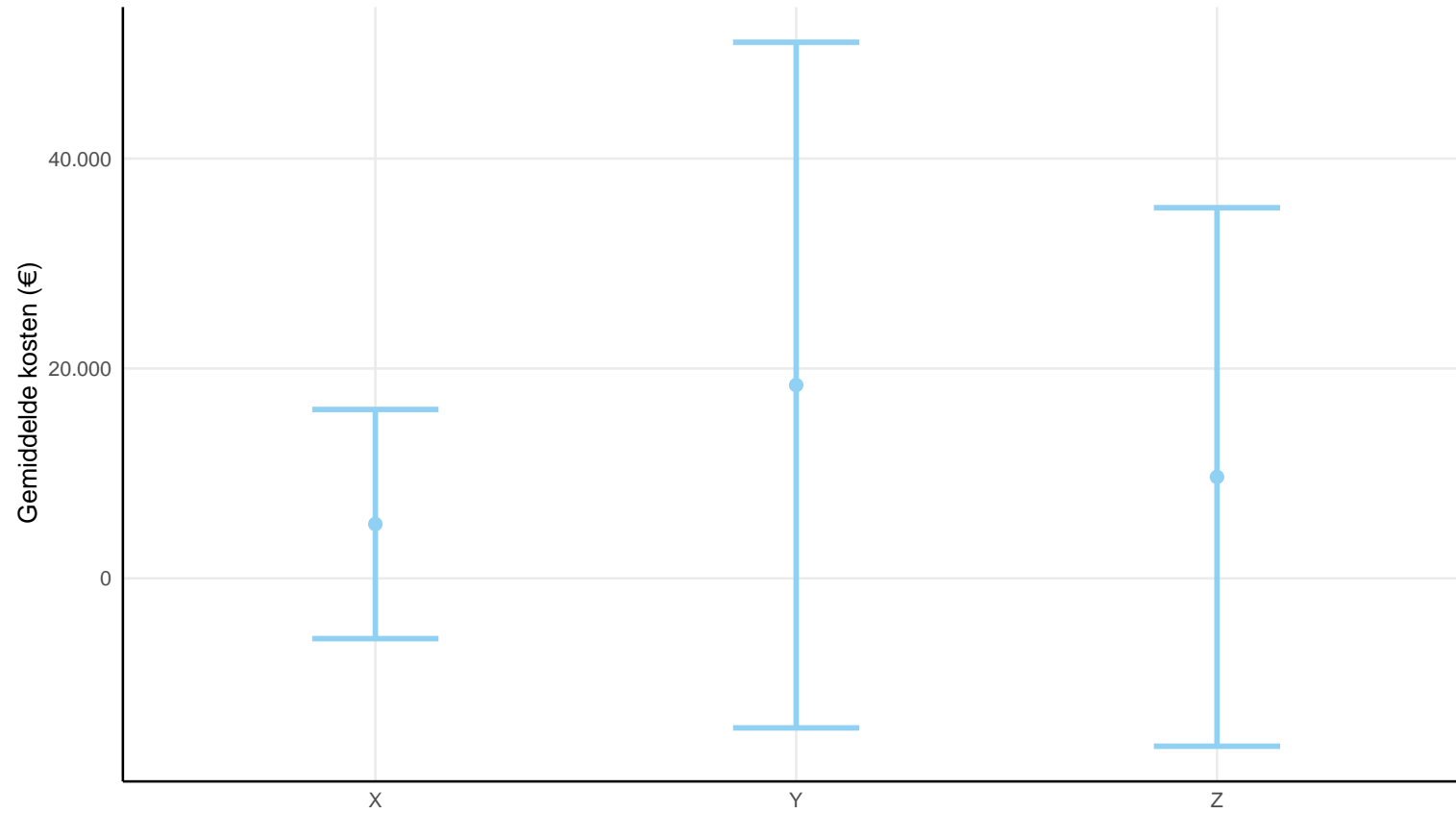


### 3.8 Samenvoegen zorgvraagtypes

In de adviescommissie kwam het voorstel naar voren om zorgvraagtypes samen te voegen en de homogeniteit en onderscheidendheid te onderzoeken. Er zijn verschillende manieren waarop dit kan worden uitgevoerd. Zo kunnen alle zorgvraagtypes in een hoofdgroep bij elkaar worden geveegd of er kunnen op subtielere manieren groepen worden samengevoegd. Hieronder werken we enkele varianten uit. Deze varianten zijn verderop beoordeeld met de kwaliteitsparameters die de adviescommissie heeft gekozen. Het is goed ook te kijken naar samenvoegingen samen met splitsingen uit de zorgvraagtyperingsgegevens, zoals beschreven in tabel 12 en bijbehorende figuren 46 tot en met 49.

