

Auteursrechterlijke overeenkomst

Opdat de Universiteit Hasselt uw eindverhandeling wereldwijd kan reproduceren, vertalen en distribueren is uw akkoord voor deze overeenkomst noodzakelijk. Gelieve de tijd te nemen om deze overeenkomst door te nemen, de gevraagde informatie in te vullen (en de overeenkomst te ondertekenen en af te geven).

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling met

Titel: Vergelijken van fuzzy preferentie relaties

Richting: 2de masterjaar in de toegepaste economische wetenschappen: handelsingenieur in de beleidsinformatica
Jaar: 2009

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Ik ga akkoord,

VANHOENSHOVEN, Frank

Datum: 14.12.2009

Vergelijken van fuzzy preferentie relaties

Frank Vanhoenshoven

promotor :
Prof. dr. Koen VANHOOF

Woord vooraf

Dit onderzoek naar aggregatiemethoden en de robuustheid ervan, toegepast op een dataset betreffende de Europese verkeersveiligheid, is gevoerd in het kader van een masterproef aan de Universiteit Hasselt.

Ik hoop met dit onderzoek naar verkeersveiligheid een tastbare bijdrage te leveren aan de samenleving, hoe minimaal die bijdrage ook mag zijn. Naar ons gevoel onderscheidde deze kans op tastbare resultaten dit onderzoek van de klassieke onderwerpen in het kader van een masterproef, die meestal vrij descriptief van aard zijn. Bovendien was er het vooruitzicht op het creëren van een oplossing die een bredere toepassing kan hebben dan de onderzoekssteekproef. De resultaten van dit onderzoek hoeven zich dus niet strikt te beperken tot het gebied van Europese verkeersveiligheid.

Via deze weg, zou ik graag iedereen bedanken die een bijdrage geleverd heeft bij het tot stand komen van de masterproef. In de eerste plaats denk ik hierbij aan mijn begeleidster, dr. Elke Hermans, en promotor Prof. Dr. Koen Vanhoof, die bereid waren hun tijd en energie te steken in het steunen van deze eindverhandeling en het verstrekken van de benodigde informatie. Ook de heer Yongjun Shen ben ik zeer dankbaar voor het beschikbaar stellen van details uit vorig onderzoek. Ten slotte wil ik studiegenoten en familie bedanken voor de nodige tips en ondersteuning.

Samenvatting

De Europese verkeerveiligheid is allesbehalve een onontgonnen onderzoeksgebied. Vorig onderzoek heeft uitgewezen welke de voornaamste risicofactoren van verkeersveiligheid zijn: alcohol en drugs, snelheid, gebruik van de autogordel, het gebruik van de motorvoertuigverlichting overdag, voertuig, weginfrastructuur, en trauma management. Vervolgens heeft men passende indicatoren, Safety Performance Indicators (SPI), opgesteld om deze elementen van verkeersveiligheid meetbaar te maken. Voor een dataset van 20 Europese landen, werd er voor elke SPI een bepaalde score verzameld.

Uit deze gegevens over Safety Performance Indicators kunnen voor beleidsmakers twee belangrijke vormen van informatie afgeleid worden. Ten eerste kan het interessant zijn om te weten welke SPI de grootste invloed heeft op de verkeersveiligheid. Ten tweede kan er interesse zijn in een bepaalde landenrangschikking op basis van verkeersveiligheid. Door gewichten toe te kennen aan de indicatoren kunnen beide doelstellingen verwezenlijkt worden. De SPI met het hoogste gewicht, kan aangeduid worden als de indicator met de meeste invloed, terwijl de gewichten in combinatie met de scores op de indicatoren leidt tot een Road Safety Score per land, op basis waarvan een rangschikking opgemaakt kan worden.

Er bestaan verschillende manieren om gewichten toe te wijzen aan de indicatoren, waarvan er enkele reeds toegepast zijn op de Europese verkeersveiligheidsdataset. Al deze wegingmethodes hebben echter voor- en nadelen, waardoor er geen duidelijke consensus bestaat welke methode te verkiezen valt boven de anderen. Een betrouwbare methode om al deze wegingmethodes te aggregeren, kan dus nuttig zijn. Aangezien aggregatiemethodes gewichten toekennen aan wegingmethodes, kan een aggregatiemethode beschouwd worden als een wegingmethode op een hoger niveau.

Deze eindverhandeling kan opgedeeld worden in 2 delen. In het eerste deel presenteren we een aantal bruikbare aggregatiemethodes. Het tweede deel zal zich concentreren op het onderzoeken van de robuustheid van deze aggregatiemethodes.

De voorgestelde aggregatiemethodes (negen in totaal) kunnen onderverdeeld worden in vier categorieën. Ten eerste kan er een eenvoudige techniek van 'equal weighting' toegepast worden waarbij elke wegingmethode eenzelfde gewicht toegewezen krijgt. In

de praktijk komt dit erop neer dat er met gemiddeldes gewerkt kan worden. We onderscheiden hier het wiskundig en het meetkundig gemiddelde. De 2de categorie omvat de deviation methods: Absolute Deviation Method en Standard Deviation Method. Deze methodes kennen gewichten toe op basis van variatie in de gegevens. Ten derde presenteren we twee aggregatiemethodes die trachten het verschil tussen de consensusranglijst en de rangschikking op basis van de road safety scores te minimaliseren. De consensusranglijst is de landenrangschikking, gebaseerd op het aantal verkeersdoden per miljoen inwoners. De methode MinScore minimaliseert het verschil tussen de scores op de consensusranglijst en de road safety scores. De methode MinRank minimaliseert het verschil in positie op de consensusranglijst en de ranglijst op basis van de road safety score. Ten slotte beschrijven we de methodes die gewichten toekennen op basis van correlaties. Hoe groter de correlatie tussen de consensusranglijst en de ranglijst op basis van de road safety scores, hoe groter het gewicht dat toegewezen wordt. Onder deze methodes verstaan we: correlaties tussen rangschikkingen, correlaties tussen relatieve posities en Classical Preference Structures.

De minimalisatieproblemen maken optimaal gebruik van alle informatie die in de dataset vervat zit en zijn dus ook potentieel het meest kwetsbaar voor wijzigingen in die data. We zullen ons dan ook vooral op MinScore en MinRank concentreren bij het onderzoeken van de robuustheid.

De robuustheid wordt onderzocht door achtereenvolgens wijzigingen aan te brengen in de consensusranglijst, de selectie van landen en het detailniveau van de data. Vervolgens gaan we na in welke mate de output van de methodes verschilt met de output uit de oorspronkelijke, ongewijzigde, situatie. Deze verschillen kwantificeren we met behulp van 3 vaak gebruikte correlatiecoëfficiënten.

Over het algemeen blijken de minimalisatiemethoden vrij robuust te zijn. Vooral de methode MinRank laat hoge correlatiecoëfficiënten optekenen. De methode MinScore blijkt dan weer gevoelig te zijn voor zowel de keuze van de consensusranglijst als het detailniveau van de data. Wanneer er met een selectie van landen gewerkt wordt, is MinScore robuuster dan MinRank. We kunnen concluderen dat de methode MinRank relatief ongevoelig is voor de keuze van de consensusranglijst en het detailniveau van de data. Dit is uiteraard een belangrijk pluspunt in een onderzoeksomgeving waarin er vaak met benaderingen en schattingen van de realiteit gewerkt wordt. De gevolgen van

eventuele foute schattingen zijn minimaal bij een methode die robuuste resultaten produceert.

Inhoud

Woord vooraf	i
Samenvatting	ii
Inhoud	v
0 Inleiding.....	- 1 -
0.1 Probleemstelling	- 1 -
0.2 Kadering van het onderzoek	- 4 -
0.2.1 Voorgaand onderzoek	- 4 -
0.2.2 Overzicht van het onderzoek	- 5 -
0.2.3 Begrippen	- 6 -
0.3 Onderzoeksvragen	- 8 -
0.3.1 Centrale Onderzoeksvraag (COV).....	- 8 -
0.3.2 Deelvragen	- 9 -
1 Gebruikte informatie.....	- 11 -
1.1 Data	- 11 -
1.2 Analyse van de data	- 13 -
1.3 Wegingmethodes	- 17 -
1.3.1 Factor Analysis (FA)	- 18 -
1.3.2 Analytical Hierarchy Process (AHP)	- 18 -
1.3.3 Budget Allocation (BA)	- 19 -
1.3.4 Data Envelopment Analysis (DEA).....	- 19 -
1.3.5 Equal Weighting (EW)	- 19 -
2 Additionele wegingmethodes	- 20 -
2.1 Wegingmethodes zonder consensusranglijst	- 21 -
2.1.1 Absolute Deviation Method (ADM).....	- 21 -
2.1.2 Standard Deviation Method (SDM).....	- 21 -
2.2 Wegingmethodes met consensusranglijst	- 21 -
2.2.1 Classical Preference Structures (CPS).....	- 21 -
2.2.2 Fuzzy Preference Structures (FPS).....	- 22 -
2.2.3 Adapted Data Envelopment Analysis (DEA+)	- 22 -
2.3 Resulterende gewichten voor de verschillende wegingmethodes	- 23 -
3 Aggregatiemethoden	- 25 -
3.1 Beschrijving van de aggregatiemethoden.....	- 27 -
3.1.1 Equal Weighting	- 27 -
3.1.2 Deviation Methods.....	- 33 -
3.1.3 Minimization problems	- 38 -
3.1.4 Correlation-based aggregation	- 44 -
3.2 Algemene bespreking.....	- 53 -
3.3 Andere mogelijke aggregatiemethodes	- 56 -
3.4 Conclusies.....	- 57 -

4	Robuustheid	- 58 -
4.1	Waarom testen op robuustheid?	- 58 -
4.2	Werkwijze.....	- 58 -
4.3	Maatstaven van robuustheid	- 60 -
5	Testen van de robuustheid van de aggregatiemethodes.....	- 62 -
5.1	Robuustheid bij verandering van consensusranglijst.....	- 62 -
5.1.1	Aantal verkeersdoden per miljoen afgelegde kilometers	- 64 -
5.1.2	Aantal verkeersdoden per 100.000 personenwagens	- 66 -
5.1.3	Conclusie	- 67 -
5.2	Robuustheid bij inkrimping van het aantal landen in de dataset	- 68 -
5.2.1	Werkwijze.....	- 69 -
5.2.2	Robuustheid	- 69 -
5.2.3	Conclusie	- 70 -
5.3	Robuustheid bij vermindering van het detailniveau in de dataset	- 71 -
5.3.1	Werkwijze.....	- 72 -
5.3.2	Robuustheid	- 73 -
5.3.3	Conclusie	- 74 -
6	Conclusies.....	- 75 -
6.1	Verder onderzoek.....	- 77 -
7	Lijst geraadpleegde werken	- 79 -
8	Bijlagen	- 81 -
8.1	Kadering van het onderzoek	- 81 -
8.2	Gebruikte informatie.....	- 82 -
8.2.1	Data	- 82 -
8.2.2	Analyse van de data	- 84 -
8.3	Aggregatiemethodes: Landenrangschikkingen	- 91 -
8.4	Resultaten van de wijzigingen.....	- 97 -
8.4.1	MinRank & MinScore: gewichten.....	- 97 -
8.4.2	MinRank & MinScore: Rangschikkingen.....	- 98 -
8.4.3	Overige aggregatiemethodes: Gewichten	- 99 -
8.5	Robuustheid	- 101 -
8.5.1	MinScore & MinRank	- 101 -
8.5.2	Overige aggregatiemethodes (enkel correlaties tussen gewichtensets)	- 103 -

0 Inleiding

0.1 Probleemstelling

Beleidsvormers hebben een voorkeur voor kwantificeerbare gegevens om hun beleid op te baseren. Cijfermateriaal kan gebruikt worden om bepaalde beslissingen te objectiveren en te verantwoorden. Zelfs wanneer dat cijfermateriaal niet verkregen is door een bepaalde, objectieve meting, kan er nog bijzonder veel belang aan gehecht worden. Denken we bijvoorbeeld maar aan de rapportcijfers op school, een score gebaseerd op een set van steekproeven over kennis en dus geen absoluut 'kenniscijfer'. Of aan de scores bij vele sporten, inclusief strafpuntensystemen en verschillende juryleden. Voorbeelden van cijfers die wel verkregen worden via objectieve metingen, zijn werkloosheidscijfers, de grootte van de overheidsuitgaven in de gezondheidszorg, het aantal geboortes per duizend inwoners...

Verkeersveiligheid is een begrip dat moeilijk in kwantitatieve en objectieve termen uitgedrukt kan worden. Er bestaat immers geen universeel aanvaarde maat voor verkeersveiligheid. We kunnen niet spreken over liters, meters, inches of graden veiligheid. Wanneer we een dergelijk begrip willen kwantificeren, moeten we onze toevlucht zoeken in andere methodes.

In het domein van Performance Management wordt gebruik gemaakt van de Balanced Scorecard om het begrip performance te meten (Kaplan & Norton, 2002). We kunnen het principe van de Balanced Scorecard op twee manieren gebruiken als onderbouw voor het meten van verkeersveiligheid. Een eerste methode is om op zoek te gaan naar uitingen van verkeersveiligheid. We zoeken dus bepaalde, meetbare, indicatoren die het resultaat zijn van verkeersonveiligheid. Vervolgens gaan we er vanuit dat de verkeersveiligheid weergegeven wordt door de prestaties op de gekozen indicator. Zo kan men bijvoorbeeld het aantal verkeersdoden per miljoen inwoners als indicator gebruiken voor de verkeersveiligheid in een land. (OECD, 2009)

Een tweede methode die we kunnen gebruiken, vertrekt niet vanuit de resultaten of uitingen van verkeersveiligheid, maar vanuit de oorzaken. In een eerste fase, worden een aantal onderliggende variabelen geïdentificeerd die de verkeersveiligheid in een land bepalen. We beschouwen verkeersveiligheid dus als een functie van deze variabelen. Zo

is bijvoorbeeld het gebruik van veiligheidssystemen, zoals de autogordel, een variabele die de verkeersveiligheid kan verhogen. (Hakkert, Gitelman, & Vis, 2007) Eenmaal deze variabelen geïdentificeerd zijn, kan er op zoek gegaan worden naar uitingen van deze variabelen. Zo is het percentage inzittenden dat gebruik maakt van de autogordel, een indicator voor het gebruik van veiligheidssystemen. De volgende stap bestaat erin de invloed van al deze indicatoren op de verkeersveiligheid in een land te kwantificeren. Dit kan door op een bepaalde manier gewichten toe te kennen aan de variabelen. Je maakt hier dus een onderscheid in het effect van de verklarende variabelen op de verkeersveiligheid. In een laatste fase wordt elke indicator vermenigvuldigd met zijn toegekende gewicht. De geaggregeerde score (bijvoorbeeld de som van deze gewogen indicatoren) kan dan gebruikt worden als een waardemeter voor verkeersveiligheid. Deze combinatie van indicatoren wordt ook wel verkeersveiligheidsprestatie-index genoemd. (Hermans, Van den Bossche, & Wets, 2008a)

Zowel via de oorzaken als via de gevolgen kan met er dus in slagen om verkeersveiligheid te kwantificeren. Er rest ons dus enkel nog de vraag waarom dat kwantificeren zo belangrijk is.

Het grote voordeel van kwantificeerbare gegevens is dat je er een hele waaier aan wiskundige bewerkingen op kunt uitvoeren. Zo kun je er allerlei statistische gegevens uit afleiden, gemiddeldes, mediaanwaarden, maxima, minima, varianties... Bij deze statistische gegevens moet echter onmiddellijk een kanttekening geplaatst worden. Het is niet zo dat we alle cijfers in de categorie 'kwantitatieve gegevens' onderbrengen. Sommige cijfers zijn eerder kwalitatief dan kwantitatief van aard, wat wil zeggen dat het eigenlijk betekenisloos is om er statistische bewerkingen op uit te voeren (van der Zee, 2004). Voorbeelden van cijfers die kwalitatief van aard zijn, zijn bijvoorbeeld identificatienummers of cijfers als attributieve gegevens (binaire gegevens om iemands geslacht weer te geven of een cijfercode om bijvoorbeeld beroepscategorieën aan te duiden). Het spreekt voor zich dat het zinloos is om bijvoorbeeld een gemiddeld identificatienummer of geslacht te berekenen, of om bijvoorbeeld de variantie van de beroepscategorieën te berekenen. Hoewel het wiskundig perfect mogelijk is om statistische gegevens uit een getallenreeks te berekenen, hebben deze cijfers in de bovengenoemde gevallen geen enkele betekenis in de praktijk.

Behalve de tegenstelling kwantitatieve versus kwalitatieve gegevens, halen we ook het verschil aan tussen gegevens die verkregen zijn via objectieve metingen en deze verkregen via subjectieve metingen.

Heli Kirvesoja citeert enkele onderzoeken naar objectieve en subjectieve metingen bij het peilen naar feedback van gebruikers van producten. Kirvesoja citeert McLelland (1995) die stelt dat bij objectieve metingen, de waarde direct afgelezen kan worden (Kirvesoja, 2001). De waarde staat dan ook niet ter discussie en wordt in principe niet beïnvloed door de gebruikte meetmethode. Indien men bij twee objectieve metingen van dezelfde waarde significante verschillen vaststelt, is dat een teken dat één van de gebruikte methoden foutief is.

Bij subjectieve metingen ligt dat anders. Het gaat hier over gegevens die niet objectief vastgesteld kunnen worden. We kunnen deze gegevens enkel schatten door te zoeken naar domeinen waarop deze 'waarden' tot uiting komen. In Performance Management maakt men gebruik van kritische succesfactoren om prestatieniveaus in te kunnen schatten. (Kaplan & Norton, 2002) De waarden die men vindt voor subjectieve metingen zijn juist wel erg onderhevig aan de gekozen meetmethode. Het kiezen van de juiste meetmethode is dus van cruciaal belang voor het bekomen van de juiste, 'reële' waarden van de gemeten factoren. Wanneer er beslissingen genomen worden op basis van subjectieve gegevens, hangt de kwaliteit van de beslissing in sterke mate samen met de kwaliteit van de gekozen meetmethode.

Juist omdat we de verkeersveiligheid meten met behulp van subjectieve metingen, is het belangrijk dat we een beeld krijgen van de betrouwbaarheid van de gebruikte meetmethodes. Zoals al gesteld, hebben deze een invloed op het meetresultaat. Het doel van deze eindverhandeling is om na te gaan hoe robuust, of hoe betrouwbaar, methodes om verkeersveiligheid te meten, zijn.

0.2 Kadering van het onderzoek

0.2.1 Voorgaand onderzoek

In het kader van verkeersveiligheid heeft IMOB verbonden aan de U Hasselt, de prestaties op vlak van verkeersveiligheid voor 21 Europese landen in één score proberen te integreren. Deze score wordt door Hermans, Van den Bossche en Wets Safety Performance Index genoemd. (Hermans, Van den Bossche, & Wets, 2008a) We beschrijven kort het tot stand komen hiervan.

Er wordt vertrokken van zeven Safety Performance Indicators (SPI), dit zijn indicatoren waarvan we mogen aannemen dat ze een significante invloed uitoefenen op het aantal verkeersdoden in een bepaald land (Hakkert, Gitelman, & Vis, 2007). We gaan er dus vanuit dat de prestatie van een Europees land op vlak van verkeersveiligheid bepaald wordt door zijn prestaties op de zeven Safety Performance Indicators.

Het verzamelen van data is niet altijd even eenvoudig. De gewenste gegevens zijn niet altijd beschikbaar, betrouwbaar en/of vergelijkbaar. Om deze problemen op te vangen, veronderstellen we in eerste instantie dat een domein van verkeersveiligheid vervat wordt in een SPI. Zo is de kwaliteit van het wagenpark uitgedrukt in het percentage voertuigen dat jonger is dan 6 jaar. Een hoger aandeel van recente wagens komt de verkeersveiligheid ten goede. Voor een set van 20 Europese landen – de dataset van 21 met uitzondering van Zwitserland – worden de prestaties op elke SPI verzameld.

Op deze data worden vervolgens een aantal technieken toegepast om tot één totale score – een maatstaf voor verkeersveiligheid – per land te komen. Dit houdt in dat men elke SPI een bepaald gewicht toebedeeld. De SPI's die belangrijker zijn in termen van verkeersveiligheid krijgen een groter gewicht dan de minder belangrijke. (Hermans, Van den Bossche, & Wets, 2008a). Aan de hand van deze gewichten kunnen beleidsmakers afleiden in welke SPI best geïnvesteerd wordt om de verkeersveiligheid te verhogen.

De methodes die gebruikt worden om gewichten te bepalen voor de SPI's kunnen onderverdeeld worden in twee categorieën. De eerste categorie maakt geen gebruik van het aantal verkeersdoden per land, maar enkel van de informatie die we verkrijgen uit de prestatie-indicatoren. De gewichten worden toegewezen met behulp van de volgende

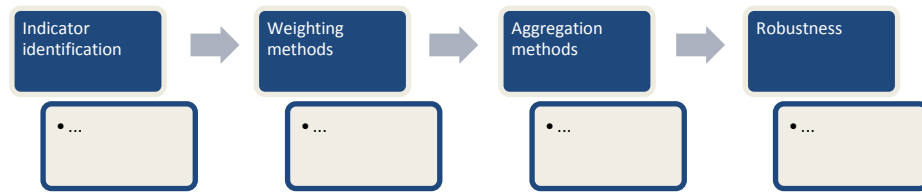
methodes: Factor Analysis, Analytic Hierarchy Process, Budget Allocation, Data Envelopment Analysis of Equal Weighting (Hermans, Van den Bossche, & Wets, 2008a). De andere categorie maakt wel gebruik van een uiting van verkeersveiligheid, namelijk het aantal verkeersdoden per miljoen inwoners. (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008) Deze kan dan gebruikt worden om de gewichten te 'modellieren' in een poging om tot zo realistisch mogelijke resultaten te komen.

Zoals Hermans, Van den Bossche & Wets al aangeven in hun artikel Impact of Methodological Choices on Road Safety Ranking (2008b), beïnvloedt de gekozen methodologie de grootte van de toegewezen gewichten in zekere mate. Zij halen dan ook terecht aan dat er zeer voorzichtig met de gewichten moet worden omgesprongen. Onzekerheids- en sensitiviteitsanalyses zijn voor hen belangrijke instrumenten om de sensitiviteit van de gewichten na te gaan.

Het onderzoek dat in het kader van deze eindverhandeling zal uitgevoerd worden, heeft tot doel de betrouwbaarheid van de methodes die men gebruikt om gewichten te bepalen, vast te kunnen stellen. Het is de bedoeling dat we ons een goed beeld kunnen vormen van de prestaties van de verschillende methodes.

0.2.2 Overzicht van het onderzoek

Ter verduidelijking zullen we een schematisch overzicht van het onderzoek geven en hoe het kadert in voorgaand werk. Onder het schema in Figuur 1 wordt puntsgewijs een toelichting gegeven. Figuur 1 zal naargelang het onderzoek vordert verder aangevuld worden. Op die manier wordt het eenvoudiger om het overzicht te bewaren.



Figuur 1: Overzicht van het onderzoek

- 1) Identificatie indicatoren: Dit is het verkennend onderzoek waarin relevante indicatoren geïdentificeerd worden. Ook het verzamelen van de nodige gegevens brengen we in deze stap onder. Sectie 1.1 bevat meer informatie over deze stap.
- 2) Weging van de indicatoren: Er wordt aan de verschillende indicatoren uit stap 1 een bepaald gewicht toegewezen. Meer informatie over de wegingmethodes die we gebruikt hebben, staat in sectie 1.3.
Het eerste deel van het onderzoek begint in deze stap. We hebben zelf een aantal wegingmethodes toegevoegd aan deze die we al ter beschikking hadden (Hoofdstuk 2).
- 3) Aggregatie van de wegingmethodes: Stap 3 omvat het aggregeren van de gewichten die door elk van de verschillende wegingmethodes toegekend wordt, in 1 gewicht per indicator. Een aantal aggregatiemethodes worden beschreven in hoofdstuk 3.
- 4) Robuustheid: Als laatste stap zullen we de robuustheid van de aggregatiemethodes testen. Dit wordt beschreven in hoofdstuk 4 (theorie) en hoofdstuk 5 (resultaten).

0.2.3 Begrippen

Vervolgens verklaren we enkele begrippen die in de loop van het onderzoek aan bod zullen komen.

- Safety performance indicators (SPIs): we gebruiken de definitie van het European Road Safety Observatory.

[SPIs] are measures (indicators), reflecting those operational conditions of the road traffic system, which influence the system's safety performance. Basic features of SPIs are their ability to measure unsafe operational conditions of the road traffic system and their independence from specific safety interventions. SPIs are aimed to serve as assisting tools in assessing the current safety conditions of a road traffic system, monitoring the progress, measuring impacts of various safety interventions, making comparisons, and for other purposes. (ERSO)
- Road Safety Score: de score die elk land toebedeeld krijgt en bestaat uit de SPIs en hun gewichten. (Hermans, Van den Bossche, & Wets, 2008a)
- Landenrangschikking: de rangschikking van landen die opgemaakt kan worden aan de hand van de Road Safety Scores.
- Consensusranglijst: een uiting van verkeersveiligheid, die we mogen beschouwen als zeer representatief en algemeen aanvaard voor verkeersveiligheid. De consensus is gebaseerd op het aantal verkeersdoden per land. (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008), (Cook, 2005) We zullen de consensusranglijst ook ideale representatie of kortweg consensus noemen.
- Wegingmethoden: de methodes die gebruikt worden om aan elke SPI een bepaald gewicht toe te kennen.
 - o Wegingmethodes die gebruik maken van een consensusranglijst: we zullen dit ook kortweg een wegingmethode met consensus noemen.
 - o Wegingmethodes die geen gebruik maken van een consensusranglijst: waarnaar we ook zullen verwijzen als een wegingmethode zonder consensus.
- Aggregatiemethodes: manieren om de verschillende wegingmethodes te combineren. Aggregatiemethodes kennen gewichten, en dus een mate van belangrijkheid, toe aan verschillende wegingmethodes en kunnen ook op zichzelf als wegingmethode bekeken worden.
 - o Geaggregeerde gewichten: het gewicht dat aan een indicator toegewezen wordt nadat de wegingmethodes met elkaar gecombineerd zijn.
- Robuustheid: de mate waarin een methode erin slaagt om stabiele resultaten af te leveren. Dat wil zeggen dat er geen grote schommelingen in de resultaten mogen zijn bij een lichte aanpassing in de dataset.

0.3 Onderzoeksvragen

0.3.1 Centrale Onderzoeksvraag (COV)

De centrale onderzoeksvraag waar deze eindeverhandeling een antwoord op wil bieden, is de volgende:

Hoe stabiel zijn de verschillende aggregatiemethodes voor gewichtsbepaling ten opzichte van de presentatie van de consensusranglijst?