### 3.8.1 Samenvoegen tot hoofdgroepen

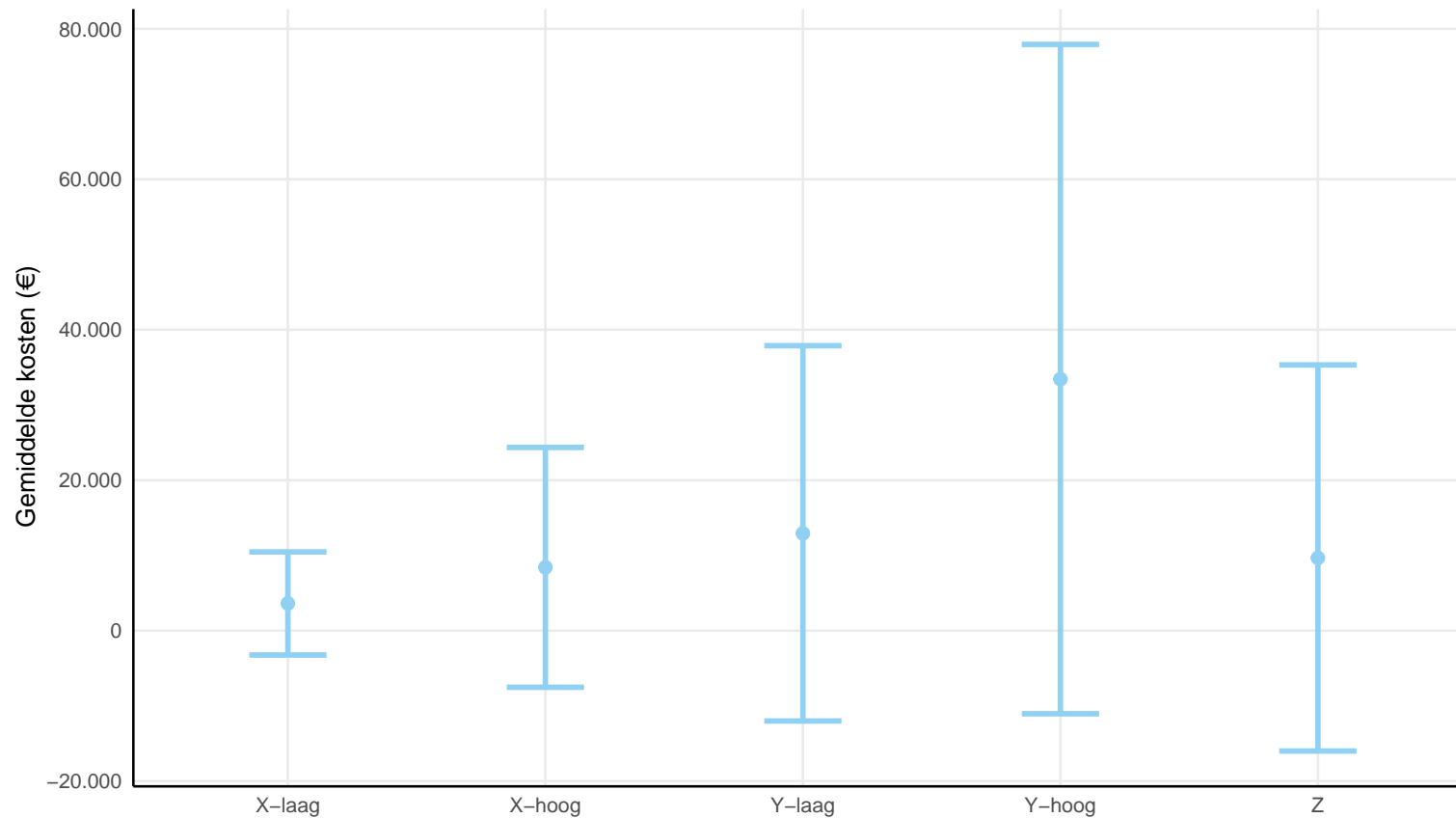
Figuur 88: Gemiddelde kosten en spreiding bij samenvoegen op het niveau van de hoofdgroepen



### 3.8.2 Samenvoegen subgroep

Alternatief kunnen we samenvoegen op het niveau van groepen binnen een hoofdgroep. In dit geval hebben we gekozen voor de groepen zoals ze hierboven ook in afzonderlijke subplots zijn weergegeven. Specifiek voegen ZT01-ZT04 samen tot X-laag, ZT05-ZT08 tot X-hoog, ZT10-ZT13 tot Y-laag, ZT14-ZT17 tot Y-hoog en hoofdgroep Z. Alternatief kan gedacht worden aan samenvoegen van groepen in hoofdgroep Y tot groepen met en zonder acuut psychotische episode. Dit hebben we niet uitgewerkt, omdat uit eerdere analyses blijkt dat gemiddelde kosten van ZT10 (eerste acute episode) beduidend lager liggen dan van de andere twee groepen met acute episode (ZT14 en ZT15), terwijl ZT16 en ZT17 juist hogere kosten hebben.

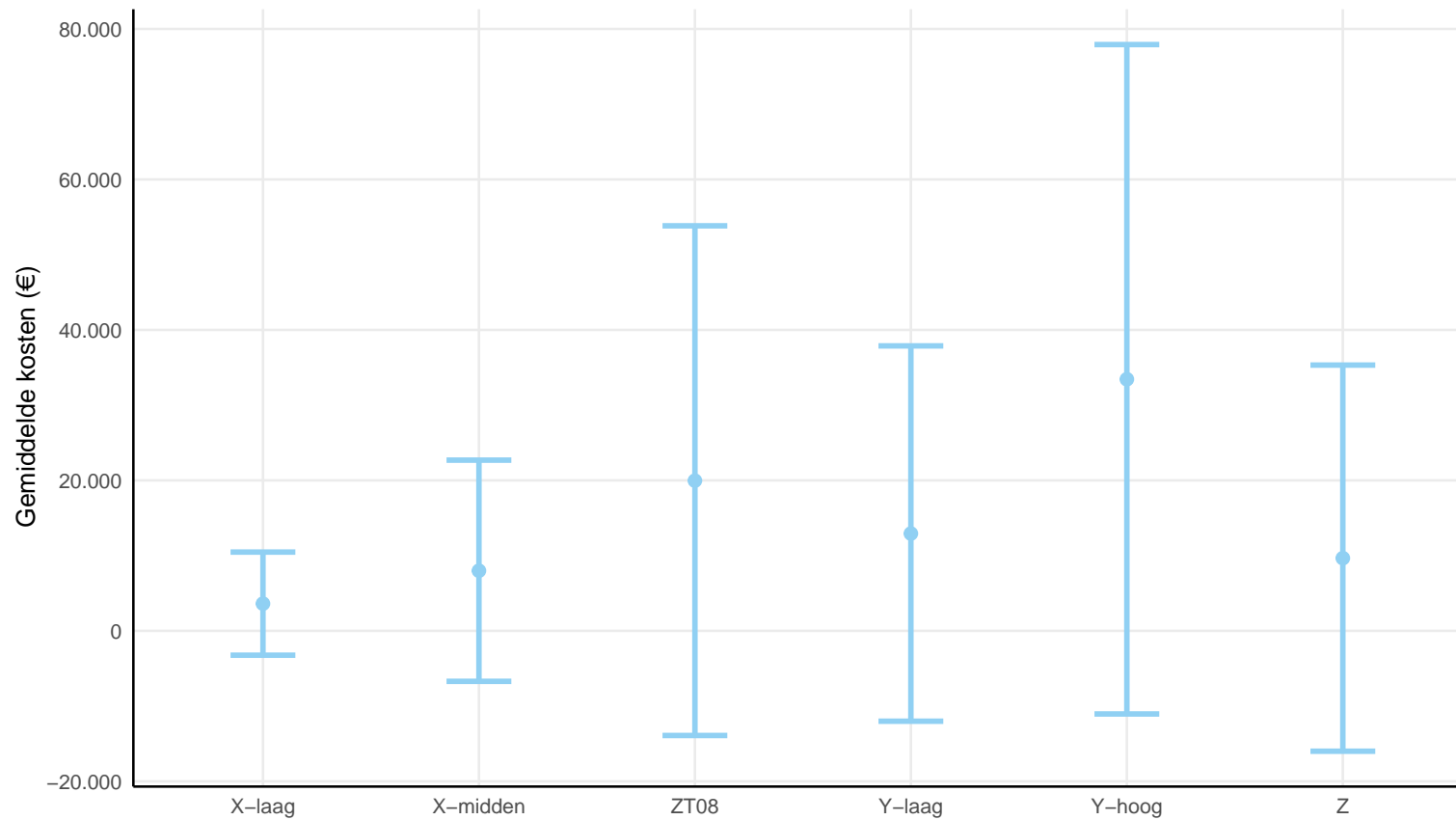
Figuur 89: Gemiddelde kosten en spreiding bij samenvoegen op het niveau van groepen binnen de hoofdgroep



### 3.8.3 Samenvoegen obv clustering

We kunnen de computer zorgvraagtypes laten clusteren op basis van de gemiddelde kosten bij het zorgvraagtype. Op deze manier komen we uit op drie groepen voor hoofdgroep X, namelijk ZT05-ZT07 samen, ZT08 los en als laatste X-laag, zoals hierboven. De andere hoofdgroepen blijven ingedeeld in Y-laag en Y-hoog en Z.

Figuur 90: Gemiddelde kosten en spreiding bij samenvoegen op het niveau van groepen binnen de hoofdgroep obv clustering



Tabel 18: Kwaliteitsparameters bij samenvoegen

Voorspeller	MAE (€)	Verklaarde basisfout ( )
Hoofdgroep	2.911	28.1
Subgroep	2.546	42.7
Clustergroep	2.494	44.8
Zvt	2.389	47.8

## 4 Beoordelen kwaliteit model

De kwaliteit van het model kan beoordeeld worden met een tweetal kwaliteitsparameters waarover de werkgroep data heeft geadviseerd. Voor deze parameters wordt een voorspelling van de zorgkosten gedaan op basis van het zorgvraagtype (of DSM ter vergelijking). Ten eerste wordt gekeken naar het absolute verschil in kosten tussen de voorspelling en de daadwerkelijke kosten per patiënt op het gekozen aggregatieniveau. Ten tweede wordt gekeken naar de hoeveelheid (kwadratische) voorspelfout na voorspelling met simpele gemiddelde kosten per patiënt kan worden verklaard door voorspellingen te doen met zorgvraagtype (of DSM).

Uitgangspunt bij de berekeningen is dat bij de beoordeling twee gegevens worden gebruikt van elke patiënt. Ten eerste nemen we de daadwerkelijke kosten die zijn gemaakt in een afgebakende periode van het zorgtraject. Ten tweede wordt het zorgvraagtype (of DSM) genomen dat de behandelaar heeft gekozen en dat op de factuur is vermeld. Bij dit zorgvraagtype kunnen de gemiddelde kosten die horen bij het zorgvraagtype worden genomen als voorspelde kosten. Het verschil (voorspelfout) tussen de voorspelde en daadwerkelijke kosten ligt aan de basis van de beoordeling van het model.

### 4.1 Niveau beoordeling

Bij het tot stand komen van het zorgvraagtyperingsmodel is steeds duidelijk benoemd dat het model niet bedoeld is om de zorg van het individu te beoordelen, maar om het gesprek te voeren over de zorg van groepen patiënten. Het lijkt daarom niet logisch om de kwaliteit van het model op het niveau van de individuele patiënt en zijn behandeling te doen, maar eerder om de individuele voorspellingen te aggregeren op een hoger niveau. Er bestaan verschillende alternatieve niveau's waarop de beoordeling kan worden uitgevoerd. Zo kan de aggregatie worden uitgevoerd op het niveau van een zorgaanbieder, de zorgverzekeraar, het contract tussen verzekeraar en aanbieder, de regio of de setting. Hoe hoger het aggregatie-niveau hoe betrouwbaarder de uitleeswaarde wordt, omdat er meer gegevens per groep worden samengevoegd in de aggregatie. Tegelijk verliezen we met een keuze voor een hoog aggregatieniveau de mogelijkheid om patronen en knelpunten te herkennen en om verdere uitsplitsingen te doen, zoals naar type aanbieder, regio of setting. Voor onderstaande analyses is op advies van de werkgroep data gekozen voor het aggregeren op het niveau van de zorgaanbieder.

## 4.2 Methoden beoordeling

Er zijn verschillende rekenkundige methoden mogelijk om de kwaliteit van het model te beoorelen. Zoals ook hierboven beschreven zijn al deze methoden gestoeld op de vergelijking tussen de voorspelde kosten (gemiddelde van het zorgvraagtype) en de daadwerkelijke kosten in een (afgebakende periode van) een zorgtraject. Niet alle onderstaande methoden zijn uitgevoerd, maar de beschrijving laten we hier staan als naslag. De werkgroep data heeft geadviseerd kwaliteit te beoordelen op basis van de MAE en een afgeleide van de R-kwadraat.

- Mean Absolute Error (MAE): Dit meet de gemiddelde absolute afwijking tussen de voorspelde en werkelijke kosten op het gekozen aggregatieniveau. Het geeft inzicht in de gemiddelde fout zonder rekening te houden met de richting van de fout. Een overschatting van kosten met €1.000 weegt net zo zwaar als een onderschatting met €1.000. De MAE is eenvoudig te interpreteren, omdat het euro's als eenheid heeft. In tegenstelling tot kwadratische methodes is de MAE niet extra gevoelig voor grote afwijkingen.
- Mean Squared Error (MSE): Dit kwantificeert de gemiddelde kwadratische fout op het gekozen aggregatieniveau, waarbij grotere fouten zwaarder worden bestraft. Het is gevoelig voor grote afwijkingen en uitbijters. De waarde van de MSE is moeilijk te interpreteren, omdat het niet in euro's wordt gegeven.
- Root Mean Squared Error (RMSE): Dit is de wortel van de MSE en is interpreteerbaar in dezelfde eenheden als de kosten. Het helpt bij het begrijpen van de afwijking van de fouten. De RMSE is gevoelig voor uitbijters.
- Mean Absolute Percentage Error (MAPE): Dit meet het gemiddelde percentage afwijking tussen voorspelde en werkelijke kosten. Het is nuttig voor het begrijpen van de relatieve fout, dit is nuttig wanneer op een aggregatieniveau zowel lage als zeer hoge kosten voorkomen. De MAPE kan verstoord worden door kleine waarden, €1.000 is immers een groot percentage ten opzichte van ZT01 (~€2.500), maar een laag percentage ten opzichte van ZT17 (~€55.000). Daarnaast is de MAPE niet symmetrisch, overschattingen en onderschattingen worden niet even zwaar gewogen.
- $R^2$  (R-kwadraat): Dit meet de verklaring van de variantie in de werkelijke kosten door de voorspelde kosten. De R-kwadraat is gestandaardiseerd en kan daarom gebruikt worden voor het vergelijken van verschillende modellen. Daarnaast is de R-kwadraat zo opgezet dat grote voorspelfouten zwaarder wegen. De R-kwadraat verklaart variantie, maar is geen directe maatstaf voor de voorspellingsfout.