De stabiliteit van de methodes verwijst naar de betrouwbaarheid van de gevonden geaggregeerde gewichten. We moeten er kunnen op vertrouwen dat de verkregen gewichten een juiste weergave van de werkelijkheid zijn. De gebruikte methode moet ook voor andere of licht aangepaste datasets dezelfde resultaten blijven geven. Het mag niet voorkomen dat de resultaten enkel voor de huidige dataset blijken te kloppen. De gewichten van de Safety Performance Indicators moeten ook voor andere jaren gelden. Indien dit niet zo is, kunnen we spreken van 'overfitting'. De uitkomsten zijn dan te afhankelijk van de gekozen dataset en kunnen niet veralgemeend worden (Witten & Frank, 1999). Anderzijds mag men ook niet gaan 'underfitten'. Bij underfitting wordt de beschikbare dataset niet genoeg benut om er juiste resultaten uit te halen (Witten & Frank, 1999). De gevonden gewichten bezitten een te grote toevalsfactor, wat eveneens vermeden dient te worden. Het gewicht van een SPI duidt namelijk op de impact die de SPI heeft op de verkeersveiligheid. Wanneer deze gewichten instabiel blijken te zijn, kan de gebruikte methode weinig of niets bijdragen tot het beslissingsproces in verkeersveiligheid. Goede methodes leveren dus stabiele gewichten die de invloed van elke SPI op de verkeersveiligheid correct kunnen voorspellen.

De ideale rangschikking verwijst naar het werkelijke niveau van verkeersveiligheid van een land. Omdat deze ideale rangschikking niet gekend is, misschien zelfs niet bestaat, moeten we teruggrijpen naar een bepaalde uiting van dit ideaal. Een dergelijke benadering van de ideale rangschikking wordt ook wel eens de consensusranglijst genoemd (Cook, 2005). De consensus is een uiting van de werkelijkheid die algemeen als representatief aanvaard wordt. De uiting die het algemeenst aanvaard kan worden voor de verkeersveiligheid in een land, is het aantal verkeersdoden. Deze gegevens

worden uitgedrukt per miljoen inwoners. Dit wordt gedaan om vergelijkingen mogelijk te maken tussen landen van verschillende grootte. (Hermans, Van den Bossche, & Wets, 2008a).

De juistheid van de gewichten kan worden uitgedrukt als mate van overeenkomst met de consensusranglijst. Aan de hand van de individuele scores op de SPI's en het gewicht van elke SPI, kan er voor elk land een Road Safety Score berekend worden. De overeenkomst tussen de ranglijst van landen op basis van de Road Safety Score en de consensusranglijst, geeft een beeld van de waarde van de gekozen gewichten. Wanneer de methode een correct gewicht toegewezen heeft aan elke SPI, zou de rangschikking van de landen op basis van de Safety Performance Index in sterke mate moeten overeenkomen met die van de consensusranglijst. We veronderstellen hierbij dat de gekozen SPI's alle aspecten van verkeersveiligheid bevatten. Het spreekt dan ook voor zich dat de gewichten niet significant mogen verschillen als er een andere, geldige, consensus gekozen wordt.

0.3.2 Deelvragen

- 1. Op welke manieren kunnen de verschillende gewichten geaggregeerd worden in één gewicht?*

Gewichten aggregeren (weight aggregation) houdt in dat men de gewichten op een bepaalde manier probeert te combineren zodat er betere resultaten uit volgen. Bij weight aggregation komt het er in feite op neer dat er gewichten toegekend worden aan de verschillende methodes. Beter presterende methodes moeten zwaarder doorwegen dan minder presterende methodes. Weight aggregation is dus een manier waarop men de stabiliteit van de gewichten probeert te vergroten. De zwakke punten van elke methode probeert men uit te schakelen door er andere methodes aan te koppelen. Ook in data mining wordt er op dezelfde manier over modellen gedacht. In de wetenschap dat geen enkel model perfect is, maar altijd wel ergens een zwakkere plek heeft, worden technieken als 'bagging' en 'boosting' gebruikt om tot betere resultaten te komen (Witten & Frank, 1999)

2. *Wat is de impact op de gekozen gewichten en de daaruit volgende landenrangschikking wanneer de overeenkomst met een andere consensus als criterium gebruikt wordt?*

Wanneer er meerdere betrouwbare consensusranglijsten zijn, zou de keuze voor een bepaalde consensus, de grootte, maar vooral de ordering, van de gekozen gewichten niet significant mogen beïnvloeden.

3. *Heeft het selecteren van een aantal landen uit de dataset een invloed op de gekozen gewichten en de daaruit volgende landenrangschikking?*

De gewichten die men toekent aan de verschillende SPI's mogen niet teveel verschillen naargelang de keuze van de dataset. Een goede manier om de stabiliteit van de gewichten te bepalen, is dan ook de dataset aan te passen. Als de gewichten erg variëren wanneer er met slechts een aantal landen uit de dataset gewerkt wordt, wil dat zeggen dat de gekozen dataset de uitkomst te erg beïnvloedt en de gebruikte methode niet robuust genoeg is.

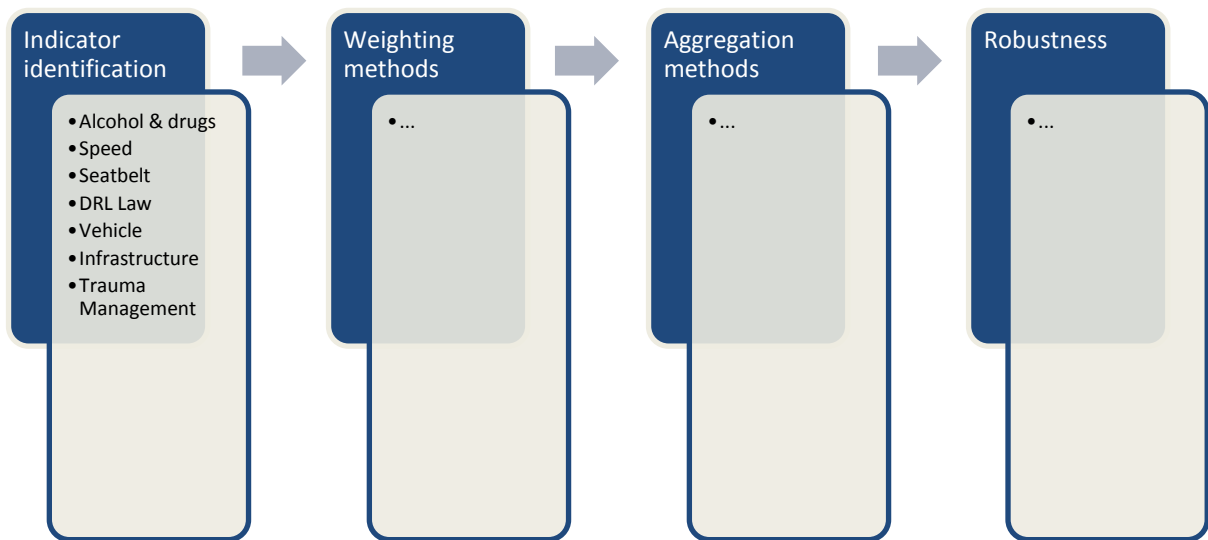
4. *Veranderen de gekozen gewichten en de daaruit volgende landenrangschikking significant wanneer men het detailniveau van de Safety Performance Indicators aanpast?*

Hoe betrouwbaar de gegevens ook mogen zijn, we verliezen best niet uit het oog dat de score van een aantal landen op een bepaalde SPI vertekend kunnen zijn. Deze vertekening kan optreden omwille van een aantal factoren zoals een verschil in definitie of meetmethode. (Vis & Van Gent, 2007a).

Een methode om lichte vertekeningen tegen te gaan, is het verkleinen van het detailniveau van de Safety Performance Indicators en zelfs de consensusranglijsten. In plaats van met de werkelijke cijfers te werken, delen we de landen in een aantal categorieën in. Deze methode wordt ook wel discretization genoemd (Clarke & Barton). Op die manier worden landen die gelijkaardige cijfers hebben voor een bepaalde indicator, als gelijken behandeld. We gaan ervan uit dat de toevalscomponent in dat geval in grote mate uitgeschakeld wordt.

1 Gebruikte informatie

In dit hoofdstuk zullen we de eerste fase, de identificatie van de relevante indicatoren, uit het onderzoeksopzet bespreken. Deze fase staat aangeduid in de aangevulde figuur 1.



Figuur 1 a: Overzicht van het onderzoek

1.1 Data

De data die we zullen gebruiken in dit onderzoek, is terug te vinden in bijlage en hebben allen betrekking op het jaar 2003. De volgende Europese landen maken deel uit van dit onderzoek: België, Cyprus, Denemarken, Duitsland, Estland, Finland, Frankrijk, Griekenland, Hongarije, Ierland, Italië, Nederland, Oostenrijk, Polen, Portugal, Spanje, Slovenië, Tsjechië, het Verenigd Koninkrijk en Zweden. In Tabel 1 in bijlage kunnen gegevens over elke Safety Performance Indicator per land gevonden worden. Er wordt een positieve relatie verwacht tussen elk van de zeven SPI's – alcohol en drugs, snelheid, beschermende systemen, de wetgeving op het gebruik van de motorvoertuigverlichting overdag, voertuig, weginfrastructuur en trauma management - en verkeersveiligheid.

- De kolom "Alcohol" geeft het percentage van de bestuurders weer dat aangeeft zich aan de BAC-limiet (Blood Alcohol Content) te houden. In de dataset blijkt Zweden op deze SPI het best presterende land te zijn. Maar liefst 100% van de

- bestuurders geeft aan zich aan de BAC limiet te houden. Cyprus sluit de rij met 78,2%, wat zeer laag is in vergelijking met het gemiddelde van 96,11%.
- In de kolom "Speed" staat het percentage van de bestuurders dat aangeeft zich aan de geldende snelheidsbeperking te houden. Ierland is met 96% het best presterende land in de dataset. In Estland geeft 80% van de bestuurders aan zich aan de geldende snelheidsbeperking te houden, wat op deze SPI meteen het laagste percentage uit de dataset is. De gemiddelde waarde is 86,95%.
 - "Seat belt" geeft het percentage bestuurders weer dat de autogordel draagt. In Frankrijk draagt het hoogste percentage bestuurders, nl. 97%, de autogordel. Griekenland doet met slechts 40% opmerkelijk slechter en is het minst presterende land wat deze SPI betreft. De gemiddelde waarde uit de dataset bedraagt 78,85%.
 - DRL Law staat voor Daytime Running Lights Law en is een codering voor de geldende wetgeving met betrekking tot het gebruik van motorvoertuigverlichting overdag. Hoe hoger de codering, hoe strenger de geldende wetgeving. Denemarken, Estland, Finland, Oostenrijk, Slovenië en Zweden krijgen de hoogste codering toegewezen. Cyprus, Griekenland, Ierland en het Verenigd Koninkrijk worden het laagst gecodeerd.
 - In "Vehicle" kan het percentage van de ingeschreven personenwagens dat jonger is dan 6 jaar teruggevonden worden. In de dataset is Cyprus het land met het laagste percentage ingeschreven personenvoertuigen dat jonger is dan 6 jaar. Al komen ook Estland en Tsjechië in de buurt van die lage 17,08%. Ierland heeft met 54,45% een opmerkelijk hoog percentage jongere personenwagens ingeschreven. De gemiddelde waarde voor deze SPI is 33,82%.
 - "Infrastructure" wordt gevat in de dichtheid van het wegennet. In deze kolom wordt de totale lengte van het wegennet in een land, ten opzichte van de totale oppervlakte van het land, weergegeven. In de landenrangschikking op basis van deze SPI, torent België hoog boven alle andere landen uit, met maar liefst 4,91 km bestrating per km² land. De gemiddelde waarde uit de dataset, 1,58 km per km², blijft daar ver onder. In Finland is 0,23 km bestrating per km² terug te vinden.
 - Onder de hoofding "Trauma Care" staat het percentage van het Bruto Binnenlands Product dat uitgegeven wordt aan gezondheidszorg. Duitsland geeft met 10,9% van het BBP, procentueel het meest uit aan gezondheidszorg. Estland is op deze

SPI het minst presterende land met 5,1%. Gemiddeld spenderen de landen in de dataset 8,15% van het BBP aan gezondheidszorg.

De bijgevoegde Tabel 3 geeft voor elk land het aantal dodelijke verkeersslachtoffers in 2003 weer. In de eerste kolom, staan de 20 landen alfabetisch gerangschikt. De kolommen 2, 3 en 4 geven respectievelijk het aantal verkeersdoden per miljoen inwoners, per miljoen afgelegde kilometers (door personenwagens en gemotoriseerde tweewielers) en per 100.000 ingeschreven personenwagens (OECD, 2009). Zweden komt in al deze ranglijsten naar voor als het veiligste land. In de ranglijst op basis van het aantal verkeersdoden per miljoen afgelegde kilometers moet het deze 1^{ste} plaats delen met het Verenigd Koninkrijk. Polen wordt in elk van de 3 rangschikkingen bestempeld als het onveiligste land.

Om vergelijkbaarheid tussen de indicatoren mogelijk te maken, worden alle gegevens genormaliseerd op een schaal van 0 tot 1. Dit gebeurt met de formule

$$N_i = \frac{X_i - \text{Min}(X_1, X_2 \dots X_n)}{\text{Max}(X_1, X_2 \dots X_n) - \text{Min}(X_1, X_2 \dots X_n)}$$

Met N_i de genormaliseerde waarde voor land i en X_i de werkelijke waarde. De genormaliseerde waarden kunnen teruggevonden worden in Tabel 2 en Tabel 4. In deze tabellen staat de waarde 1 telkens voor het best presterende land in de dataset. De waarde 0 geeft dan het minst presterende land weer. Alle andere landen bevinden zich op een schaal tussen deze twee uitersten in¹.

1.2 Analyse van de data

Een analyse van de data helpt ons om latere resultaten in het onderzoek beter te begrijpen. Het hele proces van het bepalen en aggregeren van gewichten, wordt volledig bepaald door de data die we beschikbaar hebben. Omdat de resultaten volledig afhankelijk zijn van de beschikbare dataset, kan een analyse van de data ons helpen om eventuele vertekeningen op voorhand op te sporen.

¹ Omdat een lager aantal verkeersdoden overeen komt met een betere prestatie op vlak van verkeersveiligheid, wordt er voor de ranglijsten op basis van het aantal verkeersdoden gewerkt met $1 - N_i$ als genormaliseerde waarde. Hierdoor komt een lager aantal verkeersdoden overeen met een hogere genormaliseerde waarde.

We werken met de gegevens en het framework dat gebruikt wordt in Hermans et al. 2008a. Een eerste veronderstelling die we gemaakt hebben, is, dat de data een correcte en waarheidsgetrouwe weergave zijn van de werkelijkheid. Deze veronderstelling is tweeledig. Enerzijds gaan we uit van de betrouwbaarheid van de metingen. Dit wil zeggen dat de gegevens waarmee we werken, de werkelijke waarden van de metingen zijn en dat deze metingen representatief zijn voor het hele land. Anderzijds gaan we er ook van uit dat de gebruikte metingen een goed beeld geven van de indicatoren. Concreet betekent dit niet alleen dat de score van België voor de indicator Vehicle, 40,84% (zie Tabel 1), ook effectief het werkelijke percentage is van het aantal voertuigen jonger dan 6 jaar in België. Maar ook dat die score van 40,84% een goede weergave geeft van de prestatie van België op vlak van voertuigen..

Incorrecte, niet-representatieve waarden zullen ertoe leiden dat de resultaten niet veralgemeend kunnen worden naar de werkelijkheid. Het onderzoek zou in dat geval immers gebaseerd zijn op onjuiste, misschien zelfs fictieve, waarden.

De tweede veronderstelling die we maken, bevindt zich een stapje hoger in het framework. In dit framework stelt men dat de verkeersveiligheid in een land, gemeten aan de hand van het aantal dodelijke slachtoffers per miljoen inwoners, voldoende bepaald wordt door zeven aspecten: rijden onder invloed (Alcohol), onaangepaste snelheid (Speed), gebruik van veiligheidssystemen (Seatbelt), zichtbaarheid (DRL Law), de kwaliteit van het wagenpark (Vehicle), de kwaliteit van het wegennet (Infrastructure) en aandacht aan nazorg. (Trauma Care) (Hakkert, Gitelman, & Vis, 2007).

Het eventuele ontbreken van een belangrijk aspect, kan eveneens voor vertekening zorgen. Het effect van die ontbrekende indicator zal door de methodes immers op een of andere manier aan een andere indicator of indicatoren toegewezen worden. Dit overschat het belang van de bevoordeelde indicator. Het overschatten van de ene indicator, samen met het ontbreken van een andere, zal weerspiegeld worden in de score van de landen. Deze score zal geen 'werkelijke' weergave zijn van de verkeersveiligheid in een land.

Een analyse van de data kan ons wat wijzer maken. We gaan op zoek naar landen die gekenmerkt worden door een hoog aantal verkeersdoden maar toch hoge scores behalen en dus goed presteren op de indicatoren op de indicatoren, of omgekeerd. Dergelijke

gevallen kunnen wijzen op eventuele fouten in de gegevens of in het framework. We kunnen in ieder geval stellen, dat de methodes problemen zullen hebben om voor zulke landen een juiste voorspelling te maken van de plaats in de rangschikking op basis van de road safety scores. Met een juiste voorspelling bedoelen we dan een rangschikking die ongeveer overeen komt met de consensusranglijst. Er zal dus een vorm van vertekening optreden. De precieze oorzaken van deze vertekening zijn echter zeer moeilijk te bepalen en vallen buiten het kader van het onderzoek.

Het is dus de bedoeling om grote verschillen tussen de prestaties op de indicatoren en de prestaties in termen van verkeersdoden bloot te leggen. In het vakgebied Kwaliteitstechnieken wordt er om de beheersbaarheid van productieprocessen te controleren, gebruik gemaakt van regelkaarten, ontwikkeld door W A Shewhart in 1931 (Banens, 1994). We passen een soortgelijke techniek toe. We zetten per land een grafiek uit waarop de rangschikking van dat land voor elke indicator geplot wordt. Hieraan kan afgelezen worden hoe een bepaald land ten opzichte van de andere landen presteert. Vervolgens tekenen we drie referentielijnen. De rangschikking op basis van het aantal verkeersdoden per miljoen inwoners, doet dienst als vergelijkingspunt (cfr. gemiddelde bij regelkaarten). Als boven- en ondergrens gebruiken we de rangschikking op basis van verkeersdoden ± 5 plaatsen (cfr. Upper Control Limit en Lower Control Limit bij regelkaarten). Het idee hierachter is, dat de consensusrangschikking voor verkeersveiligheid niet teveel mag verschillen van de rangschikkingen op de prestatieindicatoren. Landen die mogelijk een vertekening in de resultaten kunnen veroorzaken, kunnen met een blik op de 20 grafieken, aangeduid worden.

Vertekening kan zich voordoen wanneer een land op de indicatoren systematisch beter of slechter presteert, dan op basis van het aantal verkeersdoden. De geplotte scores moeten dus niet allemaal precies binnen de grenzen vallen. Sterke variatie in de scores op de indicatoren in combinatie met een gemiddeld aantal verkeersdoden of een enkele uitbijter zijn ook aanvaardbaar.

Cyprus (Figuur 4) en **Denemarken** (Figuur 5) zijn landen waar men zeer duidelijk kan zien dat er een grote overeenkomst is tussen de plaats op de rangschikking van de prestatie-indicatoren en plaats op de consensusranglijst. Een voorbeeld van een land met een gemiddeld aantal verkeersdoden en een grote variatie in prestatie op de prestatie-indicatoren is **Estland** (Figuur 7), dat een lage middenpositie inneemt op de

consensusranglijst en zeer lage prestaties op de indicatoren Speed, Vehicle en Trauma Care combineert met relatief tot zeer hoge scores op Alcohol en DRL Law. Bij **Ierland** (Figuur 8) zien we iets soortgelijks. Een hoge middenpositie op basis van het aantal verkeersdoden per miljoen inwoners wordt gecombineerd met zeer hoge scores op Speed en Vehicle, terwijl er slechte scores zijn voor DRL Law en Trauma Care. In Figuur 3 (**België**) en figuur 10 (**Griekenland**) zijn voorbeelden te vinden van landen die één uitbijter laten noteren en voor het overige een grote overeenkomst lijken te hebben tussen consensusranglijst en de ranglijsten op basis van de prestatie-indicatoren. Ook **Zweden**, het meest verkeersveilige land uit de selectie, haalt enkel een ondergemiddelde score voor de indicator Infrastructure (Figuur 22).

Finland (Figuur 8) is een wat eigenaardig geval. Wanneer we enkel de prestaties op de indicatoren Alcohol, Speed, Seatbelt en DRL Law in beschouwing nemen, merken we een sterke overeenkomst met de consensusranglijst. De prestaties op de indicatoren Vehicle, Infrastructure en Trauma Care zijn dan weer opvallend laag in vergelijking met het aantal verkeersdoden per miljoen inwoners.

Het **Verenigd Koninkrijk** (Figuur 21) en in mindere mate **Italië** (Figuur 13) worden op basis van de indicatoren lager gerangschikt dan op basis van het aantal verkeersdoden. **Nederland** (Figuur 14), dat op basis van het aantal verkeersdoden per miljoen inwoners op een derde plaats gerangschikt wordt, wordt enkel op basis van de SPI Infrastructure hoog gerangschikt. Alle andere prestatie-indicatoren rangschikken Nederland onder die derde plaats. Het is dan ook niet onwaarschijnlijk dat voor deze landen, de road safety score, en dan vooral hun plaats in de landenrangschikking, een onderschatting zal zijn van de werkelijke situatie.

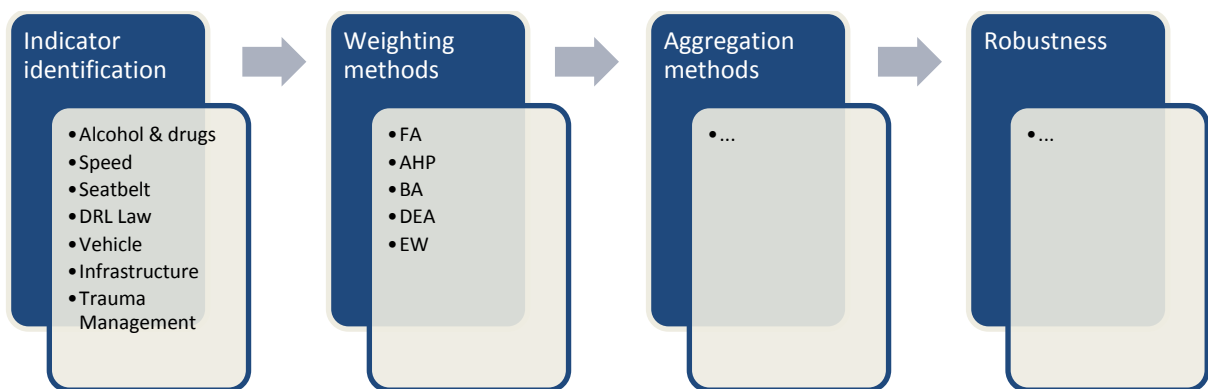
Er zijn ook een aantal landen die zich in de omgekeerde situatie bevinden. **Polen** (Figuur 16), **Portugal** (Figuur 17), **Slovenië** (Figuur 18) en **Tsjechië** (Figuur 20), zijn landen die, in vergelijking met hun prestaties op de indicatoren, minder presteren op basis van het aantal verkeersdoden per miljoen inwoners.

Het zijn deze laatste acht landen, die de hypothese van samenhang tussen rangschikkingen op basis van de indicatoren en de consensusranglijst wat ondermijnen. In hun geval worden hoge scores op de prestatie-indicatoren blijkbaar niet vertaald in een relatief laag aantal verkeersdoden per miljoen inwoners, of gaat een laag aantal

verkeersdoden gepaard met lage scores op de prestatie-indicatoren.² Er kan dus enige vorm van vertekening optreden. We kunnen vermoeden dat er een verklarende indicator ontbreekt in de analyse.

1.3 Wegingmethodes

De resultaten van dit onderzoek, zullen afhankelijk zijn van de wegingmethodes die zullen geaggregeerd worden. Deze methodes vallen buiten het bereik van dit onderzoek. We zullen een kort overzicht geven van de betrokken wegingmethodes, inclusief vermeldingen van bron en resultaten. Deze resultaten zijn in de regel afkomstig van voorafgaand onderzoek.



Figuur 1 b: Overzicht van het onderzoek

Het doel van de wegingmethodes is om een bepaald gewicht toe te kennen aan de SPI's. Aan de hand van deze gewichten kunnen we dan conclusies trekken met betrekking tot het belang van elke SPI in de verkeersonveiligheid. Hoe groter het toegekende gewicht, hoe meer de betreffende SPI een invloed zal uitoefenen op het aantal verkeersdoden.

De wegingmethodes maken allen gebruik van de gegevens die teruggevonden kunnen worden in Tabel 2. Sommige methodes maken daarbij aansluitend ook gebruik van de gegevens in Tabel 4. In dit laatste geval wordt de rangschikking van de landen op basis

² Zowel een laag aantal verkeersdoden per miljoen inwoners als een hoge score op de prestatie-indicatoren, worden vertaald in een hoge genormaliseerde score. Een score van 1 in de consensusranglijst, duidt dus op een laag aantal verkeersdoden, terwijl diezelfde score van 1 op een prestatie-indicator, een hoge indicatorwaarde betekent. Zowel hoge waarden op de prestatie-indicatoren als een laag dodencijfer, wordt immers als een goede prestatie beschouwd.

van het aantal verkeersdoden per miljoen inwoners, gebruikt als ideale representatie van verkeersveiligheid (Hermans, Van den Bossche, & Wets, 2008b). De methodes die gebruik maken van een dergelijke consensus zullen we dan ook de *wegingmethodes met consensusranglijst* noemen.

Er moet opgemerkt worden dat de data waarop de wegingmethodes zijn gebaseerd, ook gegevens van Zwitserland bevatten. Wegens het ontbreken van gegevens over Zwitserland in Tabel 4, meer bepaald de gegevens met betrekking tot het aantal verkeersdoden per miljoen kilometer en per 100.000 personenwagens, hebben we besloten Zwitserland niet langer te beschouwen. Weging en aggregatie zijn dus niet op basis van dezelfde gegevensset gebeurd. We veronderstellen dat het weglaten van 1 land, de resultaten niet significant zullen beïnvloeden.

Hierna volgt een overzicht van de gebruikte wegingmethodes. We maken een onderscheid tussen methodes met en methodes zonder consensusranglijst.

1.3.1 Factor Analysis (FA)

Factor Analysis wordt beschreven door Nardo et al. (2005). De methode is gebaseerd op het reduceren van het aantal dimensies dat de totale variantie, zijnde de mate van verkeersveiligheid in een land ten opzichte van de gemiddelde verkeersveiligheid voor de dataset, verklaart. Meer over FA kan gevonden worden in Nardo et al. (2005) en Hermans et al. (2008a). FA is toegepast op de gegevens in Tabel 2 door Hermans, Van den Bossche & Wets. De resulterende gewichten hebben we overgenomen uit dit onderzoek (Hermans, Van den Bossche, & Wets, 2008a) en worden getoond in Tabel 6.