Voor de MAE, de MSE en de RMSE is het nuttig te overwegen of de afwijking op het aggregatieniveau nog moet worden gecompenseerd voor het aantal patiënten. Door bijvoorbeeld de MAE op het niveau van een zorgaanbieder te delen door het aantal patiënten bij de zorgaanbieder kan een afwijking per patiënt worden berekend. Dit kan helpen wanneer het functioneren van het model moet worden vergeleken tussen twee groepen met zeer verschillende aantallen patiënten, zoals vergelijkingen tussen grote en kleine zorgaanbieders of regio's met veel en weinig bewoners.

### 4.3 Uitvoering kwaliteitsberekeningen

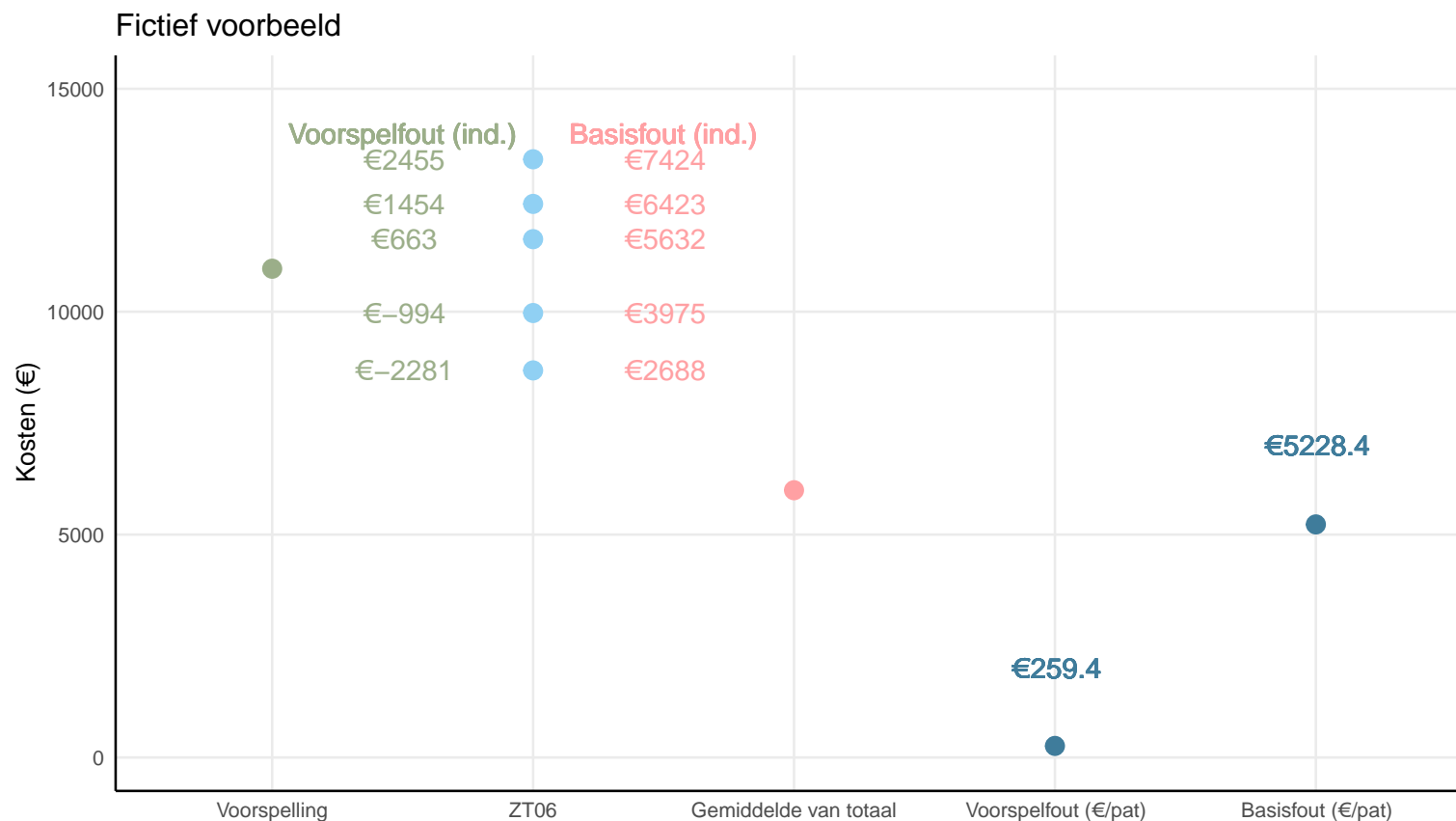
In de werkgroep data is gesproken over de manieren om de kwaliteit van het model te toetsen. De werkgroep heeft geadviseerd om in de beoordeling van het model en verbeteringen van het model een tweetal kwaliteitsparameters door te rekenen. Om precies te zijn adviseert de werkgroep om de gemiddelde absolute (voorspel-)fout per patiënt (MAE hierboven) te berekenen op het niveau van de zorgaanbieder. Om precies te zijn berekenen we het verschil tussen de totaal voorspelde kosten en de totale daadwerkelijke kosten bij een aanbieder en dit verschil delen we door het aantal patiënten dat is behandeld om een *fout per patiënt* te krijgen. Deze waarde maken we absoluut (negatieve waarden worden positief gemaakt) voor het (gewogen naar het aantal patiënten van de zorgaanbieders) te middelen. De absolute waarde zorgt ervoor dat positieve voorspelfouten in het gemiddelde niet worden 'hersteld' door negatieve voorspelfouten. Met deze waarde kan de adviescommissie een gevoel krijgen van het verschil in euro's tussen de voorspelde kosten op basis van zorgvraagtype en de daadwerkelijke kosten.

Daarnaast adviseert de werkgroep om de voorspelling in relatie te brengen tot een situatie zonder gegevens over de zorgvraag. Op advies van de werkgroep is dit uitgevoerd door per zorgaanbieder de fout per patiënt te nemen en daarnaast dezelfde berekening uit te voeren met de gemiddelde kosten per patiënt als voorspelling. Dit levert per zorgaanbieder twee waarden op, namelijk de voorspelfout en de 'basisfout'. Op deze manier kan worden bekeken welk percentage van de 'basisfout' 'afgedekt' wordt door de voorspelling. Om grote fouten zwaarder mee te wegen is gekozen voor een berekening vergelijkbaar met de R-kwadraat. Wiskundig bezien nemen we het verschil tussen de som van de kwadraten van de voorspelfout en de kwadraten van de basisfout (verklaarde kwadratische fout) en delen we dit door de som van de kwadraten van de basisfout om uit te komen op een percentage van de basisfout die is verklaard. Houd er rekening mee dat het kan zijn dat de voorspelfout groter kan zijn dan de basisfout, in dat geval is het percentage van de verklaarde fout negatief. Dit komt voor als de voorspelling op basis van het gemiddelde dichter bij de daadwerkelijke kosten ligt dan voorspelling op basis van het zorgvraagtype.

NB: Zoals gebruikelijk sluiten we kleine aanbieders uit, specifiek vrijgevestigden met op basis van de geselecteerde trajecten een omzet < €30.000 en instellingen met een omzet < €250.000.



Figuur 91: Voorbeeld van de berekening van de voorspelfout en basisfout



**Voorbeeld van kosten en voorspelling** Hierboven geven we een (fictief) voorbeeld van een aanbieder die vijf patiënten met zorgvraagtype ZT06 heeft behandeld en de kosten van de behandeling zijn opgeteld tot de kosten van een jaar van het traject (blauwe bolletjes). ZT06 heeft gemiddelde kosten van ongeveer €10.400, dit is de voorspelling (groene bolletje). De gedeclareerde zorg van de patiënten verschilt van de voorspelling, dit is de (individuele) voorspelfout. Deze fout middelen we en we nemen de absolute waarde van de uitkomst voor de absolute fout (absolute error; AE) van de aanbieder, € 259.4 in dit geval. Bij het beoordelen van het model middelen we de absolute fout van alle zorgaanbieder voor de gemiddelde absolute fout (MAE).

Naast de voorspelfout hebben we ook de gemiddelde kosten per traject (ongeveer €5.800, rode bolletje) en de individuele basisfout, het verschil tussen de daadwerkelijke kosten en de gemiddelde kosten. De absolute basisfout van de aanbieder berekenen we door de basisfout te middelen de absolute waarde te nemen, in dit geval €5228.4. De verklaarde fout is in dit fictieve voorbeeld ongeveer 99%, met echte gegevens ligt dit percentage voor vrijwel alle aanbieder vele malen lager.

### 4.3.1 Gemiddelde absolute fout (MAE)

Figuur 92: Absolute fout per aanbieder en MAE obv zvt en dsm



We hebben voor elk van de zorgaanbieders de absolute fout per patiënt berekend op basis van het zorgvraagtype, zoals hierboven beschreven. De individuele zorgaanbieders zijn in het figuur te zien als puntenwolk boven 'Absolute fout (aanb.)'. Het grootste deel van de aanbieders heeft voorspelfouten per patiënt van rond de €2.500, al komen ook waardes van €10.000 voor.

We berekenen de gemiddelde absolute fout door het gewogen gemiddelde te nemen van de absolute fout, gewogen voor het aantal zorgtrajecten van de zorgaanbieder dat in de dataset zit (€2373 grote groene bol). Voor de vergelijking met een vroeger bestaand alternatief hebben we ook de gemiddelde

absolute fout op basis van de diagnosehoofdgroep weergegeven (€2804 paarse bol). Deze vergelijking laten we voor de overzichtelijkheid alleen op totaalniveau zien.

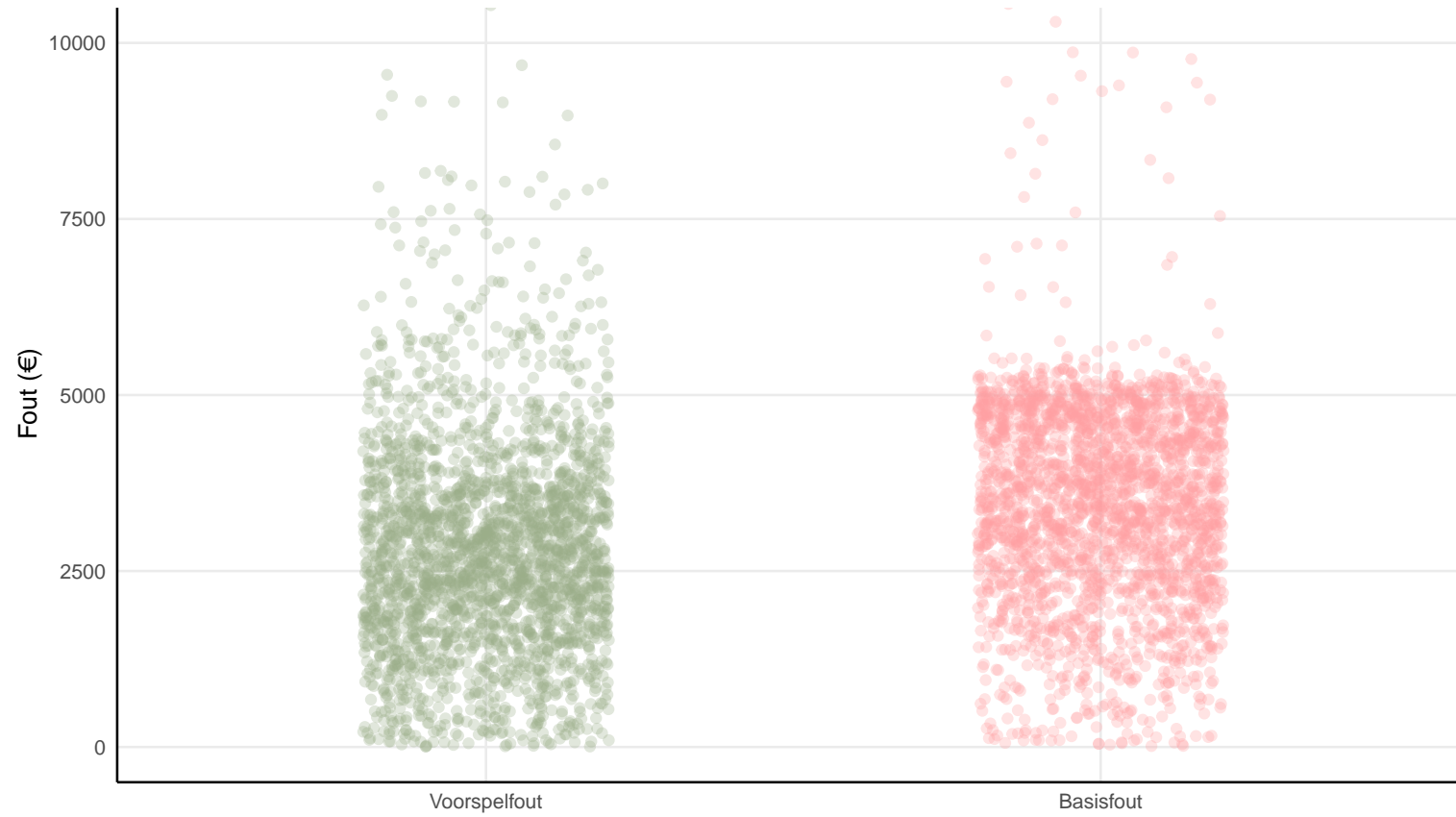
NB. Voor de vergelijkbaarheid tussen voorspelling op basis van dsm en hebben we voor dit figuur alle zorgaanbieders uitgesloten die geen dsm hebben geregistreerd.