1.3.2 Analytical Hierarchy Process (AHP)

Analytical Process Hierarchy is in de jaren 70 bedacht door Saaty (Hermans, Van den Bossche, & Wets, 2008a). Bij deze methode wordt het probleem vertaald in een algemene doelstelling met ondergeschikte subdoelen. Vervolgens wordt er aan experts gevraagd in welke mate elke indicator bijdraagt tot het behalen van deze doelen. Vervolgens worden preference matrices opgesteld om ten slotte met behulp van de eigenvector waarde tot gewichten te komen. (Hermans, Van den Bossche, & Wets,

2008b) De resulterende gewichten in Tabel 6 zijn overgenomen uit Hermans, Van den Bossche & Wets (2008a).

1.3.3 Budget Allocation (BA)

Budget Allocation is, vanuit technisch oogpunt, een eenvoudige methode die beschreven wordt door Nardo et al. (2005) Er wordt aan een panel van experts gevraagd om een beschikbaar budget te verdelen over de verschillende indicatoren. Hoe belangrijker men de indicator vindt, hoe meer gewicht men er aan dient toe te kennen. (Hermans, Van den Bossche, & Wets, 2008a) De resultaten van de methode BA hebben we eveneens overgenomen uit het artikel Combining Road Safety Information in a Performance Index (Hermans, Van den Bossche, & Wets, 2008a) en worden getoond in Tabel 6.

1.3.4 Data Envelopment Analysis (DEA)

Data Envelopment Analysis is een methode die werkt als een maximalisatieprobleem en is ontworpen door Charnes, Cooper & Rhodes (1978). In ons geval is een aangepaste versie gebruikt van het model dat voorgesteld werd door Cherchye et al. (2006) Deze aanpassing was nodig aangezien het vorige model voor elk land een gewicht per SPI vast stelde, in plaats van één voor alle landen geldend gewicht per SPI. Als extra beperkingen werden er boven en benedengrenzen aan de gewichten bepaald door experts. De doelfunctie van het maximalisatieprobleem maximaliseert de totale som van het product van elke SPI-waarde en het toegekende gewicht aan de SPI. (Hermans, Van den Bossche, & Wets, 2008a) Merk op dat DEA geen gebruik maakt van een consensusrangschikking. De resultaten van DEA (zie Tabel 6) hebben we overgenomen uit Hermans, Van den Bossche & Wets (2008a)

1.3.5 Equal Weighting (EW)

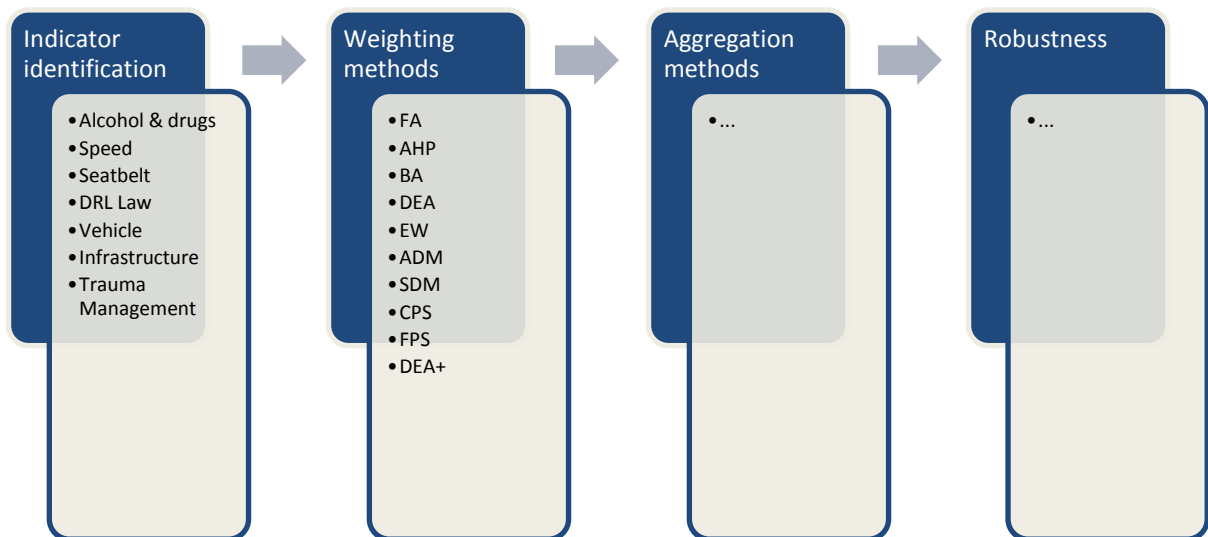
Een eenvoudige manier om SPI's te wegen, is de methode Equal Weighting. Deze wegingmethode kent aan elke SPI hetzelfde gewicht toe en maakt dus geen onderscheid tussen de indicatoren (Hermans, Van den Bossche, & Wets, 2008a). In Tabel 6 is een overzicht te vinden van de resulterende gewichten voor de wegingmethodes zonder consensusranglijst.

2 Additionele wegingmethodes

Om de prestatie van zoveel mogelijk wegingmethodes te bepalen, hebben we ervoor gekozen om er een aantal toe te voegen. De keuze om wegingmethodes toe te voegen, wordt mede bepaald door het feit dat we onder de beschikbare informatie geen wegingmethodes terugvinden die gebruik maken van een consensusranglijst.

Er zijn al toepassingen gebeurd van wegingmethodes met consensusranglijst op verkeersveiligheidsgegevens (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008). Jammer genoeg komt de gebruikte dataset niet helemaal overeen met degene waar wij gebruik van maken. We hebben er dus voor gekozen om exact dezelfde procedure te volgen, maar deze keer wordt de methode toegepast op de dataset die we in dit onderzoek gebruiken. Om het aantal wegingmethodes met consensus nog op te drijven, hebben we er daar nog één nieuwe methode aan toegevoegd. Deze methode zal hieronder toegelicht worden.

Bovendien zijn er in de literatuur wegingmethodes te vinden die gewichten toewijzen op basis van variatie in de gegevens (Wang & Fan, 2007). Aangezien we geen methodes ter beschikking hebben die op een dergelijke manier werken, hebben we ervoor gekozen om ook deze methodes aan de lijst van wegingmethodes zonder consensus toe te voegen.



Figuur 1 c: Overzicht van het onderzoek

2.1 Wegingmethodes zonder consensusranglijst

2.1.1 Absolute Deviation Method (ADM)

De twee Deviation Methods, Absolute Deviation Method en Standard Deviation Method, zijn rekenkundig de ingewikkeldste methodes die we in dit onderzoek gebruiken. Bij ADM gaat men ervan uit dat hoe groter de verschillen in waarden voor een bepaalde SPI zijn, hoe waardevoller deze indicator is. (Wang & Fan, 2007) We kunnen deze stelling ook omdraaien door te stellen dat: hoe dichter de waarden op een SPI bij elkaar liggen, hoe minder voorspellende kracht deze SPI zal hebben in termen van verkeersveiligheid. We citeren: "Large weights should be given to those fuzzy preference relations with big absolute deviation and/or standard deviation" (Wang & Fan, 2007) ADM maakt gebruik van absolute afwijkingen. Voor een nauwkeurige beschrijving van ADM, verwijzen we naar Wan & Fan (2007). In hoofdstuk 3 zullen we zelf ook dieper ingaan op deze materie. De resultaten hebben we bekomen door hun methodes toe te passen op de gegevens in Tabel 2 en zijn terug te vinden in Tabel 6.

2.1.2 Standard Deviation Method (SDM)

De Standard Deviation Method verloopt volledig analoog met ADM, uitgezonderd het feit dat er met standaarddeviaties gewerkt wordt. (Wang & Fan, 2007) De resultaten zijn bekomen door het toepassen van de methode zoals beschreven staat door Wang & Fan op de gegevens in Tabel 6. Ook SDM wordt nader uitgelegd in hoofdstuk 3.

2.2 Wegingmethodes met consensusranglijst

2.2.1 Classical Preference Structures (CPS)

Classical Preference Relations is door Hermans et al. toegepast op de verkeersveiligheidsgegevens in Tabel 2 (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008). Zowel voor elke SPI als voor het aantal verkeersdoden per miljoen inwoners, worden matrices opgesteld met preference relations tussen elk paar van landen. Voor elke prestatie-indicator bestaat er dus een matrix waaruit afgeleid kan worden of land A beter presteert dan land B. Vervolgens wordt berekend in welke mate de matrix met

preference relations, gebaseerd op het aantal verkeersdoden, overeenkomt met de SPI-matrices. Hoe groter de overeenkomst tussen de matrix van het aantal verkeersdoden en deze van een bepaalde SPI, hoe groter het gewicht dat aan deze SPI toegekend zal worden. (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008) De redenering is dat een SPI een goede voorspeller is van verkeersveiligheid als de relatieve rangschikking van de landen voor de SPI, in sterke mate gelijk is op die van het aantal verkeersdoden per miljoen inwoners. De resultaten van deze wegingmethode kunnen teruggevonden worden in Tabel 5. In hoofdstuk 3 zal dieper ingegaan worden op de methode CPS.

2.2.2 Fuzzy Preference Structures (FPS)

Fuzzy Preference Relations werkt op dezelfde manier als CPS. Het verschil zit in het feit dat er bij FPS gradaties van voorkeur kunnen opgegeven worden. In de matrices met preference relations, kan men dus niet enkel aflezen of land A wel dan niet beter gerangschikt staat dan land B, maar ook hoe groot het verschil is tussen de landen. (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008) Merk op dat er ook hier met genormaliseerde waarden wordt gewerkt. Ook FPS is toegepast door Hermans et al. in de paper 'Ranking Road Safety Risk Factors using Preference Structures and Fuzzy Preference Structures' (2008). De resulterende gewichten zijn terug te vinden in Tabel 5.

2.2.3 Adapted Data Envelopment Analysis (DEA+)

DEA+ is gebaseerd op DEA, met een aanpassing zodat er rekening gehouden wordt met een consensusranglijst in de doelfunctie van het maximalisatieprobleem. In plaats van de totale som van het product van elke waarde op elke SPI en het bijhorende gewicht te maximaliseren, kiezen we er in dit geval voor om het verschil met de consensusranglijst te minimaliseren. Dit doen we door eerst voor elk land een score te berekenen aan de hand van de waarden op de verschillende SPI's en de bijhorende gewichten. Vervolgens maken we het verschil, in absolute waarde, tussen deze score en de genormaliseerde waarde op aantal verkeersdoden per miljoen inwoners. De som van deze verschillen trachten we dan te minimaliseren door een optimale combinatie van gewichten te vinden. Per SPI zijn er bepaalde grenzen voorzien waarbinnen de gewichten moeten blijven. De grenzen zijn identiek aan diegene die gebruikt zijn door Hermans, Van den Bossche en Wets (2008a) en zijn dezelfde als deze gebruikt in de DEA-methode. De gevonden

gewichten voor DEA+ en de twee andere wegingmethodes die gebruik maken van een consensusranglijst, zijn terug te vinden in Tabel 5.

2.3 Resulterende gewichten voor de verschillende wegingmethodes

In onderstaande tabellen is een overzicht te vinden van de gewichten die door de verschillende wegingmethodes aan elke prestatie-indicator toegekend wordt.

Tabel 1: Gewichten per wegingmethode toegewezen aan elke SPI (wegingmethodes met consensus)

Indicator	CPR	FPR	DEA+
Alc	0,124 (5)	0,115 (5)	0,078 (6)
Speed	0,162 (3)	0,124 (4)	0,158 (3)
Seat Belt	0,099 (6)	0,241 (1)	0,316 (1)
DRL law	0,159 (4)	0,096 (6)	0,023 (7)
Veh	0,096 (7)	0,227 (2)	0,158 (3)
Infr	0,196 (1)	0,068 (7)	0,100 (5)
Trauma	0,164 (2)	0,130 (3)	0,167 (2)

Tabel 2: Gewichten per wegingmethode toegewezen aan elke SPI (wegingmethodes zonder consensus)

Indicator	FA	AHP	BA	DEA	EW	ADM	SDM
Alc	0,139 (4)	0,229 (2)	0,234 (2)	0,259 (1)	0,143 (1)	0,103 (6)	0,133 (6)
Speed	0,190 (1)	0,262 (1)	0,305 (1)	0,250 (2)	0,143 (1)	0,151 (3)	0,143 (3)
Seat Belt	0,095 (7)	0,129 (4)	0,132 (3)	0,211 (3)	0,143 (1)	0,138 (4)	0,140 (4)
DRL law	0,163 (3)	0,028 (7)	0,043 (7)	0,000 (7)	0,143 (1)	0,210 (1)	0,173 (1)
Veh	0,100 (6)	0,102 (5)	0,071 (6)	0,035 (5)	0,143 (1)	0,159 (2)	0,147 (2)
Infr	0,139 (4)	0,179 (3)	0,105 (5)	0,080 (4)	0,143 (1)	0,102 (7)	0,129 (7)
Trauma	0,174 (2)	0,071 (6)	0,110 (4)	0,028 (6)	0,143 (1)	0,135 (5)	0,136 (5)

Zowel in Tabel 5 als in Tabel 6 kunnen er verschillen genoteerd worden tussen de verschillende wegingmethodes onderling. De grootte van de gewichten die toegekend worden aan de prestatie-indicatoren, verschillen van wegingmethode tot wegingmethode. Bovendien zijn ADM en SDM de enige wegingmethodes die de prestatie-indicatoren in dezelfde volgorde rangschikken.

In Tabel 5 is de volgorde van gewichten die FPR toekent aan de prestatie-indicatoren, vergelijkbaar met de volgorde aan de hand van DEA+. De wegingmethode CPR komt tot verschillende resultaten. De grootste verschillen kunnen opgetekend worden voor de prestatie-indicatoren Seatbelt, Vehicle en Infrastructure. CPR wijst het hoogste gewicht toe aan Infrastructure, terwijl FPR en DEA+ eerder een laag gewicht aan deze prestatie-

indicator toekennen. Voor Seatbelt geldt het omgekeerde. FPR en DEA+ wijzen het hoogste gewicht toe aan deze prestatie-indicator, terwijl CPR deze indicator veel minder belangrijk acht. Hetzelfde kan gezegd worden van de prestatie-indicator Vehicle, die door CPR het laagste gewicht toegewezen krijgt terwijl zowel FPR als DEA+ deze indicator als belangrijk beschouwen.

Voor de wegingmethodes zonder consensus (Tabel 6), zijn de onderlinge overeenkomsten iets groter. Op basis van de volgorde van de gewichten, kan men stellen dat, behalve ADM en SDM, ook AHP, BA en DEA onderling vergelijkbaar zijn. Toch moet er op de verschillen tussen de wegingmethodes gelet worden. Over het algemeen is enkel de volgorde van Speed, dat telkens als een belangrijke prestatie-indicator aangewezen wordt, vergelijkbaar voor de verschillende methodes. De grootste onderlinge verschillen kunnen opgetekend worden voor de prestatie-indicatoren Alcohol, DRL Law en Trauma Management, die door de verschillende wegingmethodes zeer diverse volgordes toegewezen krijgen. Voor de indicatoren Vehicle en Infrastructure zijn er vooral verschillen te merken tussen ADM en SDM enerzijds en de overige wegingmethodes anderzijds. Voor Seatbelt kan er een verschil gezien worden tussen FA enerzijds en de overige wegingmethodes, met uitzondering van EW, anderzijds.

De grote verscheidenheid in grootte en volgorde van de gewichten maakt aggregatie interessant en waardevol. Het verkiezen van de ene wegingmethode boven de andere levert immers zeer verschillende resultaten op. De manier waarop de verschillende wegingmethodes zullen gecombineerd zal dus een grote invloed hebben op de geaggregeerde gewichten en de daaruit voortkomende landenrangschikking.

3 Aggregatiemethoden

De deelvraag die we in dit hoofdstuk zullen behandelen is:

Op welke manieren kunnen de verschillende toegekende gewichten, geaggregeerd worden tot één gewicht?

Aggregatie van informatie kan omschreven worden als het bijeen voegen van een verscheidenheid aan informatie in een bepaalde vorm. Er zijn twee belangrijke toepassingsgebieden waarbinnen aggregatie zijn nut kan hebben.

Het eerste toepassingsgebied is dat van Information Management. Aggregatie is een techniek om met een overvloed aan informatie om te gaan. *Information overload* is de situatie die volgens Chaffey en Wood (2005) ontstaat wanneer "The capacity of individuals and systems [...] to derive value from information is exceeded by the volume and complexity of information." In het informatietijdperk is er haast oneindig veel informatie beschikbaar terwijl we maar een beperkte tijd hebben om die te verwerken. Een grote verscheidenheid aan informatie integreren in één enkel getal levert kostbare tijdsbesparing op, die bovendien geen afbreuk hoeft te doen aan de kwaliteit van de informatie. Bovendien wordt de graad van complexiteit verminderd waardoor interpretatie gemakkelijker wordt.

Data Mining is het tweede toepassingsgebied waarin er van aggregatie gebruik wordt gemaakt. In dit geval is het vooral de bedoeling om de beperkingen van het *systeem* weg te werken. (Bij Information Management, ligt de nadruk eigenlijk op de beperkingen van de *gebruiker*). In Data Mining kunnen er verschillende modellen gebruikt worden om informatie uit data te halen. Die modellen hebben echter elk hun beperkingen en zwakke punten. Door de modellen te combineren (aggregeren) probeert men de sterke punten van een bepaald model te benutten om de zwakke punten van een ander model op te vangen (Witten & Frank, 1999).

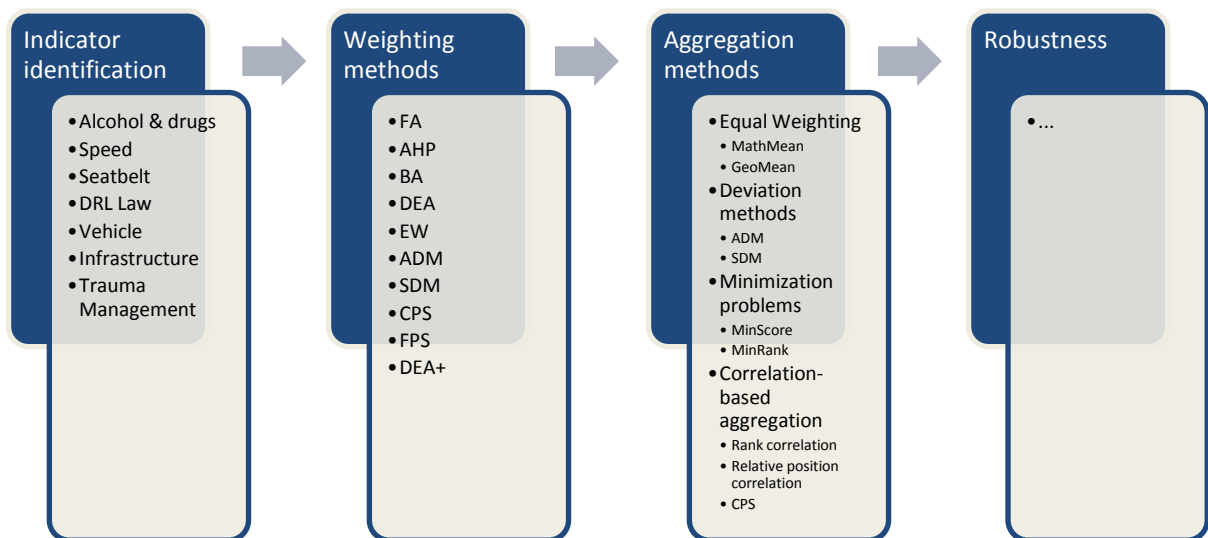
In ons geval gebruiken we de aggregatie vooral in zijn tweede hoedanigheid. Er bestaat geen 'ultieme' methode om gewichten te bepalen die in elke omstandigheid een correcte weergave van de werkelijkheid geeft. Elke methode heeft zijn voor- en nadelen. Deze

tekortkomingen willen we compenseren door de methodes op een zo optimaal mogelijke manier te combineren.

Aggregatie van de gewichten kan men bekijken als het toekennen van gewichten aan de verschillende wegingmethoden. We kunnen dus voor aggregatie dezelfde wegingmethodes gebruiken als die waarmee gewichten toegekend werden aan de SPI's.

We bespreken in de volgende paragrafen een aantal manieren waarop we aggregatie kunnen toepassen. Vervolgens wijzen we op de sterke en zwakke punten van elke methode om tenslotte de resultaten en interpretatie weer te geven.

Tot slot van dit hoofdstuk zullen we één groep aggregatiemethodes selecteren, die we in verder fasen zullen gebruiken om het onderzoek naar robuustheid verder te zetten. Hoewel we de voor- en nadelen van elke aggregatiemethode in beschouwing nemen, blijft dit een vrij subjectieve keuze. Ook bij het aggregeren bestaat er geen universeel aanvaarde optimale methode. We zouden dus ook weer de aggregatiemethodes kunnen aggregeren. Dit zou ons echter in een oneindig straatje brengen. We kiezen ervoor na de eerste aggregatiestap te stoppen.



Figuur 1 d: Overzicht van het onderzoek

3.1 Beschrijving van de aggregatiemethoden

We brengen de aggregatiemethoden in 4 categorieën onder: equal weighting, deviation methods, minimization problems en correlation-based aggregation. Equal weighting en de deviation methods onderscheiden zich van de andere twee categorieën doordat ze geen gebruik maken van een consensusranglijst om tot resultaten te komen. Minimization problems en correlation-based aggregation gebruiken dus wel een consensus ranglijst om tot resultaten (geaggregeerde gewichten en de daaruit volgende landenrangschikking) te komen.

3.1.1 Equal Weighting

Wellicht de meest eenvoudige manier om methodes te combineren is Equal Weighting toepassen. Men gaat ervan uit dat de gebruikte methodes evenwaardig zijn aan elkaar en dus allemaal even zwaar mogen doorwegen in het geaggregeerde gewicht. In de praktijk komt dit erop neer dat per indicator het gemiddelde gewicht berekend wordt.

Het grootste voordeel van equal weighting is, dat het eenvoudig te berekenen is. Als het verschil tussen de gebruikte methodes verwaarloosbaar is, is equal weighting bovendien de beste aggregatiemethode. Complexere aggregatiemethodes zouden in dat geval immers ongeveer dezelfde uitkomst moeten geven als equal weighting. Het is in dat geval niet aan te raden het aggregatieproces te compliceren.

De eenvoud, het sterke punt, van equal weighting is paradoxaal genoeg ook het zwakke punt. Er kan er niet altijd vanuit gegaan worden dat er geen verschil bestaat tussen de toegepaste methodes.

Bovendien wordt er op geen enkele manier rekening gehouden met de performantie van de wegingmethoden. Methodes die aan indicatoren gewichten toekennen die dicht bij de 'werkelijke waarden' liggen, en dus ook een betere voorspelling geven met betrekking tot verkeersveiligheid, worden daar niet voor beloond in deze aggregatiemethode. Equal weighting zal de kracht van het systeem om een indicatie te geven van verkeersveiligheid waarschijnlijk niet optimaliseren. Er zullen methodes zijn die gewichten toekennen die dicht bij de werkelijke waarden liggen, dan de geaggregeerde gewichten. Dit terwijl de doelstelling van aggregatie is om tot geaggregeerde gewichten

te komen die dichter bij de werkelijkheid aanleunen dan de wegingmethodes elk apart bereiken.

Er zijn twee vormen van equal weighting die we kunnen toepassen, namelijk het wiskundig gemiddelde en het meetkundig gemiddelde. Beide vormen zullen we hieronder kort beschrijven.

3.1.1.1 Wiskundig gemiddelde (Math. Mean)

Bij de eerste methode wordt het wiskundig gemiddelde gebruikt. We tellen de toegewezen gewichten per indicator per methode op en delen het door het aantal methodes. Het berekenen van het wiskundig gemiddelde van de gewichten, toegewezen door de verschillende wegingmethodes, per indicator, komt op hetzelfde neer als elk gewicht van elke wegingmethode voor $\frac{1}{N}$ te laten meetellen in het geaggregeerd gewicht.

$$\bar{w}_i = \frac{\sum_{j=1}^M w_{ij}}{M}$$

$\bar{w}_i$ het gewogen gewicht per indicator i

M het aantal wegingmethodes

w_{ij} het gewicht door methode j toegekend aan indicator i

Om de drie wegingmethodes die gebruik maken van een consensus ranglijst, te aggregeren door middel van het wiskundig gemiddelde, moeten per SPI de drie gewichten die toegekend worden door CPS, FPS en DEA+ opgeteld worden en vervolgens gedeeld worden door drie. De resultaten van deze bewerkingen kan u terugvinden in de onderstaande Tabel 7, kolom 2.

Tabel 3: Gewichten per SPI per aggregatiemethode bij aggregeren van wegingmethodes met consensus

	Math. mean	Geo. mean	ADM	SDM	Min score verschil	Min ranking verschil	Correlatie ranking	Correlatie relatieve positie	CPS
Alcohol	0,106 (6)	0,134 (6)	0,105 (6)	0,105 (6)	0,107 (5)	0,116 (5)	0,105 (6)	0,104 (6)	0,103 (6)
Speed	0,148 (4)	0,149 (4)	0,147 (4)	0,148 (4)	0,137 (4)	0,124 (4)	0,147 (4)	0,147 (4)	0,147 (4)
Seatbelt	0,218 (1)	0,162 (1)	0,227 (1)	0,222 (1)	0,246 (1)	0,240 (1)	0,227 (1)	0,232 (1)	0,233 (1)
DRL Law	0,093 (7)	0,119 (7)	0,087 (7)	0,090 (7)	0,083 (7)	0,096 (6)	0,088 (7)	0,085 (7)	0,084 (7)
Vehicle	0,160 (2)	0,150 (2)	0,164 (2)	0,162 (2)	0,196 (2)	0,226 (2)	0,164 (2)	0,166 (2)	0,166 (2)
Infrastr.	0,121 (5)	0,136 (5)	0,116 (5)	0,119 (5)	0,089 (6)	0,068 (7)	0,117 (5)	0,114 (5)	0,114 (5)
Trauma C.	0,154 (3)	0,150 (2)	0,153 (3)	0,153 (3)	0,143 (3)	0,130 (3)	0,153 (3)	0,153 (3)	0,153 (3)

De aggregatiemethode Math Mean duidt de SPI Seatbelt met een gewicht van 0,218 aan als de voornaamste component van verkeersveiligheid. Het promoten van het dragen van de gordel zou dus een efficiënte manier zijn om het aantal verkeersdoden per miljoen inwoners te doen afnemen. Vehicle neemt de 2^{de} plaats in op de ranglijst van Safety Performance Indicators, vlak voor Trauma Care. De wetgeving met betrekking tot het gebruik van de motorvoertuigverlichting overdag zou het minste invloed hebben op de verkeersveiligheid.