#### 4.3.2 Verklaren basisfout

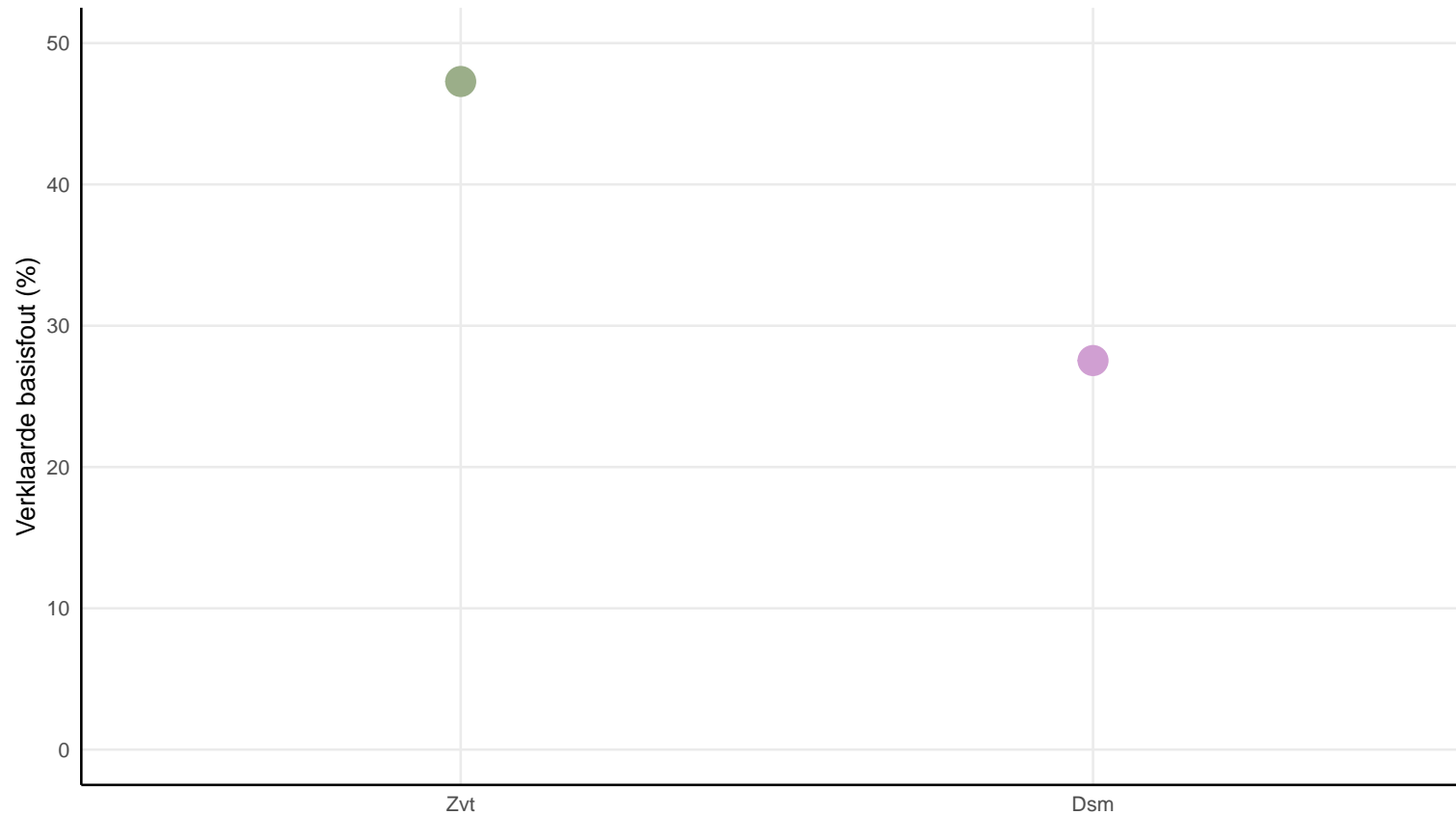
Het ook mogelijk te bekijken hoe groot de voorspelfout op basis van het model van zorgvraagtypering relatief ten opzichte van de basisfout (verschil tussen daadwerkelijke en gemiddelde kosten per patiënt). De technische uitwerking staat beschreven bij figuur 91, in de basis vergelijken we de som van de kwadraten van de voorspelfouten en de som van de kwadraten van de basisfouten. Op deze manier kunnen we zien hoeveel procent van de variatie die bestaat na 'voorspellen' op basis van de gemiddelde kosten wordt weggenomen door een voorspelling te doen op basis van het zorgvraagtype. Een perfect voorspelmodel heeft geen voorspelfout meer, dit betekent dat de gemiddelde voorspelling bij alle aanbieders exact even hoog is als de gemiddelde daadwerkelijke kosten. Dit zou een verklaarde basisfout van 100% zijn. Het is niet waarschijnlijk dat enig model dit in de zorg voor elkaar kan krijgen.

**Verdeling basis- en voorspelfout** In figuur 93 is de verdeling van voorspelfouten en basisfouten te zien. De basisfout is hoeveel het model nog kan verbeteren. Het zwaartepunt van de verdeling ligt voor de basisfout aanzienlijk hoger dan het zwaartepunt van de voorspelfout.

Figuur 93: Basisfout en absolute voorspelfout per aanbieder



Figuur 94: Aandeel van de basisfout die verklaard wordt door zvt of dsm



**Verklaren basisfout** Waar de voorspellingen met de diagnosehoofdgroep ongeveer 28% van de basisfout verklaren is dit ongeveer 47% voor voorspellingen op basis van het zorgvraagtype. Ook hier geldt dat we de vergelijking tussen zorgvraagtype en dsm alleen op totaalniveau laten zien.

NB deze percentages zijn anders dan eerder gedeeld. Dit is vooral het gevolg van wegen van de berekening op basis van het aantal patiënten dat bij een zorgaanbieder behandeld wordt, zoals geadviseerd door de werkgroep data. Aanvullende verschuiving kan het gevolg zijn van het gebruik van nieuwe en uitgebreidere bron-declaratiegegevens.

Tabel 19: Aantal per type aanbieder

Type aanbieder	Aantal
Vrijgevestigd psycholoog (94..)	2191
Vrijgevestigd psychiater (03..)	258
ZBC (22..)	272
Kleine instelling (73..)	34
Geïntegreerde instelling (0629..)	45
PAAZ/PUK (54..)	32
Overig	25
Totaal	2857



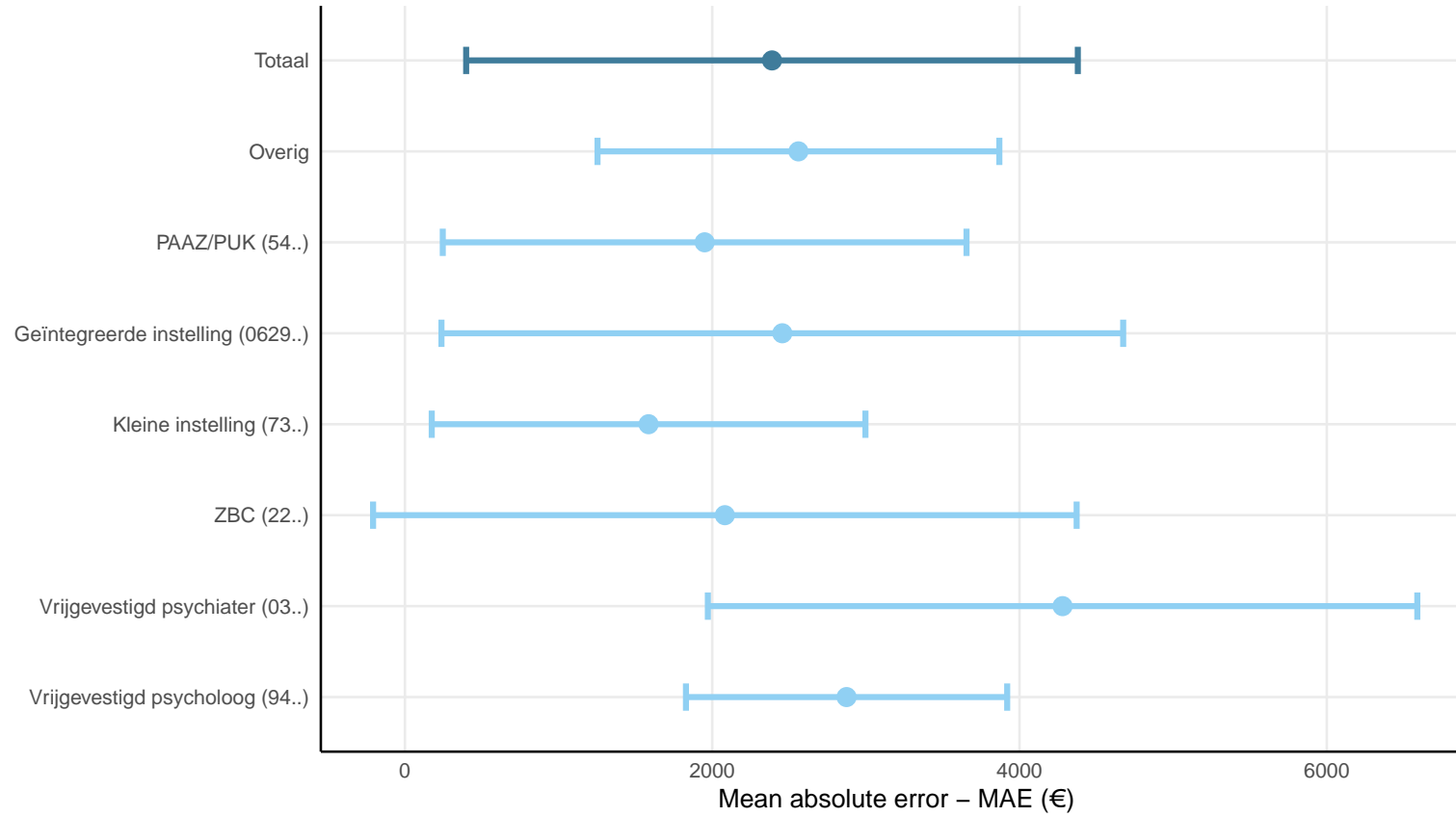
### 4.3.3 MAE per type aanbieder

We hebben gezien dat er flinke verschillen kunnen zijn van de absolute voorspelfout tussen de aanbieders. Daarom bekijken we of er ook verschillen zijn tussen een aantal type aanbieders. De aanbieders definiëren we op basis van de eerste cijfers van de AGB-code. Op deze manier onderscheiden we zes type aanbieders en een kleine groep 'overig'. In tabel 19 is te zien welke groepen we onderscheiden en hoeveel van deze aanbieders aan de omzetgrenzen voldoen. Tussen haakjes staat bij het type aanbieder welke begincijfers de AGB heeft om in de groep te vallen.

Er zit wat spreiding in de gemiddelde absolute fout van de verschillende zorgaanbieders. Vooral de hoge MAE voor vrijgevestigde psychiaters valt op. Dit lijkt vooral het gevolg van gemiddeld hoge voorspellingen en relatief lage kosten. De voorspellingen kunnen mogelijk worden verbeterd door, naast zorgvraagtype, ook de setting mee te nemen in de voorspelling, maar het zou ook een complicerende factor geven in de analyses en is daarom niet uitgevoerd.

NB. Voor deze analyse zijn de gegevens van aanbieders die geen diagnosehoofdgroep hebben geregistreerd wel meegenomen, daarom is er een kleine verschuiving van de MAE van het totaal ten opzichte van figuur 92.