Wanneer we een blik werpen op Tabel 9 in bijlage, zien we de landenrangschikking met betrekking tot verkeersveiligheid wanneer we de wegingmethodes met consensus aggregeren door middel van het wiskundig gemiddelde (Math mean). In de 1^{ste} kolom staat de naam van het betreffende land, de 2^{de} kolom geeft de road safety score weer die we verkrijgen door de prestaties van elk land voor elke SPI te vermenigvuldigen met het berekende gewicht van de SPI. De 3^{de} kolom geeft de positie in de rangschikking weer die het land inneemt op basis van de road safety score. In de 4^{de} kolom wordt de positie vermeld die het betreffende land inneemt op de ranglijst van aantal verkeersdoden per miljoen inwoners, de consensusranglijst. De laatste kolom geeft dan het verschil tussen de plaats die het land inneemt op de consensusranglijst en de plaats die het land inneemt volgens zijn prestatie-index. Op die manier kunnen we kijken hoe 'accuraat' de methodes de consensusrangschikking kunnen reconstrueren of hoe goed de methodes de rangschikking op basis van het aantal verkeersdoden per miljoen inwoners benaderen.

Tabel 9 geeft aan, op basis van de scores, dat Slovenië het verkeerveiligste land uit de dataset zou moeten zijn, voor Frankrijk en Zweden. Als we de consensusranglijst beschouwen als de werkelijke uiting van verkeersveiligheid, zien we dat vooral de positie van Slovenië slecht voorspeld wordt. Gesteld dat we een verschil van vijf posties tussen de twee rangschikkingen aanvaardbaar vinden³, zijn Slovenië, Frankrijk, Nederland, Finland, Portugal en Estland de landen waarvan de verkeersveiligheid slecht geschat wordt. Slovenië, Frankrijk en Portugal worden te hoog gerangschikt, terwijl de drie andere landen een te lage rangschikking op basis van de gewogen prestatie-index toebedeeld krijgen.

³ De tolerantiegrens om een verschil van 5 plaatsen te aanvaarden, is arbitrair gekozen. We kunnen argumenteren dat een verschil van 5 plaatsen over het algemeen het verschil maakt tussen een plaats in de topklasse, een plaats in de middenmoot en een plaats onderaan de klassering. Bij de interpretatie moeten deze grenzen voor ogen gehouden worden

De minder goede classificatie van de voornoemde landen kunnen we grotendeels verklaren door terug te kijken naar hoofdstuk 1.2. Slovenië en Portugal presteren zeer goed op vlak van de verschillende Safety Performance Indicators, terwijl ze toch een relatief hoog aantal verkeersdoden hebben per miljoen inwoners. Aangezien de scores berekend worden aan de hand van de prestaties op de SPI's, worden deze dus haast met zekerheid te hoog geschat. Nederland haalt dan weer overwegend matige scores voor de SPI's in vergelijking met het lage dodencijfer, wat onderschatting door Mathmean waarschijnlijk maakt. Ook het feit dat Frankrijk wat te hoog gerangschikt wordt, kunnen we aan de hand van de dataset nog begrijpen. We merken dat zeer goede prestaties van Frankrijk op de indicatoren Seatbelt en Trauma Care, die beiden een groot gewicht toegewezen krijgen. Als we rekening houden met de anomalieën in de dataset (zie ook Hoofdstuk 1.2), zijn het dus vooral Finland en Estland die slechter dan verwacht geclassificeerd worden. Voor deze landen kunnen we niet duidelijk een negatieve correlatie vaststellen tussen de SPI's en het aantal verkeersdoden. Doordat Vehicle en Trauma Care een groot gewicht toegewezen krijgen, worden landen met een lage score op deze prestatie-indicatoren, zoals Estland en Finland, benadeeld bij de berekening van de road safety score.

Wanneer we dezelfde aggregatieoefening uitvoeren op de wegingmethodes die geen gebruik maken van een ideale representatie van verkeersveiligheid, bekomen we de geaggregeerde gewichten uit Tabel 8, kolom 2. Speed wordt aangeduid als voornaamste SPI, met Alcohol op de 2^{de} en Seatbelt op de 3^{de} plaats. Vehicle zou nog net iets minder invloed uitoefenen op verkeersveiligheid dan DRL Law.

Tabel 4: Gewichten per SPI per aggregatiemethode bij aggregeren van wegingmethodes zonder consensus

	<u>Math. mean</u>	<u>Geo. mean</u>	<u>ADM</u>	<u>SDM</u>	<u>Min score verschil</u>	<u>Min ranking verschil</u>	<u>Correlatie ranking</u>	<u>Correlatie relatieve positie</u>	<u>CPS</u>
Alcohol	0,181 (2)	0,194 (2)	0,181 (2)	0,181 (2)	0,129 (6)	0,234 (2)	0,181 (2)	0,183 (2)	0,183 (2)
Speed	0,210 (1)	0,217 (1)	0,210 (1)	0,211 (1)	0,150 (2)	0,252 (1)	0,211 (1)	0,213 (1)	0,213 (1)
Seatbelt	0,144 (3)	0,168 (3)	0,145 (3)	0,144 (3)	0,140 (4)	0,178 (3)	0,145 (3)	0,146 (3)	0,146 (3)
DRL Law	0,111 (6)	0,000 (7)	0,111 (6)	0,111 (6)	0,173 (1)	0,053 (7)	0,110 (6)	0,109 (6)	0,109 (6)
Vehicle	0,110 (7)	0,132 (6)	0,110 (7)	0,111 (6)	0,148 (3)	0,083 (5)	0,110 (6)	0,109 (6)	0,109 (6)
Infrastr.	0,128 (4)	0,154 (4)	0,128 (4)	0,128 (4)	0,123 (7)	0,120 (4)	0,127 (4)	0,126 (4)	0,126 (4)
Trauma C.	0,116 (5)	0,135 (5)	0,116 (5)	0,116 (5)	0,136 (5)	0,080 (6)	0,116 (5)	0,115 (5)	0,114 (5)

Het valt op dat de gewichten, en vooral de rangorde van SPI's, sterk verschilt met deze waarbij enkel de wegingmethodes met consensusranglijst geaggregeerd worden. We merken ook op dat de gewichten dichter bij elkaar liggen dan in het voorgaande geval.

Als we de landen rangschikken op basis van deze gewichten, bekomen we de rangschikking uit Tabel 10. Ook hier wordt Slovenië weer als 1^{ste} gerangschikt. Deze keer wordt Slovenië echter gevolgd door Ierland en Frankrijk.

Deze rangschikking sluit misschien iets beter aan bij wat we verwachtten uit het hoofdstuk 1.3. Slovenië, Nederland, Portugal en Italië zijn landen waarvan de prestaties op de SPI's, in het algemeen, niet overeenkomen met hun prestaties op vlak van verkeersveiligheid (zijnde het aantal verkeersdoden per miljoen inwoners). Het zijn dan ook weer deze landen waarvan de rangschikking slecht voorspeld wordt door de methode. Voor het overige zien we ook Estland weer beduidend lager eindigen dan we op basis van het aantal verkeersdoden zouden mogen verwachten.

3.1.1.2 Geometrisch gemiddelde

Het geometrisch gemiddelde is een andere manier om gemiddeldes te berekenen. Hierbij maakt men eerst het product van de n verschillende gewichten. Vervolgens nemen we de n -de machtswortel van dat product.

$$\bar{w}_i = \sqrt[M]{\prod_{j=1}^M w_{ij}}$$

$\bar{w}_i$ het gewogen gewicht per indicator i

M het aantal wegingmethodes

w_{ij} het gewicht door methode j toegekend aan indicator i

We moeten opmerken dat het geometrisch gemiddelde strikt genomen niet onder equal weighting valt. Wegingmethodes die lage gewichten toekennen aan de indicatoren in vergelijking met de andere methodes, zullen zwaarder doorwegen in de eindafrekening. Het volstaat bijvoorbeeld om één enkele wegingmethode te hebben die een gewicht van 0 toebedeeld aan een bepaalde indicator om ook het geaggregeerde gewicht gelijk te stellen aan 0.

De geaggregeerde gewichten voor de wegingmethodes met consensus, worden gegeven in Tabel 7. De gewichten liggen bij Geo mean veel dichter bij elkaar dan bij Math mean,

maar de volgorde van de gewichten blijft behouden. Ook hier staat Seatbelt dus op de 1^{ste} plaats, voor Vehicle en Trauma care, met DRL als SPI met de minste invloed.

In de rangschikking van landen (Tabel 11) zien we andermaal Slovenië helemaal bovenaan de lijst opduiken, gevolgd door Frankrijk en Zweden. Ook hier is Slovenië weer een, te verwachten, probleemgeval. Ook Frankrijk, Finland, Nederland, Portugal en Estland staan wat te hoog, dan wel te laag, gerangschikt. Frankrijk dankt zijn gunstige positie aan uitstekende prestaties op de indicatoren Seatbelt en Trauma Care, terwijl de zwakke score voor de prestatie-indicator Alcohol minder doorweegt dankzij het kleine geaggregeerde gewicht. Om diezelfde redenen, staat Finland dan weer lager gepositioneerd dan in de consensusranglijst. De goede indicatorscore voor Alcohol wordt door de aggregatiemethode niet beloond, terwijl de prestaties voor Vehicle en Trauma Care zwak zijn. Ook Estland presteert op alle belangrijke prestatie-indicatoren zwak. Portugal en Nederland hebben beiden scores op de prestatie-indicatoren die globaal gezien hoger, respectievelijk lager, liggen dan hun scores op de consensusranglijst. Vooral Portugal profiteert van de hoge gewichten die toegekend worden aan Trauma Care en Seatbelt. Aan de hand van de dataset, kon ook het Verenigd Koninkrijk verwacht worden bij de landen die te laag op de ranglijst staan, wat ook blijkt uit de landenrangschikking.

In Tabel 8 staan de toegekende gewichten wanneer men de wegingmethodes zonder consensus aggregaert. In dit geval zijn de verschillen tussen de gewichten echter meer uitgesproken dan bij 'Math.Mean'. Speed wordt aangeduid als de belangrijkste SPI, voor Alcohol en Seatbelt. DRL Law krijgt een geaggregeerd gewicht gelijk aan nul doordat de wegingmethode DEA een gewicht van 0 toekende aan de indicator (zie Tabel 5).

In de rangschikking (Tabel 12) vallen onmiddellijk een aantal verschuivingen op. Hoewel het aantal verkeerdoden in Slovenië wellicht geen 4^{de} plaats rechtvaardigt, zien we dat Slovenië, dankzij het gewicht van Alcohol en Seatbelt, niet langer de rangschikking aanvoert. Ierland profiteert van het hoge gewicht dat toegekend wordt aan Speed en voert de landenrangschikking aan. Behalve Portugal en Slovenië, krijgt ook Frankrijk, dankzij de indicatoren Speed en Seatbelt, een hogere plaats toegewezen dan in de consensusranglijst. Tot de landen die te laag gerangschikt staan behoren: Zweden, Finland, Italië en Estland. De lagere klassering van Zweden kan toegeschreven worden aan mindere prestaties voor de indicator Speed, gecombineerd met het neutraliseren van

het effect van DRL Law (zie Tabel 8). Estland en Finland worden vooral benadeeld doordat er een gewicht gelijk aan nul toegekend wordt aan DRL Law. Het feit dat er relatief weinig verschil bestaat tussen de grootte van de verschillende gewichten onderling, zorgt er bovendien voor dat de mindere scores van Finland op de prestatie-indicatoren Vehicle, Infrastructure en Trauma Care niet gecompenseerd worden met lage gewichten. Italië is dan weer één van de landen die stelselmatig lager gerangschikt worden op basis van de prestatie-indicatoren dan op basis van het aantal verkeersoden per miljoen inwoners.

3.1.2 Deviation Methods

De Deviation Methods zijn ontwikkeld en beschreven door Wang & Fan (2007) en maken, zoals de naam aangeeft, gebruik van variaties in de gegevens om gewichten toe te kennen aan elke SPI. Wang en Fan beschrijven 2 varianten, enerzijds de Absolute Deviation Method (ADM) en anderzijds de Standard Deviation Method (SDM). Voor een gedetailleerde uitleg verwijzen we naar het artikel Fuzzy Preference Relations: Aggregation and Weight determination (Wang & Fan, 2007). We zullen hieronder toch een kort inzicht geven in de methodes.

Aangezien zowel ADM als SDM geen gebruik maakt van een consensusranglijst van verkeersveiligheid, moet er een andere manier gezocht worden om de gewichten te modelleren. Bij de Deviation Methods gaat men ervan uit dat gegevens met grotere variaties, of afwijkingen, ook een grotere signaalwaarde hebben. Ofwel, om het in de woorden van Wang en Fan te zeggen: *"Based on the basic idea that bigger values of A[bsolute] D[eviation] and S[tandard] D[eviation] mean more importance to decision making, it is easily inferred that large weight should be given to those fuzzy preference relations with big absolute and/or standard deviation."* (Wang & Fan, 2007)

De beide Deviation Methods zijn in essentie opgesteld als een maximalisatieprobleem, waarbij ADM de totale absolute deviatie en SDM de totale standaarddeviatie maximaliseert door een optimale set van gewichten toe te wijzen. In de praktijk kunnen deze optimale gewichten echter simpelweg berekend worden door toepassing van een afgeleide formule. (Wang & Fan, 2007) De aanloop van beiden methodes verloopt volledig analoog. Bij het overlopen hoe we tot de resultaten gekomen zijn, zullen we aanvankelijk dan ook geen onderscheid maken tussen ADM en SDM.

Bij ADM en SDM wordt er per wegingmethode een score toegekend aan elk land, aan de hand van de verkregen gewichten. Deze score kan dan gebruikt worden om voor elke wegingmethode de landen met elkaar te vergelijken. Voor elke combinatie van landen ij wordt een voorkeur r_{ij} toegekend tussen 0 en 1. Het cijfer 0,5 duidt op een gelijke score, en dus op gelijke rangschikking, van de landen i en j volgens de betreffende wegingmethodes. Het cijfer 1 wil zeggen dat land i strikt geprefereerd wordt boven land j , terwijl 0 duidt op een strikte voorkeur van land j over land i . Alle cijfers tussenin wijzen op een meer of mindere mate van voorkeur (Wang & Fan, 2007).

De mate van voorkeur berekenen we door voor elke wegingmethode het verschil te maken tussen elke combinatie van landen. Dit verschil zit altijd in het interval $[-1;1]$ aangezien alle scores tussen 0 en 1 liggen.⁴ De voorkeur r_{ij} tussen elke combinatie van landen i en j kan dus berekend worden door het verschil uit het interval $[-1;1]$ te herleiden naar een interval $[0;1]$. Vervolgens wordt $\delta_{ik} = \sum_{j=1, j \neq i}^n (r_{ij}^{(k)} - r_{ji}^{(k)})$ berekend voor elke wegingmethode k , voor elk land i . De waarde δ_{ik} duidt op de mate van voorkeur van land i over alle andere alternatieven. (Wang & Fan, 2007)

ADM en SDM verliepen tot aan de voorgaande stap volledig analoog. Het verschil tussen beiden wordt in de volgende stap duidelijk. We zullen vanaf hier dan ook een onderscheid maken.

3.1.2.1 Absolute Deviation Method

Voor elke wegingmethode k wordt de absolute deviation $AD_k = \sum_{i=1}^n \sum_{j=1}^n |\delta_{ik} - \delta_{jk}|$ gedefinieerd. Vervolgens wordt het volgende maximalisatieprobleem opgesteld.

$$\text{Max!} \sum_{k=1}^m (\bar{w}_k AD_k)$$

⁴ De som van de gewichten is altijd genormaliseerd tussen 0 en 1. Ook de prestatie op een SPI is genormaliseerd binnen het interval $[0;1]$. Aangezien het maximum van de road safety score, $RSS = w_{alc} \times Alc + w_{speed} \times Speed + w_{seatbelt} \times Seatbelt + w_{DRL} \times DRL + w_{Vehicle} \times Vehicle + w_{Infr} \times Infr + w_{Trauma} \times Trauma$, behaald wordt wanneer elke SPI = 1 en aangezien $\sum w_i = 1$, kan RSS nooit groter dan 1 zijn. De redenering waarom de minimumwaarde van RSS 0 is, verloopt analoog.

$$s. t. \sum_{k=1}^m \bar{w}_k = 1 \text{ and } \bar{w}_k \geq 0$$

Hierbij staat $\bar{w}_k$ voor het genormaliseerde gewicht dat toegekend wordt aan wegingmethode k . Eenmaal $\bar{w}_k$ verkregen is voor elke wegingmethode k , kunnen aan de hand hiervan de geaggregeerde gewichten per SPI berekend worden.⁵

De optimale en genormaliseerde gewichten $\bar{w}_k^*$ kunnen berekend worden als

$$\bar{w}_k^* = \frac{(\sum_{i=1}^n \sum_{j=1}^n |\delta_{ik} - \delta_{jk}|)^{\frac{1}{p-1}}}{\sum_{l=1}^m (\sum_{i=1}^n \sum_{j=1}^n |\delta_{ik} - \delta_{jk}|)^{\frac{1}{p-1}}}$$

In deze formule geldt steeds dat: $p > 1$. De grootte van p bepaalt in welke mate de gewichten van elkaar verschillen: hoe groter de p -waarde, hoe kleiner de verschillen.

De geaggregeerde gewichten die men bekomt met een p -waarde van 2, zijn terug te vinden in Tabel 7 en Tabel 8. De p -waarde van 2 hebben we overgenomen uit het numeriek voorbeeld dat uitgewerkt werd door Wang en Fan (2007)

De resultaten voor de wegingmethodes met consensus (Tabel 7) liggen in de lijn van onze voorgaande vaststellingen. Het hoogste gewicht wordt toegekend aan de indicator Seatbelt, daarna volgen Vehicle en Trauma Care. DRL Law krijgt het laagste gewicht toegewezen.

De rangschikking die aan de hand van deze gewichten opgemaakt kan worden, staat in Tabel 13. De tabel toont een eerste plaats voor Slovenië gevolgd Frankrijk en Ierland. Zowel Slovenië als Frankrijk worden door ADM wat te hoog geklasseerd in vergelijking met hun positie op de lijst met verkeersdoden per miljoen inwoners. Net zoals in voorgaande situaties profiteert Frankrijk van de hoge gewichten die toegekend worden aan Seatbelt en Trauma Care en het lage gewicht van de indicator Alcohol. Ook Portugal krijgt een te hoge plaats toegekend. De landen die te laag gerangschikt staan vergeleken met het aantal verkeersdoden zijn Nederland, Finland en Estland. Finland en Estland zijn

⁵ Bijvoorbeeld: het geaggregeerde gewicht van de SPI Alcohol $\bar{w}_{alc}$ berekenen we door de wegingmethodes met consensus te aggregeren. $\bar{w}_{alc} = \bar{w}_{CPS} \times w_{Alc}^{CPS} + \bar{w}_{FPS} \times w_{Alc}^{FPS} + \bar{w}_{DEA+} \times w_{Alc}^{DEA+}$. Hierbij staat $\bar{w}_{CPS}$ voor het gewicht dat de wegingmethode CPS toegewezen krijgt door de aggregatiemethode en w_{Alc}^{CPS} voor het gewicht dat aan Alcohol is toegewezen door de wegingmethode CPS.

erg vergelijkbaar doordat beiden hoge scores halen voor de indicator Alcohol, wat een laag gewicht toegekend krijgt, terwijl de prestaties voor Vehicle en Trauma Care ondermaats zijn in vergelijking met hun score op de consensusranglijst.

Speed is de SPI die het hoogste gewicht toegewezen krijgt bij aggregatie van de wegingmethodes zonder consensus (Tabel 8). De tweede plaats is weggelegd voor Alcohol, gevolgd door Seatbelt. Vehicle is in dit geval de indicator die wordt geacht het minste invloed te hebben op verkeersveiligheid. We merken opnieuw op dat de gewichten bij de wegingmethodes zonder consensus dichter bijeen liggen dan de gewichten bij de wegingmethodes met consensus.

De rangschikking in TABEL 14 toont Slovenië voor Ierland en Frankrijk. Slovenië en Portugal zijn opnieuw de landen die te hoog gerangschikt staan. Nederland, Italië en Estland zouden dan weer een hogere plaats in de rangschikking mogen bekleden. Speed, Alcohol en Seatbelt, de drie belangrijkste indicatoren volgens deze aggregatiemethode, zijn aspecten waarvoor zowel Nederland als Italië mindere scores op halen in vergelijking met hun scores op de consensusranglijst. Beide landen hebben bovendien een algemene tendens om minder te presteren op de prestatie-indicatoren dan op basis van het aantal verkeersdoden per miljoen inwoners. Voor Estland worden de goede prestaties voor de indicator Alcohol meer dan geneutraliseerd door lage scores op de andere belangrijke prestatie-indicatoren.

3.1.2.2 Standard Deviation Method

Voor elke wegingmethode k wordt de standard deviation $SD_k = \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_{ik}^2}$ gedefinieerd. Vervolgens wordt het volgende maximalisatieprobleem opgesteld.

$$\begin{aligned} & \text{Max! } \sum_{k=1}^m (\bar{w}_k SD_k) \\ & \text{s. t. } \sum_{k=1}^m \bar{w}_k = 1 \text{ and } \bar{w}_k \geq 0 \end{aligned}$$

In deze formule staat $\bar{w}_k$ voor het genormaliseerde gewicht dat toegekend wordt aan wegingmethode k . Eenmaal $\bar{w}_k$ verkregen is voor elke wegingmethode k , kunnen aan de hand hiervan de geaggregeerde gewichten per SPI berekend worden.⁶

De optimale en genormaliseerde gewichten $\bar{w}_k^*$ kunnen berekend worden als

$$\bar{w}_k^* = \frac{\left(\frac{1}{n} \sum_{i=1}^n \delta_{ik}^2\right)^{\frac{1}{2(p-1)}}}{\sum_{l=1}^m \left(\frac{1}{n} \sum_{i=1}^n \delta_{il}^2\right)^{\frac{1}{2(p-1)}}}$$

Er geldt dat: $p > 1$. De grootte van p bepaalt in welke mate de gewichten van elkaar verschillen: hoe groter de p -waarde, hoe kleiner de verschillen.

De geaggregeerde gewichten die men bekomt met een p -waarde van 2, zijn terug te vinden in Tabel 7 en Tabel 8. De gewichten die toegewezen worden door SDM vertonen opvallend veel gelijkenissen met de gewichten die toegewezen werden door ADM.

In Tabel 7 kunnen de gewichten voor de SPI's teruggevonden worden als men de wegingmethodes met consensus aggregeert. Ook hier wordt Seatbelt geacht de meeste invloed te hebben op verkeersveiligheid, terwijl DRL Law als minst invloedrijke wordt aangeduid.

Hoewel de veranderingen in gewichten ten opzichte van ADM klein zijn, zien we toch enkele verschuivingen in de landenrangschikking in Tabel 15. In de tabel neemt Slovenië de eerste plaats in voor Frankrijk en Zweden. Slovenië, Frankrijk en Portugal worden andermaal te hoog gerangschikt. Nederland, Finland en Estland zijn, ondertussen naar goede gewoonte, wat te laag op de ranglijst terug te vinden. Gezien het feit dat er nauwelijks verschillen te noteren zijn tussen de gewichten volgens SDM en de gewichten volgens ADM, hoeft het niet te verbazen dat ook de rangschikkingen nauwelijks verschillen vertonen. De enige verschuiving die we opmerken in de rangschikking van

⁶ Bijvoorbeeld: het geaggregeerde gewicht van de SPI Alcohol $\bar{w}_{alc}$ berekenen we door de wegingmethodes met consensus te aggregeren. $\bar{w}_{alc} = \bar{w}_{CPS} \times w_{Alc}^{CPS} + \bar{w}_{FPS} \times w_{Alc}^{FPS} + \bar{w}_{DEA+} \times w_{Alc}^{DEA+}$. Hierbij staat $\bar{w}_{CPS}$ voor het gewicht dat de wegingmethode CPS toegewezen krijgt door de aggregatiemethode en w_{Alc}^{CPS} voor het gewicht dat aan Alcohol is toegewezen door de wegingmethode CPS.

SDM ten opzichte van de ranglijst volgens ADM, is dat Ierland en Zweden van plaats gewisseld hebben. Het verschil in score is echter minimaal te noemen (zie tabel 15).

Ook het aggregeren van de wegingmethodes zonder consensus (Tabel 8) door SDM resulteert in gelijkaardige gewichten als bij ADM. Dit houdt in dat Speed andermaal voor Alcohol en Seatbelt geklasseerd wordt. DRL Law en Vehicle delen de laatste plaats op de rangschikking van toegekende gewichten.

De minieme verschillen ten opzichte van de gewichten bij ADM vertalen zich ook in de landenrangschikking in Tabel 16. De koppositie is voor Slovenië, voor Frankrijk en Ierland. In dit geval worden echter enkel Portugal en, uiteraard, Slovenië te hoog ingeschat. Nederland, Estland en dit maal Italië zijn de landen die lager gerangschikt worden dan de ranglijst van het aantal verkeersdoden per miljoen inwoners suggereert. Er valt geen enkele verschuiving te noteren ten opzichte van de landenrangschikking die opgesteld kan worden aan de hand van ADM.

3.1.3 Minimization problems

We kunnen de aggregatie ook behandelen als een minimalisatie probleem. Als we ervan uit gaan dat het de bedoeling is om de ranglijst op basis van het aantal verkeersdoden per miljoen inwoners zo goed mogelijk te kunnen reconstrueren, dan zijn de beste gewichten diegene die ervoor zorgen dat de nieuwe ranglijst zo sterk mogelijk lijkt op de ideale ranglijst. We moeten dus eerst een bepaalde 'voorspellingsfout' definiëren om te meten hoe dicht we bij de ideale ranglijst kunnen geraken. Het is dan uiteraard de bedoeling om de 'voorspellingsfout' te minimaliseren. We stellen twee manieren voor om een geschikt minimalisatieprobleem op te stellen. Beide manieren zullen hieronder kort uitgelegd worden.

3.1.3.1 Minimalisatie van het scoreverschil

In deze variant werk je met de genormaliseerde waarden op de consensusranglijst. Omdat we deze lijst beschouwen als een uiting van verkeersveiligheid in een bepaald land, zou de road safety score van elk land (getal tussen 0 en 1) moeten overeenkomen met de genormaliseerde waarden in de lijst van het aantal verkeersdoden per miljoen inwoners (eveneens een getal tussen 0 en 1). Landen die hoog scoren in de Safety

Performance Index, moeten ook hoge scores halen op de consensusranglijst (landen met weinig verkeersdoden).

We kiezen ervoor om ook de scores die we verkrijgen aan de hand van de gewichten, uiteindelijk te normaliseren. Door de scores op eenzelfde schaal te brengen als de waarden op de consensusranglijst, kunnen we directe vergelijking eenvoudiger maken. Deze normalisatiestap kan echter eventueel ook achterwege gehouden worden.

Het doel van het minimalisatieprobleem is het minimaliseren van de totale afwijking tussen de consensusscore en de Indexscore. De doelfunctie is dus de som van de absolute waarde van het verschil tussen RSS en consensuswaarde voor elk land. We definiëren $RSS_i = \bar{w}_1 \times S_i^1 + \bar{w}_2 \times S_i^2 + \dots + \bar{w}_M \times S_i^M$ als de Road Safety Score voor land i . De waarde S_i^m staat voor de score die land i toegekend kreeg door wegingmethode m .⁷ $\bar{w}_m$ is het gewicht dat de aggregatiemethode toekent aan wegingmethode m . RSS_i combineert dus voor elk land de scores die gehaald zijn voor de verschillende wegingmethodes.

Eenmaal RSS_i voor elk land in de dataset berekend is, is de volgende stap het normaliseren van RSS. Na normalisatie liggen niet alleen alle waarden tussen 0 en 1, maar is het ook zo dat het land met de hoogste score, een score van 1 toegewezen krijgt en het land met de laagste score een waarde 0 krijgt. De waarden van de overige landen liggen hier tussenin. De scores van deze landen zijn dan in feite relatieve scores ten opzichte van het beste en het slechtste scorende land. Aangezien ook het aantal verkeersdoden per miljoen inwoners op dezelfde schaal wordt gezet, is een directe vergelijking eenvoudig te maken.

Aanvankelijk dient een initiële set van gewichten toegewezen te worden aan de wegingmethodes. Deze set zal vervolgens geoptimaliseerd worden door het algoritme van het minimalisatieprobleem. De initiële set kan zowel een set gewichten uit een andere aggregatiemethode als een willekeurig gekozen set, zijn. In onze toepassing zijn we van start gegaan met equal weighting, waarbij elke wegingmethode aanvankelijk hetzelfde gewicht kreeg.

⁷ Merk op dat we de gewichten uit de wegingmethodes als gegeven beschouwen.

We kunnen het minimalisatieprobleem nu wiskundig schrijven als:

$$\boxed{\text{Min! } \sum_{i=1}^N |\overline{RSS}_i - \overline{F}_i|}$$

$$\text{s.t. } \sum_{m=1}^M \overline{w}_m = 1$$

Hierbij werden de volgende symbolen gebruikt:

$\overline{F}_i$ de genormaliseerde waarde voor land i op de ranglijst van het aantal verkeersdoden per miljoen inwoners

$\overline{RSS}_i$ de genormaliseerde waarde van de road safety score (RSS_i) voor land i die gedefinieerd wordt als $RSS_i = \overline{w}_1 \times S_i^1 + \overline{w}_2 \times S_i^2 + \dots + \overline{w}_M \times S_i^M$

N het aantal landen in de selectie.

M het aantal te aggregeren wegingmethodes

Het is mogelijk om extra beperkingen toe te voegen aan de gewichten die toegekend worden aan de verschillende wegingmethodes. Hierbij wordt dan voor elke wegingmethode m een onder- en bovenlimiet wordt opgelegd aan de gewichten. Deze limieten kunnen afgeleid worden uit vorige resultaten ofwel vastgesteld worden met de hulp van experts op het gebied van wegingmethodes.

Het geaggregeerde gewicht van een SPI wordt berekend door het gewicht dat elke wegingmethode toekent aan die SPI te vermenigvuldigen met het gewicht dat het minimalisatieprobleem toekent aan de wegingmethode⁸. Deze gewichten kunnen teruggevonden worden in Tabel 7 voor de wegingmethodes met consensus en Tabel 8 voor de wegingmethodes zonder consensus.

Het aggregeren van de wegingmethodes met consensus (Tabel 7) geeft opnieuw Seatbelt als indicator met het grootste toegekende gewicht. Vehicle en Trauma Care krijgen respectievelijk het 2^{de} en 3^{de} grootste gewicht toegewezen. Opvallend is dat de SPI Infrastructure een lager gewicht meekrijgt dan bij de vorige gevallen. Het onderscheid tussen de minst belangrijke indicator DRL Law en infrastructure is echter klein.

⁸ Zie voetnoot 6.

De gewichten die toegewezen worden door de aggregatiemethode “minimalisatie van het scoreverschil” vertalen zich in de rangschikking in TABEL 17. We zien dat de aggregatiemethode andermaal een hoge rangschikking toekent aan Slovenië. Ook Frankrijk, dat volgens de rangschikking van het aantal verkeersdoden per miljoen inwoners maar een goede middenmotor (8^{ste}) is, krijgt een hoge plaats (2^{de}) toegewezen. Dit gebeurt dankzij het hoge gewicht dat toegekend wordt aan Trauma Care en – vooral – Seatbelt. Portugal wordt eveneens wat te hoog geklasseerd, terwijl Estland in de onderste regionen van de rangschikking te vinden is (12^{de} vs. 18^{de}). Zowel Estland, Portugal als Slovenië heeft een aantal verkeersdoden per miljoen inwoners dat moeilijk te verklaren valt met zijn scores op de verschillende prestatie-indicatoren.

TABEL 8, de gewichten die toegekend worden als de wegingmethodes zonder consensus geaggregeerd worden, toont opmerkelijk weinig onderscheid tussen de gewichten. DRL Law blijkt zowaar de SPI te zijn die het grootste gewicht toegewezen krijgt. Op korte afstand volgen Speed en Vehicle. Infrastructure is de minst invloedrijke SPI op gebied van verkeersveiligheid.

De landenrangschikking volgens de scores die berekend worden aan de hand van voorgaande gewichten, (Tabel 18) laat zich vrij eenvoudig beoordelen. In de toplaag van goed presterende landen op vlak van verkeersveiligheid zit 1 land dat we daar afgaand op het aantal verkeersdoden niet zouden verwachten. We hebben het hier over Slovenië. Ook onderaan de rangschikking staat met Estland een land dat volgens de rangschikking op basis van het aantal verkeersdoden per miljoen inwoners, eerder in de middenmoot verwacht mag worden. Tenslotte hebben we op de plaatsen 10 en 11 het Verenigd Koninkrijk en Nederland, die beiden zeer weinig verkeersdoden laten noteren en dus ook hoger op de ranglijst verwacht zouden worden. Daar net onder staat Portugal met een 12^{de} plaats die toch een stuk hoger is dan de Portugese 19^{de} stek op de ranglijst van verkeersdoden. Voor Nederland en Estland vallen de lagere positioneringen te verklaren doordat de prestatie-indicatoren waarop ze het best scoren, Infrastructure respectievelijk Alcohol, beiden een laag gewicht toegewezen krijgen. Slovenië en Portugal behoren tot de landen die stelselmatig een gunstigere rangschikking krijgen op basis van de prestatie-indicatoren dan op basis van het aantal verkeersdoden per miljoen inwoners verwacht kan worden. Het Verenigd Koninkrijk ondervindt vooral nadeel van het hoge gewicht dat aan DRL Law gekoppeld wordt.

3.1.3.2 Minimalisatie van het verschil in ranking

In plaats van de scores op de consensus en de Safety Performance Index te vergelijken, vergelijken we in deze variant de rangschikking van beide scores. Het minimalisatieprobleem verloopt volledig analoog met het voorgaande. Het enige verschil zit in het opstellen van de doelfunctie. Het probleem laat zich moeilijker vertalen in een wiskundige vorm dan de vorige variant, waar de scores met elkaar vergeleken werden.

$$\text{Min! } \sum_{i=1}^N |RRSS_i - RF_i|$$

$$\text{s.t. } \sum_{m=1}^M \bar{w}_m = 1$$

- RRSS_i de plaats van land i in de rangschikking op basis van de road safety scores
- RF_i de plaats van land i in de rangschikking op basis van het aantal verkeersdoden per miljoen inwoners
- N het aantal landen in de selectie
- M het aantal te aggregeren wegingmethodes
- $\bar{w}_m$ het gewicht dat aan wegingmethode m toegekend wordt

RRSS_i is afhankelijk van RSS_i en tegelijk ook van elke andere RSS die berekend werd voor elk ander land i . RRSS_i is de plaats die RSS_i inneemt in de aflopend gesorteerde rij [RSS₁ ... RSS_N] waarin de road safety score van alle N landen opgenomen is. Aangezien RSS_i berekend wordt zoals in de vorige sectie werd aangegeven, is ook RRSS_i afhankelijk van de gewichten die toegekend worden aan de wegingmethodes. Deze set van gewichten is wat het minimalisatieprobleem tracht te optimaliseren.

Het minimalisatieprobleem werkt onder de beperking dat de som van de gewichten die toegekend worden aan de wegingmethodes, gelijk moet zijn aan 1. Net zoals in het voorgaande geval, kunnen er ook hier bijkomende beperkingen toegevoegd worden.

Eens de optimale set van gewichten gevonden is, zou het verschil tussen de consensusranglijst op basis het aantal verkeersdoden per miljoen inwoners, en de rangschikking op basis van de road safety scores minimaal moeten zijn. Met andere

woorden: de som van het verschil in positie⁹ dat ieder land inneemt in de ranglijst op basis van de road safety score en de positie van dat land op de consensusranglijst, wordt geminimaliseerd. Dit is precies wat er in de doelfunctie wiskundig uitgedrukt staat.

Men zou er ook voor kunnen kiezen om in plaats van het verschil in posities te sommeren, de gekwadrateerde verschillen bij elkaar op te tellen. Op die manier worden grote afwijkingen zwaarder afgestraft. Daar staat wellicht tegenover dat er meer landen zullen zijn waarvoor er lichte verschillen tussen de posities zullen optreden. Wanneer er echter vooral naar groepen van landen gekeken wordt, en niet zozeer naar de specifieke plaats in de rangschikking, hoeft dit echter geen probleem te zijn.

In Tabel 7 kunnen de geaggregeerde gewichten voor de wegingmethodes met consensus teruggevonden worden. Hoewel er overeenkomsten zijn met de aggregatiemethode minimalisatie van het scoreverschil (MinScore), merken we toch enkele verschillen in de gewichten. Niettemin is Seatbelt ook hier de SPI dat het grootste gewicht krijgt toegekend. Vehicle en Trauma Care nemen net als bij MinScore de 2^{de} en 3^{de} plaats in. Het verschil zit echter in de grootte van de gewichten. Seatbelt en Vehicle liggen kort bij elkaar, vervolgens is er een groepje van SPI's met vergelijkbare gewichten, terwijl DRL Law en Infrastructure het laagste gewicht hebben.

In de rangschikking in Tabel 19 zie we dat ook een aggregatiemethode die specifiek ontworpen is om een zo identiek mogelijke klassering als de consensusranglijst te maken, nog een aantal voorspellingsfouten met zich meebrengt. Als we de rangschikkingen analyseren, lijkt het alsof de aggregatiemethode tweemaal twee landen verwisseld heeft. Slovenië, dat een gemiddeld aantal verkeersdoden per miljoen inwoners kent, staat helemaal bovenaan de rangschikking terwijl Nederland slechts een 10^{de} plaats krijgt toebedeeld in vergelijking met een 3^{de} plaats op de ranglijst van verkeersdoden. Onderaan de rangschikking vinden we Estland andermaal terug op een 18^{de} plaats. Op de consensusranglijst staat Estland echter gevoelig hoger genoteerd (12^{de}). Die 12^{de} plaats wordt op basis van de scores ingenomen door Portugal, dat op de ranglijst van verkeersdoden met een 19^{de} plaats aanzienlijk lager staat. We stellen vast het de landen met de grootste voorspellingsfout, die landen zijn waarvoor er een stelselmatig verschil bestaat tussen de prestaties op de indicatoren en de prestaties op de consensusranglijst.

⁹ Merk op dat het om de absolute waarde van het verschil gaat, niet om het verschil zelf.

De geaggregeerde gewichten voor de wegingmethodes zonder consensus (Tabel 8), tonen een volledig ander beeld dan de aggregatiemethode Minimalisatie dan het scoreverschil. Er zijn drie indicatoren die een relatief laag gewicht toegekend krijgen. Deze indicatoren zijn, in aflopende volgorde, Vehicle, Trauma Care en DRL Law. Speed wordt verondersteld de grootste invloed te hebben op de verkeersveiligheid en krijgt aldus het grootste gewicht toegewezen. Na Speed volgen de indicatoren Alcohol en Seatbelt.

De veranderde gewichten uiteten zich ook meteen in de landenrangschikking in Tabel 20. Slovenië vinden we pas terug op de 3^{de} plaats, wat nog altijd 9 plaatsen hoger is dan de Sloveense rangschikking op de ranglijst van het aantal verkeersdoden per miljoen inwoners. We kunnen dit verklaren door het relatief hoge gewicht dat toegewezen wordt aan de indicator Alcohol. Ook het lage gewicht voor DRL Law zorgt voor een lagere road safety score voor Slovenië en dus een minder goede klassering op de landenrangschikking. De eerste plaats wordt hier ingenomen door Ierland, op de consensusranglijst slechts een goede middenmotor (7^{de}), dat profiteert van het hoge gewicht dat aan de prestatie-indicator Speed gegeven wordt. Op de 13^{de} plaats vinden we Portugal terug. Ook Portugal is één van de landen aan wie de aggregatiemethoden stelselmatig een hogere rangschikking toekennen in vergelijking met de ranglijst op basis van verkeersdoden per miljoen inwoners. Estland en Italië worden met een 18^{de} en een 16^{de} plaats dan weer respectievelijk 6 en 7 plaatsen lager geklasseerd dan op de consensusranglijst. Speed, Alcohol en Seatbelt, de indicatoren met het hoogste gewicht, zijn voor Italië juist de drie zwakste indicatoren, wat de mindere rangschikking verklaart.

3.1.4 Correlation-based aggregation

Het gebruiken van correlaties om tot geaggregeerde gewichten te komen, houdt eveneens in dat er gebruik wordt gemaakt van een consensusranglijst. Het is de bedoeling dat de wegingmethodes die er het best in slagen om de consensus te benaderen zwaarder zullen doorwegen in van de geaggregeerde gewichten. Het verschil met de minimalisatieproblemen (zie 3.1.3) zit in het feit dat er geen optimalisatie van de gewichten plaatsvindt. De gewichten die aan de wegingmethodes toegekend worden, zijn proportioneel met de correlatie tussen de consensusranglijst en de rangschikking gebaseerd op de road safety scores, die door elke wegingmethode berekend worden.

We zullen eerst twee manieren voorstellen waarin effectief de correlatie tussen rangschikkingen wordt berekend, namelijk de consensusranglijst en de ranglijsten op basis van de scores. De eerste aggregatiemethode is een methode waarin we gebruik maken van de correlaties tussen twee rangschikkingen. De tweede methode berekent de correlaties aan de hand van relatieve posities op die rangschikkingen. Tenslotte zullen we Classical Preference Structures behandelen als aggregatiemethode.

3.1.4.1 Correlatie tussen rankings

Het aggregeren van de wegingmethodes door gebruik te maken van de correlatie tussen rangschikkingen, gebeurt door N maal de correlatie te berekenen tussen twee matrices. De eerste van deze matrices is stevast de consensusranglijst. Dit is in ons geval dus de ranglijst van het aantal verkeerdoden per miljoen inwoners. De tweede ranglijst is deze die opgesteld kan worden aan de hand van de gewichten die door wegingmethode m toegewezen werden.

De matrices nemen telkens de vorm aan van een $N \times 1$ matrix met N het aantal landen in de dataset. De getallenreeks in deze matrices duiden op de klassering van de landen in de betreffende rangschikking. Bijvoorbeeld: Als we F beschouwen als de matrix met de klassering in de consensusranglijst, dan is f_{11} de plaats die land 1 inneemt op de ranglijst van het aantal verkeersdoden per miljoen inwoners. Met land 1 hebben we het hier concreet over België dat op een 11^{de} plaats in de consensusranglijst prijkt. Dus, $f_{11}=11$. Op dezelfde manier kan er een matrix RS^m opgesteld worden voor elke wegingmethode m .

$$F = \begin{bmatrix} f_{11} \\ f_{21} \\ \dots \\ f_{N1} \end{bmatrix} \qquad RS^m = \begin{bmatrix} r_{S11} \\ r_{S21} \\ \dots \\ r_{SN1} \end{bmatrix}$$

De correlatie $\rho(F, RS^m)$ wordt berekend voor iedere wegingmethode m . Deze correlaties liggen dan later aan de basis van de gewichten die toegekend zullen worden aan de wegingmethodes.

$$\rho(F, RS^m) = \frac{Cov(F, RS^m)}{\sigma_F \sigma_{RS^m}}$$

We gaan ervan uit dat de wegingmethode van wie de rangschikking op basis van de scores de hoogste correlatie vertoont met de consensusranglijst, de best presterende wegingmethode is en dus ook het grootste gewicht moet toebedeeld krijgen. We kiezen dan ook voor gewichten, die proportioneel zijn met de toegekende correlatie. Specifiek komt dit erop neer dat we elke $\rho(F, RS^m)$ bij elkaar optellen en iedere wegingmethode m een gewicht w_m meegeven dat gelijk is aan de verhouding waarin $\rho(F, RS^m)$ bijdraagt tot de som van correlaties.

$$w_m = \frac{\rho(F, RS^m)}{\rho(F, RS^1) + \rho(F, RS^2) + \dots + \rho(F, RS^M)}^{10}$$

De gewichten die vervolgens toegewezen worden aan de indicatoren kunnen eenvoudig berekend worden uit de set van gewichten voor de wegingmethodes.¹¹ In Tabel 7 en 8 kunnen deze geaggregeerde gewichten teruggevonden worden.

Wat de wegingmethodes met consensus betreft (Tabel 7), zien we dat de indicator Seatbelt het hoogste geaggregeerde gewicht toegewezen krijgt. Na Seatbelt volgen Vehicle en Trauma Care. Met ook DRL Law helemaal onderaan, wijkt de ordening van de gewichten niet veel af ten opzichte van eerdere aggregatiemethoden.

Ondanks het feit dat de gewichten gebaseerd zijn op de correlaties met de consensusranglijst merken we in de rangschikking (Tabel 21) toch een redelijk hoog aantal landen op met een klassering die meer dan vijf plaatsen verschilt van hun klassering op de consensusranglijst. Helemaal bovenaan merken we met Slovenië, Frankrijk en Ierland drie landen op die op de ranglijst van het aantal verkeersdoden per miljoen inwoners slechts gemiddeld scoren. In de middenmoot van de rangschikking vallen vooral Nederland (9^{de}) en Finland (11^{de}) op, die respectievelijk 6 en 7 plaatsen lager geklasseerd staan dan op de consensusranglijst. Ook het als 18^{de} geklasseerde Estland laat een hogere rangschikking noteren op deze consensusranglijst (12^{de}). Voorts krijgt een land met relatief veel verkeersdoden per miljoen inwoners als Portugal (19^{de}) toch een 12^{de} plaats toegewezen. Het lage gewicht voor de indicator Alcohol is in het voordeel van landen als Slovenië, Frankrijk en Ierland terwijl het in het nadeel werkt van Estland en Finland. Op basis van de prestaties op de indicatoren, kan Nederland enkel

¹⁰ Merk op dat $\sum_{m=1}^M w_m = \frac{\rho(F, RS^1) + \rho(F, RS^2) + \dots + \rho(F, RS^M)}{\rho(F, RS^1) + \rho(F, RS^2) + \dots + \rho(F, RS^M)} = 1$

¹¹ Zie voetnoot 6

een gunstige rangschikking krijgen wanneer er veel belang gehecht wordt aan Infrastructure, de enige SPI waarvoor Nederland in de top 3 plaats neemt. Met slechts een 5^{de} plaats in de rangschikking van de indicatoren, is er hier niet aan deze voorwaarde voldaan.

De geaggregeerde gewichten voor de wegingmethodes zonder consensus (Tabel 8) liggen allemaal zeer dicht bij elkaar. De SPI Speed wordt als de meest invloedrijke indicator beschouwd. Ook Alcohol en Seatbelt krijgen een relatief hoger gewicht toegewezen. De aggregatiemethode is onbeslist als het gaat tussen DRL Law en Vehicle als minst invloedrijke SPI.

De rangschikking voor de wegingmethodes zonder consensus (Tabel 22), laat enkele verschillen noteren in vergelijking met zijn voorganger (Tabel 21). Slovenië blijft de ranglijst aanvoeren, maar deze maal zijn Ierland (2^{de}) en Frankrijk (3^{de}) van plaats gewisseld. Wanneer men het barema hanteert dat een verschil van vijf plaatsen aanvaardbaar is, komt de klassering van beide landen daarmee binnen de tolerantiezone terecht. Nederland, Portugal en Estland zijn andermaal landen die buiten de tolerantiezones vallen. Ook Italië wordt in TABEL 18 zeven plaatsen lager geklasseerd dan in de consensusrangschikking. De lage rangschikking van Nederland kunnen we toeschrijven aan het lage gewicht dat toegekend wordt aan Infrastructure en Trauma Care. Portugal en Slovenië hebben over het algemeen betere indicatorscores dan de score op basis van het aantal verkeersdoden per miljoen inwoners. Estland laat ten opzichte van de consensusranglijst dan weer overwegend mindere prestaties op basis van de indicatoren optekenen. Een combinatie van een zeer hoog gewicht voor DRL Law en Alcohol, zou voor Estland in de meest gunstige indexscore resulteren. Een hoog gewicht voor Speed, gecombineerd met een laag gewicht voor DRL Law, zijn echter nefast voor de positie van Estland in de landenrangschikking.

3.1.4.2 Correlatie tussen relatieve posities

De aggregatiemethode Correlatie tussen relatieve posities is in sterke mate gebaseerd op de Classical Preference Structures uit Hermans, Shen, Vanhoof, Wets & Brijs. (2008) Het grootste verschil is dat de gewichten voor de wegingmethodes berekend worden door middel van een correlatie. Correlatie tussen relatieve posities zal dan ook ongeveer het midden houden tussen Correlatie tussen rankings en CPS.

Correlatie tussen relatieve posities neemt exact dezelfde aanvang als CPS. Dit houdt in dat er, voor een dataset van N landen, $N \times N$ matrices opgesteld worden waarin telkens de posities van twee landen in een bepaalde rangschikking met elkaar vergeleken worden. Vervolgens wordt aangeduid of het ene land al dan niet hoger geplaatst staat dan het andere. Er wordt voor elke wegingmethode m een matrix opgesteld waarin voor elke mogelijke combinatie van landen aangegeven wordt welk land het hoogst gerangschikt staat. Ook op basis van het aantal verkeersdoden per miljoen inwoners wordt een dergelijke matrix opgesteld.

Als F de matrix is waarin de relatieve posities tussen twee landen aangegeven wordt, dan is $f_{12} = 1$ wanneer land 2 beter gerangschikt wordt dan land 1 op basis van het aantal verkeersdoden per miljoen inwoners. Wanneer dat niet het geval is, m.a.w. land 2 staat slechter of gelijk gerangschikt met land 1, dan wordt $f_{12} = 0$. In onze toepassing, waar België land 1 vertegenwoordigt, en Cyprus land 2, zou $f_{12} = 0$. België (land 1) kent immers minder verkeersdoden per miljoen inwoners dan Cyprus en is dus beter gerangschikt. Op een analoge manier wordt er voor elke wegingmethode m een gelijkaardige matrix opgesteld.

$$\hat{F} = \begin{bmatrix} f_{11} & \cdots & f_{1N} \\ \vdots & \ddots & \vdots \\ f_{N1} & \cdots & f_{NN} \end{bmatrix} \quad \hat{RS}^m = \begin{bmatrix} rs_{11} & \cdots & rs_{1N} \\ \vdots & \ddots & \vdots \\ rs_{N1} & \cdots & rs_{NN} \end{bmatrix}$$

Eenmaal de matrices opgesteld zijn voor de consensusranglijst en voor elke wegingmethode m , kan er overgegaan worden naar de volgende stap. Net zoals bij CPS, is de redenering dat een wegingmethode m goed presteert wanneer $\hat{RS}^m$ in sterke mate overeen komt met $\hat{F}$. Dit wil zeggen dat de wegingmethode m goed presteert, wanneer elk mogelijk paar landen ten opzichte van elkaar op dezelfde manier geklasseerd wordt als op de consensusranglijst. In andere woorden: als België minder verkeersdoden per miljoen inwoners telt dan Cyprus, en dus beter gerangschikt staat op de consensusranglijst, zouden we België ook graag boven Cyprus zien verschijnen in de ranglijst op basis van de road safety scores. We kunnen de mate van samenhang tussen elke matrix $\hat{RS}^m$ en $\hat{F}$ uitdrukken door de correlatie $\rho(\hat{F}, \hat{RS}^m)$ tussen de matrices te berekenen.

$$\rho(\hat{F}, \hat{RS}^m) = \frac{Cov(\hat{F}, \hat{RS}^m)}{\sigma_{\hat{F}} \sigma_{\hat{RS}^m}}$$

Deze correlaties kunnen vervolgens gebruikt worden om een gewicht w_m toe te kennen aan elke wegingmethode m . Deze gewichten worden op een identieke manier toegekend als bij Correlaties tussen rankings. Het berekenen van het gewicht per SPI, volgt uit de set van gewichten voor de wegingmethodes.

$$w_m = \frac{\rho(\hat{F}, \hat{RS}^m)}{\rho(\hat{F}, \hat{RS}^1) + \rho(\hat{F}, \hat{RS}^2) + \dots + \rho(\hat{F}, \hat{RS}^M)}$$

De geaggregeerde gewichten voor de wegingmethodes met consensus kunnen teruggevonden worden in Tabel 7. Deze gewichten blijken bijzonder dicht aan te leunen bij de gewichten die door aggregatiemethode Correlatie tussen rankings berekend werden. Ook hier wordt Seatbelt als eerste geordend, voor Vehicle en Trauma Care. DRL Law sluit het rijtje af.

De rangschikking (Tabel 23) toont opnieuw Slovenië en Frankrijk bovenaan. Onderaan de rangschikking neemt Estland andermaal de 18^{de} plaats in. Deze landen nemen op de consensusranglijst allen een positie in de middenmoot in. In het midden van deze rangschikking treffen we Finland (11^{de}) en Portugal (12^{de}). Finland is samen met Estland het land dat het meeste posities verlies boekt in vergelijking met hun rangschikking op basis van het aantal verkeersdoden per miljoen inwoners. Portugal en Slovenië zijn de landen die de meeste winst boeken. In vergelijking met de rangschikking op basis van de correlatie tussen rankings, is het enige verschilpunt dat Nederland een plaatsje opschuift ten kost van België. We kunnen hiervoor het iets hogere gewicht voor Seatbelt en het iets lagere gewicht voor Infrastructure invoeren als verklaring.

Ook in Tabel 8 merken we de grote overeenkomst op tussen Correlatie tussen relatieve posities en Correlatie tussen rankings. De geaggregeerde gewichten voor de wegingmethodes zonder consensus blijven dus dezelfde volgorde behouden als bij de vorige aggregatiemethode. DRL Law en Vehicle krijgen het laagste gewicht toebedeeld. Speed wordt als de meest invloedrijke SPI beschouwd, voor Alcohol en Seatbelt.

Gezien het geringe verschil in gewichten, vinden we in de rangschikking (Tabel 24) dan ook geen verrassingen terug. Slovenië staat op de haast traditionele eerste plaats. Nederland en Estland zijn al even traditioneel de landen die het meeste plaatsen moeten inleveren ten opzichte van de ranglijst op basis van het aantal verkeersdoden per miljoen inwoners. Ook Italië (16^{de}) vinden we onderaan terug, waar we op basis van het aantal verkeersdoden een 9^{de} klassering kunnen optekenen. Portugal wordt met een 13^{de} plaats 6 plaatsen hoger geklasseerd dan de 19^{de} positie in de consensusranglijst. Het enige verschil ten opzichte van de rangschikking op basis van de correlaties tussen rankings, is dat Oostenrijk net iets beter gerangschikt wordt dan Finland.