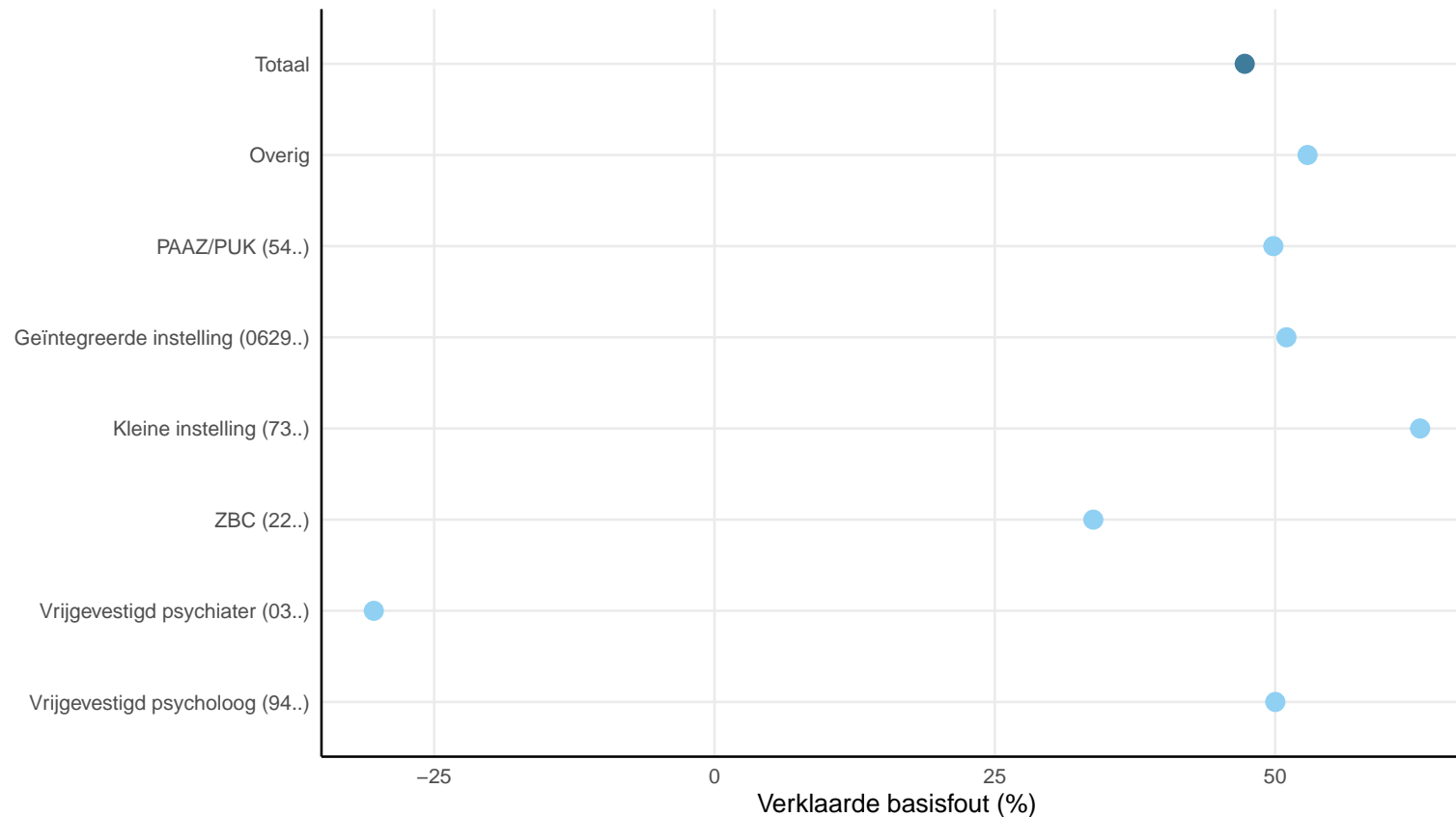
Figuur 95: MAE per type aanbieder



#### 4.3.4 Verklaarde basisfout per type aanbieder

Er zijn ook verschillen tussen de type aanbieders op het gebied van de hoeveelheid verklaarde basisfout. Voor veel aanbiedertypes wordt 20-30% van de basisfout verklaard wanneer een voorspelling wordt gedaan op basis van het zorgvraagtype. Voor de vrijgevestigde psychiaters is er een negatieve verklaring van de basisfout. Dit betekent dat de voorspelling op basis van het zorgvraagtype een grotere fout heeft dan de voorspelling op basis van het gemiddelde van de sector. Dit lijkt te komen doordat bij deze aanbieders gemiddeld relatief complexe zorgvraagtyperingen voorkomen en daar relatief lage kosten tegenover staan.

Figuur 96: Verklaarde basisfout



## 5 Vergelijking met evaluatie van het zorgclustermodel

In de evaluatie van de pilots met het zorgclustermodel zijn berekeningen uitgevoerd om in kaart te brengen hoe goed het model in die gegevens functioneerde. De adviescommissie heeft gevraagd eenzelfde set berekeningen uit te voeren om de declaraties uit het zorgprestatie-model te vergelijken met de gegevens uit de pilots. Hierbij moet rekening worden gehouden met de manier van rekenen in de evaluatie. De evaluatie verklaarde de variantie op het niveau van het individu, dit is anders dan de manier waarop we gezamenlijk hebben afgesproken dat we de zorgvraagtypering in de praktijk willen toepassen, namelijk voor groepen patiënten. Verder zijn in de evaluatie van het zorgclustermodel de trajecten afgekapt vanaf het moment dat hertypering/evaluatie van het zorgcluster plaats had moeten vinden en zijn de kosten per dag berekend door de som van alle kosten in een zorgtraject met een zorgvraagtype te nemen en deze te delen door het aantal dagen tussen de eerste zorgactiviteit en de laatste zorgactiviteit met het zorgcluster. Als laatste zijn trajecten met een looptijd van minder dan 7 dagen verwijderd. Om vergelijkbare gegevens te verkrijgen hebben we dezelfde verwerkingen uitgevoerd op de declaratiegegevens. De schattingen worden in alle gevallen gedaan met lineaire modellen zonder interactietermen.

Uit de evaluatie van het zorgclustermodel blijkt in die set gegevens 18,5% van de variantie van de kosten per dag verklaard te kunnen worden met het zorgcluster en dat dit 6% is met de diagnosehoofdgroep. Uit de declaratiegegevens komen iets lagere, maar vergelijkbare percentages naar voren. Specifiek blijkt dat het zorgvraagtype 17.6% van de variantie verklaart en dat dit 6.2% is voor de diagnosehoofdgroep. Als we een model maken dat zowel zorgvraagtype als diagnosehoofdgroep gebruikt als voorspellers, dan is 18.7% van de variantie te verklaren. Het toevoegen van de diagnosehoofdgroep aan het zorgvraagtype zorgt dus niet voor een grote verbetering van de verklaarde variantie, terwijl het wel een vergroting geeft van het aantal categorieën waarover afspraken gemaakt kan/moet worden.

NB de percentages zijn iets verschoven sinds de vorige versie van dit document door inclusie van meer gegevens.