3.1.4.3 Classical Preference Structures

Een laatste methode die we kunnen gebruiken om te aggregeren, is Classical Preference Structures (CPS). Deze methode is ook al toegepast door Hermans, Shen, Vanhoof et al. (2008) om de gewichten van de indicatoren te bepalen. CPS maakt gebruik van de gelijkens tussen de landenrangschikking van de methodes en die van de consensus. Er wordt vertrokken van het volgende principe: Als België in de consensuslijst hoger geklasseerd staat dan land Cyprus dan moet een goede methode land België ook hoger klasseren dan land Cyprus.

We rekenen dus, aan de hand van de toegekende gewichten, voor elke wegingmethode een landenscore uit. Vervolgens worden matrices opgesteld, waarin de landen met elkaar vergeleken worden. Voor elke wegingmethode stellen we een matrix op die aangeeft of land X een hogere of lagere score behaald heeft dan land Y. Deze matrices worden dan vergeleken met een soortgelijke matrix die is opgesteld aan de hand van de consensusranglijst. Tenslotte tellen we het aantal keer dat de matrices overeenkomen. Er wordt dus geteld hoeveel keer elke mogelijke combinatie van twee landen zowel volgens de wegingmethode als in de consensus op dezelfde manier gerangschikt wordt. Dit getal delen we dan door het aantal keer dat er geen overeenkomst is tussen consensus en methode. Dit getal is het gewicht dat CPS toekent aan de wegingmethode. Methodes die een ranglijst opstellen die in sterkere mate overeenkomt met de consensusranglijst, zullen een hoger gewicht toegekend krijgen.

Er worden dus matrices opgesteld met paarsgewijze vergelijkingen tussen landen. Deze matrices komen exact overeen met diegene die we opgesteld hebben onder de sectie

3.1.4.2. We stellen voor de consensusranglijst de matrix F op, waarin wordt aangegeven of land j beter gerangschikt staat dan land i . Ook voor elke wegingmethode m wordt een soortgelijke matrix RS^m opgesteld.

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1N} \\ \vdots & \ddots & \vdots \\ f_{N1} & \cdots & f_{NN} \end{bmatrix} \quad RS^m = \begin{bmatrix} rS_{11} & \cdots & rS_{1N} \\ \vdots & \ddots & \vdots \\ rS_{N1} & \cdots & rS_{NN} \end{bmatrix}$$

Tot hiertoe verliep alles analoog aan de vorige sectie. De gewichten die toegekend zullen worden aan de wegingmethodes worden echter op een andere manier berekend. Dit komt doordat er andere criteria gebruikt worden om de overeenkomsten tussen F en RS^m te meten.

Nadat F en elke RS^m opgesteld is, worden er M maal 2 nieuwe matrices gecreëerd. In matrix O^m wordt gecontroleerd of F overeenkomt met RS^m . De tweede matrix, V^m , telt het aantal verschillen tussen F en RS^m . In matrix O^m is $o_{ij} = 1$ als $f_{ij} = rS_{ij}$, in het andere geval is $o_{ij} = 0$. In V^m wordt juist het omgekeerde gedaan. Namelijk: $v_{ij} = 1$ als $f_{ij} \neq rS_{ij}$, in de andere gevallen is $v_{ij} = 0$.

$$O^m = \begin{bmatrix} o_{11} & \cdots & o_{1N} \\ \vdots & \ddots & \vdots \\ o_{N1} & \cdots & o_{NN} \end{bmatrix} \quad V^m = \begin{bmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{bmatrix}$$

Vervolgens wordt het gewicht dat door CPS toegekend wordt aan de wegingmethode, berekend aan de hand van O^m en V^m . In beide matrices wordt het aantal "eentjes" geteld. Er wordt dus gekeken hoeveel maal F overeenkomt met RS^m en hoeveel keer er een verschil is tussen F en RS^m . Dit kan eenvoudig gedaan worden door de som te maken van alle elementen in de matrix.¹² Dit aantal duiden we dan aan met O^m en V^m .

Hoe groter de overeenkomst tussen F en RS^m , hoe groter het gewicht dat aan wegingmethode m toegekend moet worden. Het gewicht w_m wordt als volgt berekend:

$$w_m^* = \frac{O^m}{V^m}$$

¹² Er staat immers enkel 1 en 0 in de matrix. De som van al deze elementen is dus altijd gelijk aan het aantal keer dat 1 in de matrix opduikt.

Tenslotte vormen we w_m^* om zodat de som van de gewichten gelijk is aan 1. Op die manier maken we vergelijkingen tussen de verschillende aggregatiemethodes mogelijk. De gewichten voor de indicatoren kunnen weer eenvoudig berekend worden uit de gewichten voor de wegingmethodes.

$$w_m = \frac{w_m^*}{w_1^* + w_2^* + \dots + w_M^*}$$

CPS maakt gebruik van de consensusranglijst en beloont die wegingmethodes die de consensusranglijst het best kunnen reconstrueren. Een meer gesofisticeerde methode is Fuzzy Preference Structures (FPS). Bij FPS is er enige zin van nuance, die er bij CPS niet is. FPS houdt rekening met het verschil in score op de Safety Performance Index en die op de consensus ranking. (Hermans, Shen, Vanhoof, Wets, & Brijs, 2008) Daardoor kunnen verschillen tussen landen in meer schakeringen uitgedrukt worden dan de hogere-lager aanpak in CPS.

Voor de wegingmethodes met consensus, zijn de geaggregeerde gewichten terug te vinden in Tabel 7. De gewichten lijken zeer sterk op de gewichten die toegekend worden door de andere methodes gebaseerd op correlaties. We zien dus dat DRL Law het kleinste gewicht toegekend krijgt. Seatbelt is volgens CPS de belangrijkste indicator. Vehicle neemt voor Trauma Care de 2^{de} plaats in.

Wat de rangschikking (Tabel 25) betreft, wordt opnieuw voor 5 van de 20 landen een groot verschil in positie opgetekend. De rangschikking komt exact overeen met de ranglijst die gemaakt is op basis van de correlatie tussen relatieve posities. Dit houdt dus ook in dat ze nauwelijks verschilt met de ranglijst op basis van de correlatie tussen rankings.

In Tabel 8 bevinden zich de geaggregeerde gewichten voor de wegingmethodes zonder consensus. Ook hier zijn de gewichten haast identiek aan de gewichten die toegekend worden door de overige aggregatiemethodes op basis van correlaties. De indicator Speed neemt de 1^{ste} plaats in, voor Alcohol en Seatbelt. DRL Law en Vehicle krijgen het laagste gewicht toebedeeld.

Deze gewichten leiden ook hier tot exact dezelfde rangschikking als de rangschikking op basis van correlaties tussen relatieve posities.

3.2 Algemene bespreking

Het feit dat zelfs een aggregatiemethode die gebruikt maakt van de consensusrangschikking op basis van het aantal verkeersdoden per miljoen inwoners, een aantal landen een zeer verschillende rang toekent in vergelijking met hun plaats op de consensusrangschikking, bevestigt het vermoeden dat er allicht een of meerdere indicatoren ontbreken die de verkeersveiligheid mee bepalen. Aangezien het telkens dezelfde landen zijn van wie de verkeersveiligheid moeilijk te voorspellen blijkt op basis van de Safety Performance Indicators, vermindert de waarschijnlijkheid dat de weging- en aggregatiemethodes verantwoordelijk zijn voor dat falen.

Ter illustratie wordt rangschikking voor elk land per aggregatiemethode nog eens uitgezet in Tabel 27 en Tabel 28. Onder de kolom '+' wordt het aantal keer geteld dat de aggregatiemethode het betreffende land een hogere rangschikking toekent dan in de consensusranglijst. De kolom '-' duidt op het aantal keer dat er een lagere rangschikking uitgerekend wordt, terwijl '=' het aantal keer telt dat er een gelijke plaats toegekend wordt. In de laatste twee kolommen staan het aantal aggregatiemethodes dat een verschil van drie, respectievelijk vijf, plaatsen laat noteren in vergelijking met de consensusranglijst.

In Tabel 27, waar de aggregatieresultaten voor de wegingmethodes met consensus samengevat staan, merken we op dat Duitsland het enige land is dat door alle negen aggregatiemethodes telkens voorspeld wordt op dezelfde plaats als in de consensusranglijst. Oostenrijk wordt door zeven aggregatiemethodes op gelijke hoogte geklasseerd. Het is zelfs zo dat zowel Duitsland als Oostenrijk op een identieke positie als in de consensusranglijst verschijnen. Voor Duitsland kan dit verklaard worden doordat Alcohol en Speed, de indicatoren waarop de minste prestaties geleverd worden, nooit bij de drie belangrijkste indicators gerekend worden. Afgezien van de indicator DRL Law, heeft Oostenrijk onderling zeer vergelijkbare prestaties voor de verschillende indicatoren. Dit maakt dat wijzingen in de geaggregeerde gewichten weinig effect hebben op de uiteindelijke indexscore.

De belangrijkste vaststelling is echter dat de landenrangschikkingen telkens eenzelfde afwijking laten noteren ten opzichte van de consensusranglijst. De landen Cyprus, Estland, Finland, Griekenland, Hongarije, Italië, Nederland, Zweden en het Verenigd Koninkrijk worden door alle aggregatiemethodes stelselmatig lager geklasseerd dan in de consensusranglijst. Frankrijk, Ierland, Polen, Portugal, Slovenië en Spanje worden stelselmatig hoger geklasseerd. België krijgt een eerder hogere klassering ten opzichte van de rangschikking op basis van het aantal verkeersdoden per miljoen inwoners. Terwijl Denemarken door de aggregatiemethodes eerder lager geklasseerd wordt.

Estland, Portugal en Slovenië zijn de landen die voor elke aggregatiemethode een verschil laten optekenen van meer dan vijf plaatsen in vergelijking met de consensuslijst. Ook Frankrijk, Finland en Nederland zijn vaak in dat rijtje onder te brengen. Voor het Verenigd Koninkrijk komt die situatie slechts één maal voor. De aanwezigheid van Slovenië, Portugal en het Verenigd Koninkrijk, is niet echt een verrassing. In het hoofdstuk 1.2 werd al voorspeld dat deze landen moeilijk te voorspellen zouden zijn, omwille van de onduidelijke relatie tussen Safety Performance Indicators en aantal verkeersdoden per miljoen inwoners. Ook het feit dat Frankrijk en Nederland een verschillende klassering toebedeeld krijgen in vergelijking met de consensuslijst, kan verklaard worden doordat ze op de indicatoren over het algemeen relatief beter dan wel slechter voor de dag komen dan op basis van het aantal verkeersdoden per miljoen inwoners. De slechte prestaties van Finland op de indicatoren Trauma Care en Vehicle, zorgen ervoor dat de 4^{de} plaats op de consensusranglijst, niet geëvenaard kan worden in de rangschikking op basis van de road safety scores. Ook Estland scoort slecht op de bovengenoemde indicatoren en heeft het bijkomende probleem dat het ook wat de SPI Speed betreft, zeer ondermaats presteert.

Tabel 28, waar de aggregatie van de wegingmethodes zonder consensus samengevat staat, geeft een gelijkaardig beeld als TABEL 21. Net als in Tabel 27, vallen ook nu weer de stelselmatige afwijkingen ten opzichte van de consensusranglijst op, al is dat deze maal iets minder uitgesproken. Cyprus, Estland, Finland, Duitsland, Griekenland, Hongarije, Italië, Nederland en Zweden worden stelselmatig onder hun waarde geklasseerd, als we de consensusranglijst als model nemen. Tsjechië, Frankrijk, Polen, Portugal en Slovenië krijgen kregen stevast een hogere positie toegewezen.

Wanneer we de laatste kolom analyseren, zien we dat er minder landen zijn die door elke aggregatiemethode meer dan vijf posities naast hun rangschikking op basis van het aantal verkeersdoden per miljoen inwoners, geplaatst worden. Deze landen zijn Slovenië en Estland. Anderzijds zijn er wel meerdere landen waarbij dat minstens 1 maal voor komt. Portugal en Italië worden slechts één maal binnen de tolerantiegrens van 5 geklasseerd, terwijl Nederland zich 2 maal in die situatie bevindt. Voor het overige zien we ook Finland, Ierland (2 ×), Polen, Zweden en het Verenigd Koninkrijk in deze lijst opduiken. De aanwezigheid van de meeste van deze landen, hebben we hierboven al verklaard, terwijl we Polen en Italië in de sectie 'Analyse van de data' al aangeduid hadden als landen waarvan de relatie tussen de indicatoren en het aantal verkeersdoden niet altijd duidelijk was. Ierland belandde dankzij uitstekende scores op de indicatoren Speed en Vehicle tweemaal op de eerste plaats in de rangschikking, wat juist 6 plaatsen beter is dan de 7^{de} plaats op basis van het aantal verkeersdoden per miljoen inwoners. De lagere score van Zweden, het best presterende land uit de dataset in termen van het aantal verkeersdoden, voor de SPI's Speed en Infrastructure resulteerde dan weer in een lage 9^{de} plaats voor de aggregatiemethode GeoMean.

Het feit dat de landen, die we aangehaald hebben in 1.2, vaak op een totaal andere plaats gerangschikt worden dan op basis van het aantal verkeersdoden per miljoen inwoners zou verwacht worden, kan op twee manieren geïnterpreteerd worden.

Eenzijds is het positief dat deze landen op een dergelijke manier gerangschikt worden. We hebben immers vastgesteld dat het hier gaat om landen waarbij er niet meteen een duidelijke link bleek te bestaan tussen de prestaties op de verschillende indicatoren en het aantal verkeersdoden per miljoen inwoners. Doordat landen die goed presteren voor de verschillende SPI's, ook een hoge score en dus een goede plaats in de rangschikking toegewezen krijgen, ondanks een hoger dodencijfer, hoeft niet negatief te zijn. Dit zou immers betekenen dat de gewichten die toegekend worden aan de indicatoren, niet al te sterk beïnvloed worden door uitbijters of uitzonderingen. Wanneer elk land met uitstekende prestaties op vlak van de indicatoren, dat desondanks toch een hoog aantal verkeersdoden heeft, een gemiddelde score zou bekomen, zou er mogelijk een probleem kunnen zijn. De aggregatiemethoden zouden allicht de gewichten in dergelijke mate aangepast hebben om de uitzonderingen onder de landen te classificeren.

We proberen dit intuïtief te verduidelijken met een voorbeeld. Slovenië is het land waarvan het verschil in positie tussen de consensusranglijst en de ranglijst op basis van de road safety scores het grootst is. Om een set gewichten te bekomen waarbij Slovenië ook werkelijk een 12^{de} plaats zou bezetten op basis van de road safety scores, moet een methode ontwikkeld worden die specifiek daarop gericht is. We zouden bijvoorbeeld een minimalisatieprobleem kunnen opstellen dat het verschil tussen de 2 rangnummers van Slovenië minimaliseert. Een andere mogelijkheid is om een extra beperking toe te voegen aan onze bestaande methodes waarbij de eindpositie van Slovenië 12 moet zijn of in een bepaald interval moet vallen. Het is duidelijk dat de methodes dan zouden overfitten op de case Slovenië.

Anderzijds kan men bedenkingen hebben bij een rangschikking op basis van road safety scores, die eigenlijk verondersteld wordt de werkelijke rangschikking te zijn, waarbij een land als Slovenië als verkeersveiligste land beschouwd wordt. De kopplaats van Slovenië staat immers in contrast met de matige scores in vergelijking met de andere landen qua verkeersdoden per miljoen inwoners.

Er is dus een zekere afweging die men moet maken bij het evalueren van de resultaten. Onderzoekers die vooral geïnteresseerd zijn in de grootte en volgorde van de gewichten en dus de onderliggende oorzaken van verkeersveiligheid beter willen inschatten hechten minder belang aan de rangschikkingen.

3.3 Andere mogelijke aggregatiemethodes

Om het overzicht compleet te maken, dienen we nog te vermelden dat er nog andere aggregatiemethodes mogelijk zijn. Zoals we al aangehaald hebben, kan een aggregatiemethode beschouwd worden als een wegingmethode, toegepast op wegingmethodes. Daaruit volgt dus logisch dat ook elke wegingmethode die we geaggregeerd hebben, zou gebruikt kunnen worden als aggregatiemethode.¹³ We hebben ons echter beperkt tot de methodes beschreven in deze sectie.

¹³ FA, AHP, BA, DEA, FPS en afgeleide vormen hiervan, zouden ook gebruikt kunnen worden als aggregatiemethodes.

3.4 Conclusies

Wat de landenrangschikkingen betreft, leveren de aggregatiemethodes de te verwachten resultaten af. Het blijkt dat vooral de landen die een moeilijk vast te stellen relatie tussen indicatoren en aantal verkeersdoden hebben, ten opzichte van de consensusranglijst een verschillende positie toegewezen krijgen. Aangezien de plaats in de rangschikking bepaald wordt door de prestaties van de landen op de SPI's, mogen de resultaten als bevredigend gezien worden. Het is immers onvermijdelijk dat landen met goede prestaties op de Safety Performance Indicators een hoge score krijgen en landen met mindere prestaties voor de SPI's ook mindere scores krijgen, ongeacht het aantal verkeersdoden.

Over de geaggregeerde gewichten kunnen minder eenvoudig conclusies afgeleid worden. Het is zeer moeilijk om de correctheid van de gewichten te achterhalen. Indien we ons baseren op de landenrangschikkingen, die opgesteld worden aan de hand van de geaggregeerde gewichten, mogen we er vertrouwen in hebben dat de uitkomsten van de gewichten stroken met de realiteit. Als we de landen waarbij we in sectie 1.2 geen duidelijk verband konden vaststellen tussen consensusranglijst en prestatie-indicatoren, buiten beschouwing laten, kan de consensusranglijst immers redelijk goed gereconstrueerd worden.

We merken bovendien dat er een bepaalde overeenkomst bestaat tussen de gewichten die de verschillende aggregatiemethodes toewijzen. Elke SPI krijgt van elke aggregatiemethode ongeveer hetzelfde gewicht toegewezen. Vooral de aggregatiemethodes GeoMean en de beide minimalisatieproblemen laten geregeld gewichten optekenen die licht afwijken van de geaggregeerde gewichten van de overige aggregatiemethodes. Die overige methodes blijken merkwaardig eensgezind te zijn wat de gewichten betreft. Daarbij is het vooral opvallend dat zowel het verschil tussen de deviation methods als het verschil tussen de methodes die werken op basis van correlaties, haast verwaarloosbaar is (zie Tabel 7 en 8). De beide varianten van het minimalisatieprobleem kennen voor elke SPI dan weer wel gewichten toe die onderling verschillend zijn.

4 Robuustheid

Na een beschrijving van de verschillende aggregatiemethodes en hun resultaten, bespreken we in dit hoofdstuk de evaluatie van deze methodes.

4.1 Waarom testen op robuustheid?

De robuustheid van de aggregatiemethoden geeft ons een belangrijke indicatie van de betrouwbaarheid van de methodes en de kwaliteit van de resultaten. Zoals we eerder al aangehaald hebben, passen we de aggregatiemethodes onder een aantal assumpties toe. De eerste assumptie is dat de 7 aspecten van verkeersveiligheid – alcohol & drugs, snelheid, de autogordel, motorvoertuigverlichting overdag, voertuig, weginfrastructuur en nazorg – goed weergegeven worden door de Safety Performance Indicators in Tabel 1. We gaan er eveneens van uit dat het aantal verkeersdoden per miljoen inwoners een goede weergave is van de verkeersveiligheid in een land. De tweede assumptie is dat de waarden die we terugvinden in Tabel 1, een correcte weergave zijn van de werkelijkheid.

We zijn er ons van bewust dat de realiteit lichtjes verschilt van de dataset. Het belang van een robuuste aggregatiemethode is dat lichte wijzigingen in de dataset geen significante effecten hebben op de resultaten. Indien de door ons gebruikte aggregatiemethodes robuust zijn, zullen ze erin slagen om zelfs bij een licht gewijzigde dataset, ongeveer dezelfde gewichten toe te wijzen aan de indicatoren, wat dan resulteert in een gelijkaardige landenrangschikking. Robuuste aggregatiemethodes nemen het probleem van onzekerheid over de dataset voor een deel weg. Aangezien kleine wijzigingen in de dataset geen significante invloed hebben op de gewichten en daaruit volgende landenrangschikking, is het niet noodzakelijk om een foutloze dataset te hebben. Voor robuuste aggregatiemethodes zijn gegevens die de werkelijkheid *benaderen* voldoende om “correcte” resultaten te krijgen.

4.2 Werkwijze

De gewichten die toegekend worden door de verschillende wegingmethodes beschouwen we als gegeven en zullen dus ook geen veranderingen ondergaan. Dit impliceert meteen dat de aggregatiemethodes die gebruik maken van gemiddeldes, MathMean en GeoMean, altijd dezelfde resultaten zullen geven. Er zou immers telkens het gemiddelde berekend

worden van een onveranderende gegevensreeks. Deze aggregatiemethodes kunnen dus niet beïnvloed worden door veranderingen in de data en produceren altijd dezelfde resultaten. We kunnen deze methodes dan ook robuust noemen.

Omdat niet alle aggregatiemethodes even onderhevig zijn aan veranderingen, zullen we ons concentreren op de methodes die in de meeste opzichten van de gegevens gebruik maken. De minimalisatieproblemen proberen elke informatiewaarde die in de beschikbare gegevens vervat zit, optimaal te benutten. Eerst en vooral wordt een consensusranglijst gebruikt om de resultaten te modelleren. Bovendien hebben zowel de waarden in de dataset als de landen die er deel van uit maken, een invloed op de uitkomst. Tenslotte is het zo dat de wegingmethodes op een optimale manier gecombineerd worden, met een bepaalde doelfunctie in het achterhoofd.

Binnen de minimalisatieproblemen bestaan er dus drie bronnen van variatie, aangenomen dat de resultaten van de wegingmethodes als gegeven mogen opgevat worden. Deze drie variatiebronnen zijn: de consensusranglijst, de landen die aanwezig zijn in de dataset en de exacte waarden uit de dataset. Veranderingen in één van deze elementen zullen in ieder geval een invloed hebben op de resultaten. Dit is niet zo voor alle aggregatiemethodes. De aggregatiemethodes die werken met gemiddeldes zijn totaal niet onderhevig aan voorgenoemde veranderingen. De Deviation Methods, ADM en SDM, maken dan weer geen gebruik van een consensus.

Het grote verschil tussen de correlatiemethodes (inclusief CPS) en de minimalisatieproblemen, is de manier waarop de gewichten bepaald worden. Bij de correlatiemethodes is er een bepaalde, voorgemaakte, formule aan de hand waarvan de gewichten worden bepaald. Er is dus geen optimalisatieproces waarbij de gewichten optimaal gefit worden aan de dataset. De kans op overfitting (Witten & Frank, 1999) is dus groter bij de minimalisatieproblemen. Daarom focussen we op deze aggregatiemethodes.

We zullen in het volgende hoofdstuk drie deelvragen behandelen. Daarbij zullen we telkens onderzoeken welk effect een kleine aanpassing in hetzij consensusranglijst, hetzij landen in de dataset, hetzij de indicatorwaarden in de dataset, zal hebben op de resultaten. De eerste stap zal zijn dat we de kleine veranderingen aanbrengen, vervolgens worden de aggregatiemethodes opnieuw toegepast. De volgende stap is dan

dat we de verandering in de resultaten kwantificeren en tenslotte bespreken. De maatstaven waarmee we de veranderingen in de resultaten, en dus de robuustheid van de methodes, zullen meten, staan beschreven in de volgende sectie.

4.3 Maatstaven van robuustheid

Uiteraard hebben we een aantal maatstaven nodig waarmee we de robuustheid van de aggregatiemethodes kunnen bepalen. Aangezien we robuustheid meten aan de hand van het verschil in resultaten onder invloed van licht verschillende omstandigheden, hebben we een coëfficiënt nodig die deze mate van verschil kan vervatten. We kiezen ervoor om de omgekeerde redering te maken en juist de mate van overeenkomst te beschrijven. De mate van verschil kan hieruit vervolgens rechtstreeks afgeleid worden.

De aggregatiemethodes genereren in de eerste plaats gewichten die toegewezen worden aan de verschillende wegingmethodes. Vanuit deze gewichten kunnen dan rechtstreeks de gewichten die toegekend worden aan de Safety Performance Indicators berekend worden. Aangezien we de uitkomsten van de wegingmethodes in de analyse als gegeven beschouwen, is het voldoende om enkel de robuustheid van de gewichten aan de SPI's te onderzoeken.

Ten tweede kunnen er voor elk land in de dataset road safety scores berekend worden aan de hand van de gewichten die toegekend werden aan de SPI's en de prestaties van elk land voor deze indicatoren. Op basis van deze road safety scores kan dan een rangschikking opgemaakt worden.

We hebben dus de beschikking over vier vormen van output die we kunnen onderzoeken op robuustheid. Ten eerste kunnen we ons concentreren op de grootte van de toegewezen gewichten. Het is echter ook mogelijk om ons vooral te richten op de *volgorde* van deze gewichten, en niet zozeer op de precieze grootte ervan. We achten met andere woorden de vraag welke SPI het hoogste (resp. 2^{de}, 3^{de} etc.) gewicht toegewezen krijgt, belangrijker dan de vraag welk gewicht nu juist aan elke SPI toegewezen wordt. De derde vorm die we kunnen onderzoeken, zijn de road safety scores van elk land in de dataset. In de vierde, en laatste, vorm houden we enkel rekening met de rangschikking van de landen.

Om twee rangschikkingen met elkaar te vergelijken, bedienen we ons van twee verschillende correlatiecoëfficiënten. De eerste is **Spearman's rank correlation coefficient** (Spearman, 1904), uitgedrukt als ρ . Aangezien er nooit sprake is van een gelijke rangschikking¹⁴, kan hiervoor de volgende formule toegepast worden:

$$\rho = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)}$$

In deze formule staat n voor het aantal elementen in de dataset. De plaats in de rangschikking van de waarde X_i (resp. Y_i) wordt gegeven door x_i (resp. y_i).

De tweede correlatiecoëfficiënt die we kunnen gebruiken, is de **Kendall tau correlation coefficient** (Kendall, 1938), genoteerd als de Griekse letter τ (tau). De formule die we gebruiken ziet er als volgt uit:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

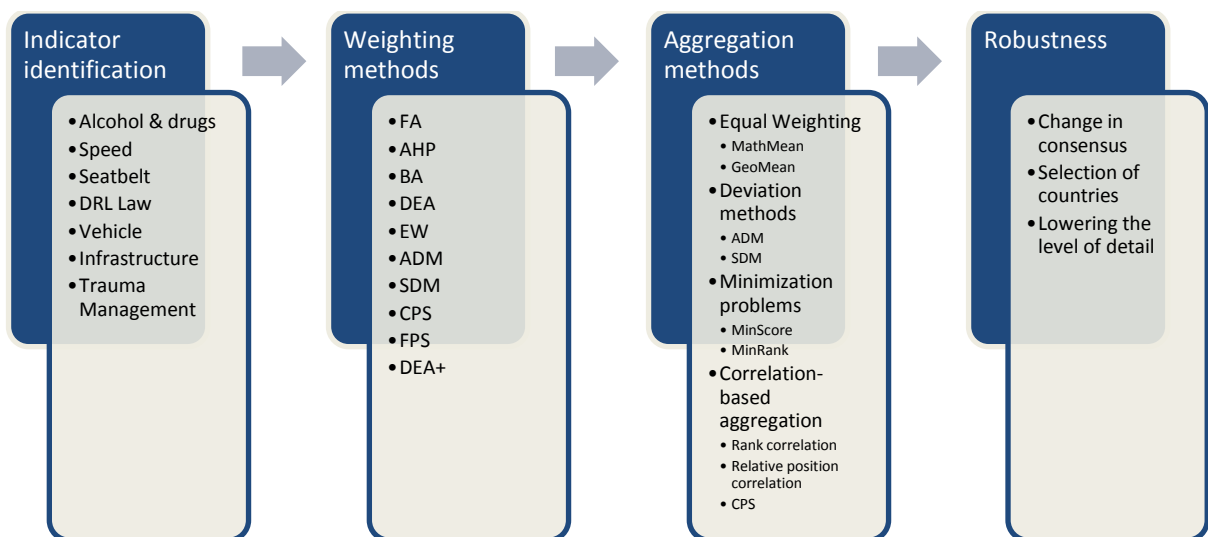
Hierbij staat n voor het aantal elementen in de dataset. De waarden n_c en n_d stellen het aantal concordante (resp. discordante) paren voor. Een paar mag concordant genoemd wanneer ze zowel in de ene als in de andere rangschikking, ten opzichte van elkaar op dezelfde manier gerangschikt worden. D.w.z. X_i en Y_i zijn concordant wanneer X_i in zowel de ene als in de andere rangschikking, hetzij boven hetzij onder Y_i geplaatst wordt.

Elke correlatie coëfficiënt is een waarde in het interval $[-1; 1]$. Hoe hoger de waarde, hoe groter de correlatie tussen de beide gegevensreeksen. Hierbij staat 1 voor een perfecte overeenkomst tussen beide. Een waarde van 0 duidt op geen enkele correlatie. Wanneer de gegevensreeksen elkaars complete spiegelbeeld zijn, wordt de waarde van de correlatiecoëfficiënt gelijk aan -1.

¹⁴ Er is altijd een verschil tussen de geaggregeerde gewichten van de SPI's. Ook bestaan er geen twee landen die exact dezelfde road safety score hebben, ook al is dit niet altijd te zien bij afrondingen tot op twee decimalen.

5 Testen van de robuustheid van de aggregatiemethodes

In dit hoofdstuk zullen we het patroon volgen dat we al in de vorige sectie voorgesteld hebben. We beginnen met de deelvraag die het onderwerp vormt van de sectie. Dan verklaren we de werkwijze om vervolgens het antwoord op de vraag te formuleren. Dit antwoord zal eruit bestaan dat we de verschillen in resultaten tussen de gewijzigde en de oorspronkelijke situatie zullen vergelijken. De oorspronkelijke resultaten kunnen teruggevonden worden in hoofdstuk 3. De gewijzigde situatie zal geval per geval uitgelegd worden. We zullen niet uitweiden over de precieze gewijzigde gewichten en rangschikkingen, maar ons concentreren op de veranderingen ten opzichte van de oorspronkelijke situatie.



Figuur 1 e: Overzicht van het onderzoek

5.1 Robuustheid bij verandering van consensusranglijst

In deze sectie onderzoeken we of de aggregatiemethodes 'Minimalisatie van de rankingverschillen (MinRank) en Minimalisatie van de scoreverschillen (MinScore), robuust mogen bevonden worden wanneer er veranderingen optreden in de consensusranglijst.

De ranglijst op basis van het aantal verkeersdoden is immers slechts één mogelijke vorm van consensus. Er kunnen een veelvoud aan alternatieve ranglijsten en cijfers gebruikt worden als consensus. Het is dus aannemelijk dat er discussie kan optreden tussen

experts over welke ideale representatie nu juist het beste is. De ideale consensus wordt immers geacht de realiteit in zake verkeersveiligheid te weerspiegelen en de "werkelijke rangschikking" van de landen weer te geven. Omdat deze werkelijke rangschikking niet gekend is, of misschien zelfs verschillend gedefinieerd kan worden, bestaan er ook geen harde criteria om de beste consensusranglijst te selecteren.

Juist omdat de consensusranglijst onderwerp van discussie kan zijn, is het belangrijk dat de verschillende voorgestelde consensusranglijsten onderling verwisselbaar zijn. De aggregatiemethodes zouden dus voor elke consensusranglijst die gebruikt wordt om de gewichten op te modelleren, vergelijkbare resultaten moeten opleveren. Dit maakt discussies over de ideale representatie van secundair belang.

Bovendien moeten we er rekening mee houden dat de mogelijkheid bestaat dat geen enkele consensusranglijst een exacte weergave is van de "werkelijke" rangschikking van de landen op basis van verkeersveiligheid. Ook daarom is de onderlinge inwisselbaarheid van verschillende consensusranglijsten van belang. Hoe meer de consensusranglijsten, die allen in theorie de werkelijke situatie sterk benaderen, voor elkaar inwisselbaar zijn, hoe groter de kans dat ook die werkelijke rangschikking gelijkaardige resultaten zou genereren.

We zouden dus graag de robuustheid van de aggregatiemethodes kennen, wanneer een andere consensusranglijst gebruikt wordt. We stellen dan ook de volgende deelvraag.

Wat is de impact op de gekozen gewichten en de daaruit volgende landenrangschikking wanneer de overeenkomst met een andere consensus als criterium gebruikt wordt?

Deze deelvraag zullen we oplossen door de aggregatiemethodes de wegingmethodes opnieuw te laten aggregeren, met tweemaal een andere landenrangschikking als consensus. De twee ranglijsten worden, samen met de resultaten van de test, apart behandeld. Als alternatieve ideale representatie maken we gebruik van:

1. De rangschikking van de landen op basis van het aantal verkeersdoden per miljoen afgelegde kilometers
2. De landenrangschikking op basis van het aantal verkeersdoden per 100.000 personenwagens

Het opzet is dat we de consensusranglijst in de doelfunctie van de aggregatiemethodes veranderen. Het is belangrijk daarbij op te merken dat er voor het overige geen wijzigingen aangebracht zullen worden. We blijven dus werken met dezelfde dataset waarmee we de oorspronkelijke resultaten berekenden. Dit houdt onder andere ook in dat ook de resultaten van de wegingmethodes met consensus niet gewijzigd werden. We hebben deze wegingmethodes dus niet opnieuw laten lopen met de alternatieve ideale representatie. Dit onderzoek concentreert zich immers voornamelijk op de aggregatiemethodes en beschouwt de wegingmethodes als gegeven.

5.1.1 Aantal verkeersdoden per miljoen afgelegde kilometers

5.1.1.1 De nieuwe consensus

Het aantal verkeersdoden per miljoen afgelegde kilometers kan, net als het aantal verkeersdoden per miljoen inwoners, beschouwd worden als een goede consensusranglijst. In deze consensus kiest men ervoor om het aantal verkeersdoden per land onderling vergelijkbaar te maken door ze te delen door het totale aantal kilometers dat in dat land door een personenwagen afgelegd werd.

De redenering om deze consensus te kiezen boven andere mogelijkheden, kan zijn dat er mogelijk geen duidelijk causaal verband is tussen het aantal inwoners in een land en het aantal verkeersdoden. Tussen het aantal afgelegde kilometers en het aantal verkeersdoden, kan wel een causale relatie vermoed worden. Hoewel we geen cijfers gebruiken om deze hypothese te staven, gaan we uit van de stelling: "Hoe meer kilometers er in een bepaald land afgelegd worden op de openbare wegen, hoe groter de kans dat er een (dodelijk) ongeval gebeurt."

Onder de voorstanders van deze consensus bevinden zich allicht de landen met weinig inwoners, maar relatief veel verkeersactiviteit en veel ongevallen. Zij zouden in deze consensusranglijst beter scoren dan op basis van het aantal verkeersdoden per miljoen inwoners. De landen met een grote populatie, maar relatief weinig verkeersactiviteit, zouden in deze rangschikking benadeeld worden ten opzichte van onze originele consensus.

5.1.1.2 Aanbrengen van de veranderingen

We passen in de aggregatiemethodes de consensusranglijst aan en maken alle nodige berekeningen opnieuw. De geaggregeerde gewichten voor de aggregatiemethodes MinRank en MinScore kunnen teruggevonden worden in Tabel 29 en Tabel 31 (wegingmethodes met consensus) & Tabel 30 en Tabel 32 (wegingmethodes zonder consensus). De nieuwe landenrangschikkingen die we op basis van deze gewichten kunnen opstellen, zijn terug te vinden in Tabel 33 en Tabel 34 voor MinRank en Tabel 35 en Tabel 36 voor MinScore. In Tabel 37 en Tabel 38, ten slotte, staan de geaggregeerde gewichten voor de andere aggregatiemethodes. Het spreekt voor zich dat deze gewichten niet verschillen ten opzichte van het origineel voor de aggregatiemethodes die geen gebruik maken van een consensusranglijst. De gewichten van MathMean, GeoMean, ADM en SDM zijn dus onveranderd gebleven.

5.1.1.3 Robuustheid

De correlatiecoëfficiënten in Tabel 45 beschrijven de robuustheid van de geaggregeerde gewichten wanneer men enkel de wegingmethodes met consensus aggregeert. Hoe dichter de correlatiecoëfficiënt de waarde 1 benadert, hoe groter de robuustheid van de aggregatiemethode. Hoge correlaties betekenen immers dat de gewichten nauwelijks veranderd zijn ten opzichte van de oorspronkelijke situatie.

We merken dat er een aanzienlijk verschil is tussen de aggregatiemethodes MinScore en MinRank. De methode MinRank blijkt relatief ongevoelig voor de precieze keuze van de consensusranglijst. Alle correlatiecoëfficiënten voor MinRank duiden op een sterke samenhang tussen de resultaten uit de oorspronkelijke en de aangepaste situatie. De correlatiecoëfficiënten voor MinScore staan daarmee in schril contrast en duiden erop dat er geen enkele correlatie merkbaar is.

De correlatiecoëfficiënten bij de wegingmethodes zonder consensus (Tabel 46) geven eenzelfde beeld. Deze keer zijn de verschillen zelfs nog meer uitgesproken. MinRank blijkt zeer robuust te zijn, met extreem hoge correlatiecoëfficiënten die de waarde één zeer dicht benaderen. Voor MinScore variëren de coëfficiënten tussen geen enkele merkbare correlatie tot zelfs licht negatieve correlaties.

Wanneer we de correlaties berekenen tussen de oorspronkelijke landenrangschikkingen en de landenrangschikkingen uit de aangepaste situatie, bekomen we de coëfficiënten uit Tabel 53 en Tabel 54. Zowel de aggregatie van de wegingmethodes met consensus als de aggregatie van de wegingmethodes zonder consensus produceert robuuste resultaten. Zowel op basis van de scores als puur op basis van rangschikking treden er erg weinig verschillen op ten opzichte van de oorspronkelijke situatie wanneer er een andere consensusranglijst gebruikt wordt.

5.1.2 Aantal verkeersdoden per 100.000 personenwagens

5.1.2.1 De nieuwe consensus

Een andere mogelijke consensusranglijst die gekozen kan worden, is de landenrangschikking op basis van het aantal verkeersdoden per 100.000 personenwagens die in het betreffende land ingeschreven zijn. Ook voor deze consensus kunnen argumenten gevonden worden. Het lijkt een redelijke aanname om te veronderstellen dat het aantal verkeersdoden in een land toeneemt met het aantal voertuigen. Immers, hoe meer wagens zich er op de openbare wegen bevinden, hoe groter de kans op een (dodelijk) ongeval.

5.1.2.2 Aanbrengen van de veranderingen

Ook hier zullen we de consensusranglijst die gebruikt wordt door de aggregatiemethoden invullen door de ranglijst op basis van het aantal verkeersdoden per 100.000 personenwagens te gebruiken. Vervolgens onderzoeken we het effect dat de veranderingen hebben op de resulterende gewichten en landenrangschikking.

5.1.2.3 Robuustheid

De correlatiecoëfficiënten in Tabel 47 en Tabel 48 tonen een volledig ander beeld dan bij de vorige verandering in consensus. Door te kiezen voor het aantal verkeersdoden per 100.000 wagens als consensusranglijst, verandert er nauwelijks iets aan de gekozen gewichten. Alle coëfficiënten wijzen op een bijna perfecte positieve correlatie tussen de geaggregeerde gewichten uit de originele situatie, en de geaggregeerde gewichten uit de aangepaste versie. We zien hetzelfde patroon opduiken zowel bij het aggregeren van de

wegingmethodes met consensus als bij het aggregeren van de wegingmethodes zonder consensus.

Aangezien er nauwelijks verschillen op te tekenen vielen tussen de gewichten, is het logisch dat ook de landenrangschikking nauwelijks wijzigt. In Tabel 55 en Tabel 56 worden deze vermoedens bevestigd. Alle correlatiecoëfficiënten benaderen de waarde 1. Gebaseerd op de resultaten uit de laatste 4 genoemde tabellen, kunnen we stellen dat zowel MinScore als MinRank robuuste gewichten genereren en dus ook robuuste rangschikkingen samenstellen.

5.1.3 Conclusie

Wanneer we een algemene beschouwing maken van het effect van een wijziging in de consensusranglijst op de uiteindelijke geaggregeerde gewichten, komen we tot de volgende conclusies.

Het is opvallend dat de correlaties voor MinScore zo sterk verschillen tussen beide situaties. Wanneer het aantal verkeersdoden per miljoen afgelegde kilometers als consensusranglijst gebruikt wordt, veranderen de geaggregeerde gewichten aanzienlijk ten opzichte van de originele situatie waar het aantal verkeersdoden per miljoen inwoners de ideale representatie vormt. Met het aantal verkeersdoden per 100.000 wagens als consensusranglijst, worden er dan weer wel zeer gelijkaardige resultaten geproduceerd.

MinScore blijkt gevoeliger te zijn voor de consensusranglijst dan MinRank. Dit is op zich allerm minst verwonderlijk. MinScore maakt immers gebruik van de genormaliseerde waarden uit de consensusranglijst, terwijl MinRank enkel de plaats op de rangschikking in rekening neemt. Het is dus perfect mogelijk dat twee potentiële consensusranglijsten de landen uit de dataset op eenzelfde manier rangschikken, terwijl ze toch een andere waarde toegewezen krijgen. De variaties tussen de scores kunnen met andere woorden hoger oplopen dan de variatie in de rankings.

We mogen dus stellen dat MinScore zeer afhankelijk lijkt te zijn van de gebruikte consensusranglijst. We kunnen deze methode op dit vlak dan ook niet als robuust beschouwen. De geaggregeerde gewichten die MinScore voortbrengt zijn immers

onvoorspelbaar wanneer er wijzigingen aangebracht worden in de ideale representatie. Het lijkt ons dan ook beter te kiezen voor de aggregatiemethode MinRank wanneer er twijfels zijn over de te gebruiken consensusranglijst. MinRank blijkt immers veel minder gevoelig te zijn voor de keuze van ideale representatie. Dit maakt dat eventuele foute of minder goede keuzes van consensusranglijst, minder invloed zal hebben op de resultaten.

Wat veranderingen in de consensusranglijst betreft, is het logisch dat de methodes erg afhankelijk zijn van de keuze van consensus. Aangezien alle consensusranglijsten een representatie van verkeersveiligheid moeten voorstellen, is enige overeenkomst tussen de varianten uiteraard een vereiste. Het blijkt duidelijk dat de aggregatiemethode MinRank op dit vlak veel robuuster is dan MinScore. Dit zijn logische resultaten aangezien MinScore zich op scores concentreert terwijl MinRank vooral focust op rangschikkingen. Tussen de scores is er meer variatie mogelijk dan tussen rangschikkingen waardoor de geaggregeerde gewichten van MinScore gevoeliger zijn voor veranderingen dan MinRank. Wanneer er twijfels zijn over de te gebruiken ideale representatie, raden we dan ook aan om gebruik te maken van de aggregatiemethode MinRank.

5.2 Robuustheid bij inkrimping van het aantal landen in de dataset

De tweede vorm van robuustheid die we zullen testen, is de robuustheid van de aggregatiemethodes wanneer de landen in de dataset veranderen. We zullen enkele landen niet opnemen in de dataset waarop de aggregatiemethoden werken. We kiezen ervoor om het aantal landen in de dataset van 20 terug te brengen tot 15. Indien de aggregatiemethodes robuuste geaggregeerde gewichten genereren, zouden de nieuwe gewichten niet teveel mogen verschillen van de oorspronkelijke.

In het vakgebied van Data Mining wordt vaak de omgekeerde redenering gebruikt om methodes te testen. De normale procedure is dat je een deel van de beschikbare dataset buiten beschouwing laat. Vervolgens laat je de methodes werken op het andere deel. Eens er resultaten berekend zijn, test je deze uit op het deel van de dataset dat je achterwege had gelaten (Witten & Frank, 1999). Concreet zou dit betekenen dat we een aantal landen uit de dataset opzij zouden moeten schuiven. Vervolgens zouden we de geaggregeerde gewichten berekenen op de overgebleven landen. Om deze gewichten dan te testen, zouden we ze toepassen op de weggelaten landen en nagaan of de scores

en rangschikking die deze landen zouden krijgen vergelijkbaar waren met de consensusranglijst. Aangezien we slechts de beschikking hebben over 20 cases in onze dataset, hebben we ervoor gekozen om deze procedure *niet* te volgen.

Wanneer het weglaten van een aantal landen uit de dataset geen significante invloed heeft op de geaggregeerde gewichten, betekent dit dat de methodes niet overfitten op één bepaalde case uit de dataset. De methodes mogen dan als robuust gezien worden. De deelvraag die we in deze sectie zullen behandelen luidt:

Heeft het selecteren van een aantal landen uit de dataset een invloed op de gekozen gewichten en de daaruit volgende landenrangschikking?

5.2.1 Werkwijze

Om de test zo efficiënt en zinvol mogelijk te maken, zullen we niet werken met een willekeurig gekozen steekproef van landen maar met een specifieke dataset. We beschouwen niet langer die landen waarvoor de aggregatiemethode MinRank het minst goed de plaats op de consensusranglijst kon reconstrueren. Dit zouden de landen moeten zijn waarvoor de samenhang tussen SPI's en aantal verkeersdoden niet duidelijk was. Het zijn ook juist deze landen die de gewichten in een bepaalde richting zouden kunnen beïnvloeden. We laten Estland (18^{de} vs. 12^{de}), Frankrijk (3^{de} vs. 8^{ste}), Nederland (10^{de} vs. 3^{de}), Portugal (12^{de} vs. 19^{de}) en Slovenië (1^{ste} vs. 12^{de}) buiten beschouwing.

Door te kiezen voor deze landen, kunnen we erop vertrouwen dat de aggregatie op basis van een willekeurig gekozen dataset van landen, steeds minder grote verschillen zal geven dan in deze opstelling. Deze vermoedens werden bevestigd nadat we dezelfde proef deden met een willekeurige dataset van 15 landen. Daarbij werd op basis van een willekeurige steekproef de landen België, Denemarken, Oostenrijk, Spanje en Tsjechië geweerd uit de dataset.

5.2.2 Robuustheid

Na het opnieuw aggregeren van de wegingmethodes met consensus, onder de aangepaste omstandigheden, bekomen we de geaggregeerde gewichten in Tabel 49. Zowel MinScore als MinRank slaagt erin ongeveer dezelfde geaggregeerde gewichten

voort te brengen als in de oorspronkelijke situatie. Op basis van deze tabel zouden we de beide aggregatiemethodes robuust kunnen noemen.

Tabel 50 geeft echter een ander beeld. MinScore blijkt erg robuust te zijn. De correlatiecoëfficiënten wijzen op een perfecte overeenkomst tussen de oorspronkelijke en de gewijzigde situatie. Die perfecte correlatie moet echter niet helemaal letterlijk genomen worden, het gaat hier om afrondingen van erg hoge coëfficiënten. In tegenstelling tot MinScore, laat MinRank geen erg hoge correlaties optekenen. MinRank blijkt dus minder robuust te zijn dan MinScore wanneer er met een selectie uit de dataset gewerkt wordt. De lage correlatiecoëfficiënten voor MinRank tussen de oorspronkelijke en de aangepaste situatie hoeven er echter niet toe te leiden dat MinRank overboord gegooid wordt. Aangezien we met een specifiek gekozen dataset werken, mag aangenomen worden dat willekeurig genomen steekproefsets van landen hogere correlaties zouden hebben met de oorspronkelijke resultaten.

De rangschikkingen die we opgesteld hebben voor de gewijzigde situatie bevatten eveneens de landen die bij het aggregeren buiten beschouwing gelaten zijn. We hebben deze rangschikking opgesteld aan de hand van de scores die men kan berekenen door voor elk land de, gewijzigde, geaggregeerde gewichten te vermenigvuldigen met de scores voor de Safety Performance Indicators. Door de 5 weggelaten landen toch in de rangschikking op te nemen, wordt vergelijken met de oorspronkelijke situatie mogelijk.

Tabel 57 en Tabel 58 tonen een vertrouwelijk beeld met betrekking tot de rangschikkingen. Blijkbaar heeft de verandering in geaggregeerde gewichten nauwelijks impact op de resulterende landenklasseringen. Wanneer we enkel de rangschikkingen beschouwen, blijken zowel MinScore als MinRank onder alle omstandigheden robuust te zijn.

5.2.3 Conclusie

Gebruik maken van een selectie van landen brengt relatief weinig verschil in resultaten met zich mee. Deze maal blijkt MinRank minder robuust te zijn dan MinScore. We kunnen echter nog steeds spreken van relatief robuustemethodes. Door juist de uitzonderingen onder de landen, die de resultaten dus het meest vertekenden, uit de selectie te weren,

is het logisch dat de aangepaste gewichten wat verschillen ten opzichte van de originele situatie. Bovendien blijken er altijd nog enige positieve correlaties te blijven bestaan.

5.3 Robuustheid bij vermindering van het detailniveau in de dataset

Zoals we al aangehaald hebben, worden de precieze gewichten die toegewezen worden aan de indicatoren, bepaald door de dataset die we ter beschikking hebben. In deze dataset kunnen echter waarden staan waarvan de kwaliteit twijfelachtig is. Bijvoorbeeld omdat ze slecht meetbaar zijn, omdat de meetmethode niet helemaal dezelfde was als bij andere cases of omdat we twijfels hebben bij de accuraatheid van het meetinstrument. Methodes die niet te zeer afhankelijk zijn van te exacte waarden in de dataset kunnen dus van grote waarde zijn. Robuuste aggregatiemethodes ondervinden minder nadeel van eventuele fouten of onjuistheden in de data.

We kunnen deze vorm van robuustheid testen door het detailniveau van de metingen te verlagen. Als we het detailniveau verlagen tot, bijvoorbeeld, 5, wil dat zeggen dat er in plaats van 100 mogelijke waarden die een SPI kan aannemen, nog maar 5 waarden overblijven. Alle waarden die dicht tegen elkaar aanleunen worden dus tot dezelfde waarde herleid. De scores van landen waarvan de waarden voor een bepaalde SPI zeer dicht bij elkaar liggen, worden op hetzelfde niveau gebracht. Door het detailniveau te verlagen tot 5 gebruiken we dus in feite gewoon een ruwere meetmethode, die enkel nog aangeeft of een land zeer goed (=1), goed (=0,75), matig (=0,5), slecht (=0,25), of zeer slecht (=0) presteert in vergelijking met de rest van de dataset.

Wanneer er weinig verschil is tussen de geaggregeerde gewichten, die berekend worden op de oorspronkelijke dataset, en de geaggregeerde gewichten die men berekent aan de hand van de minder gedetailleerde set, mag men aannemen dat de aggregatiemethode robuust is. Bij robuuste aggregatiemethoden zijn de resulterende gewichten en landenrangschikking niet te sterk afhankelijk van de precieze waarden in de dataset. Dit brengt ons dan ook bij de volgende onderzoeksvraag:

Veranderen de gekozen gewichten en de daaruit volgende landenrangschikking significant wanneer men het detailniveau van de Safety Performance Indicators aanpast?

5.3.1 Werkwijze

De eerste vraag die we ons moeten stellen, is hoe we de waarden in de vijf categorieën zullen verdelen. We hebben gekozen voor een simpele methode om de landen te verdelen. De verdeling voor elke SPI is gebeurd op basis van het bereik en het aantal beoogde categorieën.

Voor elke SPI hebben we het bereik van de waarden berekend. Vervolgens hebben we er eenvoudigweg voor gekozen om dit bereik onder te verdelen in 5 delen. Het eerste deel bevat al de waarden tussen de maximumwaarde en een vijfde van het bereik. Het tweede deel beslaat weer een vijfde van het bereik, enz. Tenslotte wordt elk land voor elke SPI aan één van de vijf categorieën toegewezen op basis van zijn score op de betreffende indicator.

Een voorbeeld zal de indeling verduidelijken. In Tabel 1 kunnen we aflezen dat de waarden voor de SPI Alcohol variëren tussen 100 en 78,2. Het bereik voor deze SPI is dus 21,8. Vervolgens delen we het bereik door het aantal categorieën. We hebben gekozen om vijf onderverdelingen te maken, wat ons op een stapgrootte van $\frac{21,8}{5} = 4,36$ brengt. We zullen het volledige bereik dus onderverdelen in blokjes van 4,36. Ten slotte brengen we elk land in de bijhorende categorie onder. Uit Tabel 1 lezen we dat België voor Alcohol een score heeft van 94,2. Deze score hoort thuis in categorie 2.¹⁵ We wijzen op een identieke manier elk land toe aan een bepaalde categorie voor elke SPI. Eens alle detailniveaus naar vijf verlaagd zijn, bekomen we, na normalisatie, Tabel 69.

Alcohol is een goed voorbeeld om onze manier van categoriseren te illustreren. Het blijkt immers duidelijk dat niet elke categorie evenveel landen bevat. Het is zelfs mogelijk dat er categorieën zijn waar geen enkel land toe behoort. Enkel de hoogste en de laagste categorie bevatten met zekerheid minstens 1 element. Zo zijn er voor de SPI alcohol maar liefst 14 landen in categorie 1 ondergebracht. In de categorie daaronder zitten 5 landen. Vervolgens zijn er 2 lege categorieën om tenslotte enkel Cyprus in de laatste categorie terug te vinden.

¹⁵ Voor Alcohol omvat de eerste categorie de waarden [95,64;100]. Categorie 2 omvat [91,28;95,64[etc. Telkens in stappen van 4,36.

Als laatste stap converteren we de categorieën naar bijhorende waarden tussen 0 en 1. Dit kan simpelweg gebeuren door de waarden te normaliseren. In de praktijk komt het erop neer dat categorie 1 een score zal krijgen van 1, categorie 2 krijgt een score van 0,75, categorie 3 wordt gelijk gesteld aan 0,5, categorie 4 wordt 0,25 en de 5^{de} categorie krijgt een score van 0.

Eenmaal alle landen een aangepaste waarde per SPI gekregen hebben, berekenen we opnieuw de road safety scores aan de hand van de nieuwe waarden en de gewichten die elke SPI toegewezen kreeg van de wegingmethodes. Het zijn deze aangepaste scores die gebruikt worden als input voor de verschillende aggregatiemethodes.

Het is belangrijk om op te merken dat we ook het detailniveau van de scores op de consensusranglijst op dezelfde wijze verlagen.

5.3.2 Robuustheid

De correlatiecoëfficiënten voor de aangepaste situatie kunnen teruggevonden worden in Tabel 51 en Tabel 52. Het aggregeren van wegingmethodes met consensus geeft zeer robuuste resultaten (Tabel 51). De geaggregeerde gewichten voor de wegingmethodes zonder consensus (Tabel 52) blijken niet altijd robuust te zijn. De aggregatiemethode MinScore produceert aan de hand van de aangepaste dataset gewichten die slechts licht correleren met de gewichten uit de oorspronkelijke situatie. MinRank lijkt echter erg robuust te zijn.

In de rangschikking (Tabel 59 en Tabel 60) is er helemaal niets terug te vinden van de lage correlaties tussen de rangschikkingen. Net als in de vorige situaties blijkt ook hier dat de aggregatiemethodes wat de rangschikkingen betreft, zeer robuust zijn. Zelfs met relatief afwijkende gewichten kan er nog altijd een rangschikking opgesteld worden die erg lijkt op deze uit de originele situatie.

Het aanpassen van de dataset, in dit geval door het detailniveau van de data te verminderen, blijkt vooral een effect te hebben op de aggregatiemethode MinScore. MinRank lijkt dan weer vrijwel ongevoelig te zijn voor veranderingen in de data. Dit kan andermaal verklaard worden doordat MinScore vooral met de scores werkt. Deze worden sterk beïnvloed door een wijziging in de data. De rangschikking die aan de hand van de

scores opgesteld wordt, wordt immers niet of nauwelijks beïnvloed door het detailniveau van de data te verminderen.

5.3.3 Conclusie

Het verlagen van het detailniveau toont aan dat MinRank op dit vlak robuuster is dan MinScore. Dit kan andermaal verklaard worden doordat het verlagen van het detailniveau de scores veel zwaarder beïnvloedt dan de rangschikking. Daardoor zijn er voor MinRank nauwelijks verschillen te merken in de resultaten. De aggregatiemethode MinScore is gevoeliger voor het veranderen van het detailniveau aangezien dit een directe impact heeft op de scores. Wanneer men niet zeker is van de correctheid van de waarden in de dataset, kan best de aggregatiemethode MinRank gebruikt worden.

6 **Conclusies**

In dit onderzoek stond het beschrijven en het testen op robuustheid van mogelijke aggregatiemethoden centraal. Verschillende methodes werden toegepast op een Europese verkeersveiligheidsdataset waarmee we het belang van zeven onderliggende risicofactoren van verkeersveiligheid trachtten te achterhalen. Aan de hand van het gewicht dat aan elk van de Safety Performance Indicators werd toegekend, gecombineerd met de prestaties van een set van 20 Europese landen voor deze indicatoren, kan er vervolgens een road safety score per land berekend worden. Deze score geeft een indicatie van de verkeersveiligheid in het betreffende land. Tenslotte kan er op basis van de road safety scores een landenrangschikking opgesteld worden.

Er bestaan vele verschillende wegingmethodes die men kan gebruiken om gewichten aan de indicatoren toe te kennen. Aangezien elke wegingmethode sterke en zwakke punten heeft en er geen overeenstemming heerst over welke methode te verkiezen valt boven alle andere methodes, kan aggregatie een oplossing bieden. In hoofdstuk 3 hebben we getracht een oplossing te formuleren op de vraag: *“Op welke manieren kunnen de verschillende gewichten geaggregeerd worden in één gewicht?”*

We hebben een aantal mogelijke aggregatiemethodes voorgesteld die we onderverdeeld hebben in 4 categorieën. De eerste categorie is de categorie ‘Equal weighting’ waarbij er geen onderscheid gemaakt wordt tussen de wegingmethodes en dus elke wegingmethode even zwaar doorweegt in het geaggregeerde gewicht. De varianten die we onder deze categorie rekenden zijn MathMean en GeoMean, die op basis van respectievelijk het wiskundig en het meetkundig gemiddelde werken. Ten tweede hebben we de categorie van de Deviation Methods waaronder we de Absolute Deviation Method en de Standard Deviation Method rekenen. Deze aggregatiemethodes maken gebruik van afwijkingen in de gegevens om gewichten te bepalen. De derde categorie bevat de Minimization problems. Hier wordt er een bepaalde voorspellingsfout gedefinieerd die vervolgens geminimaliseerd wordt. De aggregatiemethode MinScore minimaliseert het verschil tussen de road safety scores en de score voor de consensusranglijst. De methode MinRank minimaliseert het verschil in posities tussen beide rangschikkingen. De laatste categorie bevat de correlation-based aggregation. De aggregatiemethodes in deze categorie werken op basis van de correlatie tussen de rangschikking op basis van de road safety scores en de consensusranglijst. We kunnen drie van de voorgestelde

aggregatiemethodes in deze laatste categorie onderbrengen. Correlatie tussen rankings, Correlatie tussen relatieve posities en Classical Preference Structures maken alle drie gebruik van een vorm van overeenkomst tussen consensusranglijst en rangschikking op basis van road safety scores.

Aangezien de minimalisatieproblemen MinScore en MinRank het meest intensief gebruik maken van de beschikbare gegevens, zijn ze ook het gevoeligst voor wijzigingen in de dataset. We testen dan ook de robuustheid van beide aggregatiemethodes en maken een vergelijking. De robuustheid van de methodes hebben we uitgetest door lichte wijzigingen aan te brengen in de dataset om vervolgens vast te stellen in welke mate de resulterende gewichten en landenrangschikking veranderden ten opzichte van de oorspronkelijke situatie. In eerste instantie hebben we het effect van een wijziging in consensusranglijst onderzocht. De eerste nieuwe consensus die we uitprobeerden, was de ranglijst op basis van het aantal verkeersdoden per miljoen afgelegde kilometers. De tweede alternatieve consensus werd opgesteld op basis van het aantal verkeersdoden per 100.000 ingeschreven personenwagens. Als tweede test hebben we de aggregatiemethodes toegepast op een selectie van landen. De vijf landen waarvan de methode MinRank het minst goed de positie in de landenrangschikking kon voorspellen, werden uit de selectie geweerd. Deze landen waren: Estland, Frankrijk, Nederland, Portugal en Slovenië. Ten slotte hebben we het effect onderzocht dat het verlagen van het detailniveau van de data heeft op de resulterende gewichten en landenrangschikking. Hoewel deze tests niet altijd eenduidige resultaten opleveren, kunnen er toch enkele waardevolle conclusies getrokken worden.

Als antwoord op de deelvraag - *"Wat is de impact op de gekozen gewichten en de daaruit volgende landenrangschikking wanneer de overeenkomst met een andere consensus als criterium gebruikt wordt?"* - kunnen we concluderen dat de aggregatiemethode MinScore erg afhankelijk is van de gekozen consensusranglijst. MinScore slaagt er niet altijd in om robuuste resultaten te creëren bij veranderingen in de consensusranglijst. De aggregatiemethode MinRank bleek op dit vlak wel erg robuust te zijn.

Het werken met een gerichte selectie van een aantal landen uit de dataset heeft een bepaalde invloed op de dataset. De aggregatiemethode MinScore blijkt op dit vlak robuuster te zijn dan MinRank. De minimalisatiemethodes mogen vrij robuust genoemd worden wanneer er een willekeurige steekproef van landen als input uit de dataset

weerhouden wordt. De deelvraag: *"Heeft het selecteren van een aantal landen uit de dataset een invloed op de gekozen gewichten en de daaruit volgende landenrangschikking?"* hebben we hiermee beantwoord.

Wanneer het detailniveau van de dataset verlaagd wordt, blijkt vooral MinRank robuust te zijn. De resultaten van MinScore worden door een dergelijke wijziging sterker beïnvloed. Als antwoord op de deelvraag: *"Veranderen de gekozen gewichten en de daaruit volgende landenrangschikking significant wanneer men het detailniveau van de Safety Performance Indicators aanpast?"* kunnen we dan ook vooral MinRank aanduiden als een robuuste methode.

De centrale onderzoeksvraag: *"Hoe stabiel zijn de verschillende aggregatiemethodes voor gewichtsbepaling ten opzichte van de presentatie van de consensusranglijst?"* kan positief beantwoord worden. Er blijkt over het algemeen een zekere stabiliteit van de resultaten, zowel gewichten als landenrangschikking, te zijn. Toch moet er ruimte gehouden worden voor de nodige nuances. Men kan het best een oordeel vormen over de stabiliteit van de aggregatiemethodes door een beschouwing te maken van de verschillende deelvragen. Deze onderzoeken elk een aspect van robuustheid door bepaalde aanpassingen aan te brengen in de dataset, en geven gezamenlijk een beeld van de algemene robuustheid. We kunnen concluderen dat de stabiliteit van de aggregatiemethodes erg afhankelijk is van de aard van de veranderingen die aangebracht werden in de data.

6.1 Verder onderzoek

We zijn na dit onderzoek versterkt in het vermoeden dat het framework voor de verkeersveiligheidsprestatie-index in realiteit iets complexer is dan het opnemen van zeven prestatie-indicatoren. Verder onderzoek hieromtrent kan verduidelijking brengen. Zoals we in sectie 1.2 al aangaven, zijn er allicht één of meerdere ontbrekende elementen in het framework. Dit kan een onbekende extra indicator zijn die het aantal verkeersdoden in een aantal landen helpt verklaren. Het is echter ook mogelijk dat er tussen de indicatoren bepaalde interactie-effecten bestaan, die totnogtoe niet in kaart zijn gebracht. Het zou interessant zijn om te zien welke resultaten we zouden bekomen indien de methodes opnieuw toegepast zouden worden op een uitgebreider framework.

Deze studie heeft inzicht geboden in mogelijke aggregatiemethodes en hun robuustheid. Uiteraard moet er de nodige voorzichtigheid aan de dag gelegd worden bij het interpreteren van de resultaten. Er mag zeker niet vergeten worden dat de tests niet exhaustief gebeurd zijn. Zo wordt het effect van een verandering in consensus slechts 2 maal getest. De veranderingen aan de dataset zijn beide zelfs slechts 1 maal getest, zij het dan dat de veranderingen doordacht gebeurd zijn. Verder onderzoek kan er dan ook werk van maken om de robuustheid van de aggregatiemethodes intensiever te testen. Dit kan gebeuren door de methodes opnieuw toe te passen op meerdere consensusranglijsten, meerdere al dan niet willekeurige selecties van landen te maken en het detailniveau van de data verder verlagen. Daarnaast kunnen ook andere maatstaven gehanteerd worden.

Ook het bestuderen van nieuwe aggregatiemethoden kan een interessant onderwerp vormen van verder onderzoek.

Tenslotte zou het nuttig kunnen zijn om de voorgestelde aggregatiemethoden uit te testen op een totaal andere dataset buiten het toepassingsgebied van de verkeersveiligheid om de waarde van bepaalde methodes in een bredere context te bepalen.

7 Lijst geraadpleegde werken

Banens, P. D. (1994). Regelkaarten en stabiliteit. In P. D. Banens, *Industriële statistiek en kwaliteit* (pp. 135-158). Kluwer Bedrijfswetenschappen.

Chaffey, D., & Wood, S. (2005). Introduction to information management. In *Business Information Management: Improving Performance using information systems* (p. 11).

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* , 429-444.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., et al. (2006). *Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index*. Catholic University of Leuven and Joint Research Center.

Clarke, E., & Barton, B. *Entropy and MDL Discretization of Continuous Variables for Bayesian Belief Networks*. Maryland Medical Research Institute opgevraagd:
<http://sci2s.ugr.es/keel/pdf/specific/articulo/IJIS00.pdf> op 20/05/2009.

Cook, W. D. (2005). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research* .

De Baets, B., & Fodor, J. (1997). Twenty Years of Fuzzy Preference Structures (1978-1997). *Decisions in Economics and Finance* , 20(1) 45-66 .

Goldenbeld, C., & Amelink, M. (2006). Dissemination of the Road Safety Information System. In *Building the European Road Safety Observatory* (p. D3.4). SafetyNet.

Hakkert, A. S., Gitelman, V., & Vis, M. A. (2007). Road Safety Performance Indicators: Theory. In *Building the European Road Safety Observatory* (p. D 3.6). SafetyNet.

Hermans, E., Shen, Y., Vanhoof, K., Wets, G., & Brijs, T. (2008). *Ranking Road Safety Risk Factors using Preference Structures and Fuzzy Preference Structures*. Hasselt University.

Hermans, E., Van den Bossche, F., & Wets, G. (2008a). Combining Road Safety Information in a Performance Index. *Accident Analysis and Prevention* , 40(4) p.1337-1344.

Hermans, E., Van den Bossche, F., & Wets, G. (2008b). *Impact of methodological Choices on Road Safety Ranking*. TRB.

Kaplan, R. S., & Norton, D. P. (2002). The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review Jan-Feb* , 70-80.

Kendall, M. (1938). A New Measure of Rank Correlation. *Biometrika* , 30 81-89 .

Kirvesoja, H. (2001). Subjective vs Objective. In H. Kirvesoja, *Experimental ergonomic evaluation with user trails: EEE product development procedures* (pp. 32-33). University of Oulu, Finland.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Organisation for Economic Co-operation and Development.

Nicewander, W. A., & Rodgers, J. L. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, Feb 42(1) 59-66.

OECD. (2009). *Road Motor Vehicles And Road Fatalities*. Opgevraagd op: <http://www.oecd.org/dataoecd/44/48/36340933.pdf> opgevraagd te 23/05/2009.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15 72-101.

van der Zee, F. (2004). *Kennisverwerving in de Empirische Wetenschappen, de methodologie van wetenschappelijk onderzoek*. Groningen: BMOOO.

Vis, M. A. (2005). State of the art Report on Road Safety Performance Indicators. In *Building the European Road Safety Observatory* (p. D3.1). SafetyNet.

Vis, M. A., & Amelink, M. (2006). Safety Performance Indicators: posters for the first SafetyNet Conference. In *Building the European Road Safety Observatory* (p. D3.5). SafetyNet.

Vis, M. A., & Van Gent, A. L. (2007b). Road Safety Performance Indicators: Country profiles. In *Building the European Road Safety Observatory* (p. D 3.7b). SafetyNet.

Vis, M. A., & Van Gent, A. (2007a). Road Safety Performance Indicators: A country comparison. In *Building the European Road Safety Observatory* (p. D 3.7a). SafetyNet.

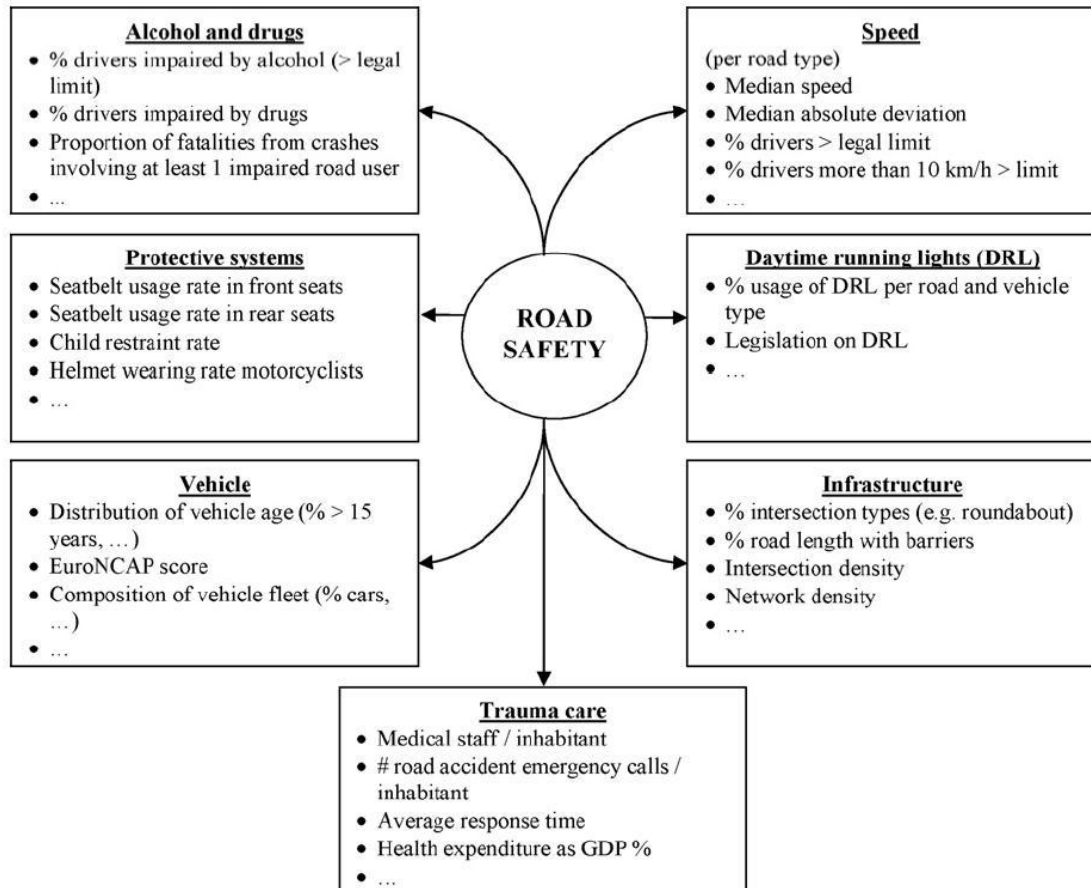
Wang, Y.-M., & Fan, Z.-P. (2007). Fuzzy preference relations: Aggregation and weight determination. *Computers & Industrial Engineering*.

Witten, I. H., & Frank, E. (1999). *DATA MINING: Practical machine learning tools and techniques with Java implementations*. University of Walkato, New Zealand.

8 Bijlagen

8.1 Kadering van het onderzoek

E. Hermans et al. / Accident Analysis and Prevention 40 (2008) 1337–1344



Figuur 2: Road Safety Framework

8.2 Gebruikte informatie

8.2.1 Data

Tabel 5: Ruwe waarden per SPI

Country	Alcohol	Speed	Seat Belt	DRL law	Vehicle	Infra-structure	Trauma Care
België	94,2	86	66	2	40,84	4,91	9,1
Cyprus	78,2	82	80	1	17,08	1,28	6,4
Denemarken	99,7	86	84	3	39,14	1,67	8,8
Duitsland	97,6	83	94	2	37,85	1,8	10,9
Estland	98,4	80	75	3	17,12	1,24	5,1
Finland	99,7	90	89	3	29	0,23	7,3
Frankrijk	94,9	90	97	2	37,18	1,81	9,7
Griekenland	92,1	81	40	1	27,8	0,89	9,5
Hongarije	98,7	83	59	2	26,28	1,73	7,8
Ierland	97,5	96	85	1	54,45	1,36	7,3
Italië	92,7	85	71	2	35,26	1,61	8,5
Nederland	98,1	86	86	1	38,9	3,04	9,1
Oostenrijk	97,4	89	77	3	35,14	1,59	7,7
Polen	99,7	89	71	2	22,29	1,36	6,1
Portugal	95,8	85	88	2	28,7	0,88	9,3
Slovenië	97,3	90	81	3	50,07	1,89	8,2
Spanje	92,8	87	86	2	36,11	0,71	7,6
Tsjechië	98	93	61	2	19,43	1,62	7,6
VK	99,4	92	93	1	44,42	1,53	7,7
Zweden	100	86	92	3	39,43	0,48	9,2

Tabel 6: Genormaliseerde waarden per SPI

Country	Alcohol	Speed	Seat Belt	DRL law	Vehicle	Infra-structure	Trauma Care
België	0,73	0,38	0,46	0,50	0,64	1,00	0,69
Cyprus	0,00	0,13	0,70	0,00	0,00	0,22	0,22
Denemarken	0,99	0,38	0,77	1,00	0,59	0,31	0,64
Duitsland	0,89	0,19	0,95	0,50	0,56	0,34	1,00
Estland	0,93	0,00	0,61	1,00	0,00	0,22	0,00
Finland	0,99	0,63	0,86	1,00	0,32	0,00	0,38
Frankrijk	0,77	0,63	1,00	0,50	0,54	0,34	0,79
Griekenland	0,64	0,06	0,00	0,00	0,29	0,14	0,76
Hongarije	0,94	0,19	0,33	0,50	0,25	0,32	0,47
Ierland	0,89	1,00	0,79	0,00	1,00	0,24	0,38
Italië	0,67	0,31	0,54	0,50	0,49	0,29	0,59
Nederland	0,91	0,38	0,81	0,00	0,58	0,60	0,69
Oostenrijk	0,88	0,56	0,65	1,00	0,48	0,29	0,45
Polen	0,99	0,56	0,54	0,50	0,14	0,24	0,17
Portugal	0,81	0,31	0,84	0,50	0,31	0,14	0,72
Slovenië	0,88	0,63	0,72	1,00	0,88	0,35	0,53
Spanje	0,67	0,44	0,81	0,50	0,51	0,10	0,43
Tsjechië	0,91	0,81	0,37	0,50	0,06	0,30	0,43
VK	0,97	0,75	0,93	0,00	0,73	0,28	0,45
Zweden	1,00	0,38	0,91	1,00	0,60	0,05	0,71

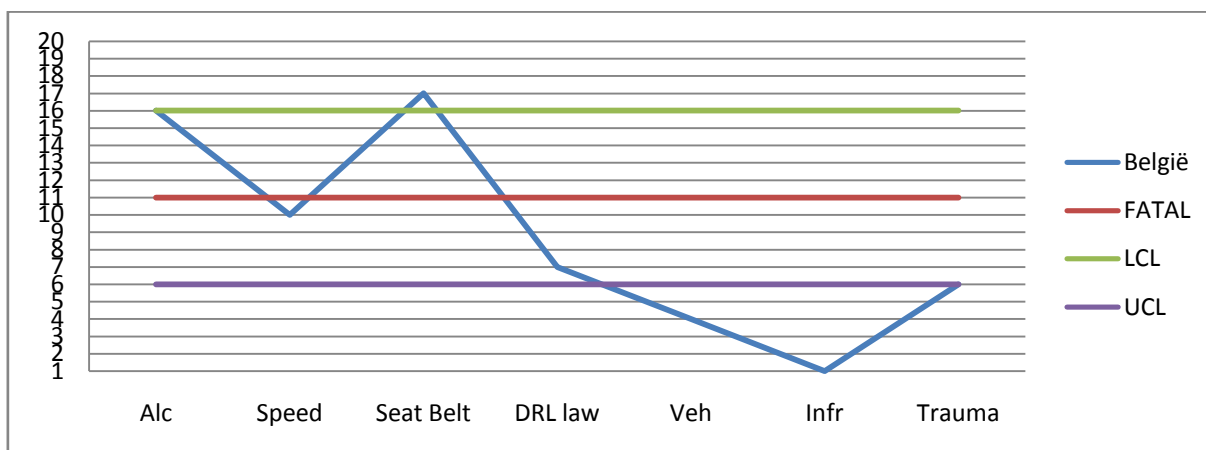
Tabel 7: Ruwe waarden voor 3 mogelijke consensusranglijsten

Country	Fatal/mln inhabitants	Fatal/mln km	Fatal/100. 000 Cars
België	117	110	252
Cyprus	134	293	321
Denemarken	80	70	228
Duitsland	80	76	148
Estland	121	162	378
Finland	73	63	167
Frankrijk	101	81	206
Griekenland	146	235	418
Hongarije	131	280	477
Ierland	84	81	224
Italië	105	78	177
Nederland	63	69	149
Oostenrijk	115	111	230
Polen	149	323	506
Portugal	148	156	257
Slovenië	121	154	272
Spanje	130	153	289
Tsjechië	142	202	390
VK	61	54	136
Zweden	59	54	130

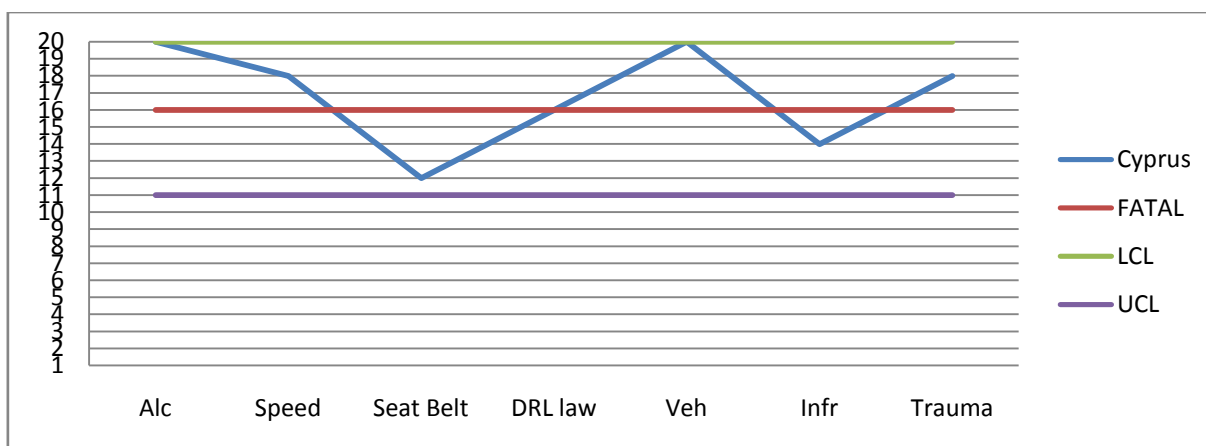
Tabel 8: Genormaliseerde waarden voor 3 mogelijke consensusranglijsten

Country	Fatal/mln inhabitants	Fatal/mln km	Fatal/100. 000 Cars
België	0,36	0,79	0,68
Cyprus	0,17	0,11	0,49
Denemarken	0,77	0,94	0,74
Duitsland	0,77	0,92	0,95
Estland	0,31	0,60	0,34
Finland	0,84	0,97	0,90
Frankrijk	0,53	0,90	0,80
Griekenland	0,03	0,33	0,23
Hongarije	0,20	0,16	0,08
Ierland	0,72	0,90	0,75
Italië	0,49	0,91	0,88
Nederland	0,96	0,94	0,95
Oostenrijk	0,38	0,79	0,73
Polen	0,00	0,00	0,00
Portugal	0,01	0,62	0,66
Slovenië	0,31	0,63	0,62
Spanje	0,21	0,63	0,58
Tsjechië	0,08	0,45	0,31
VK	0,98	1,00	0,98
Zweden	1,00	1,00	1,00

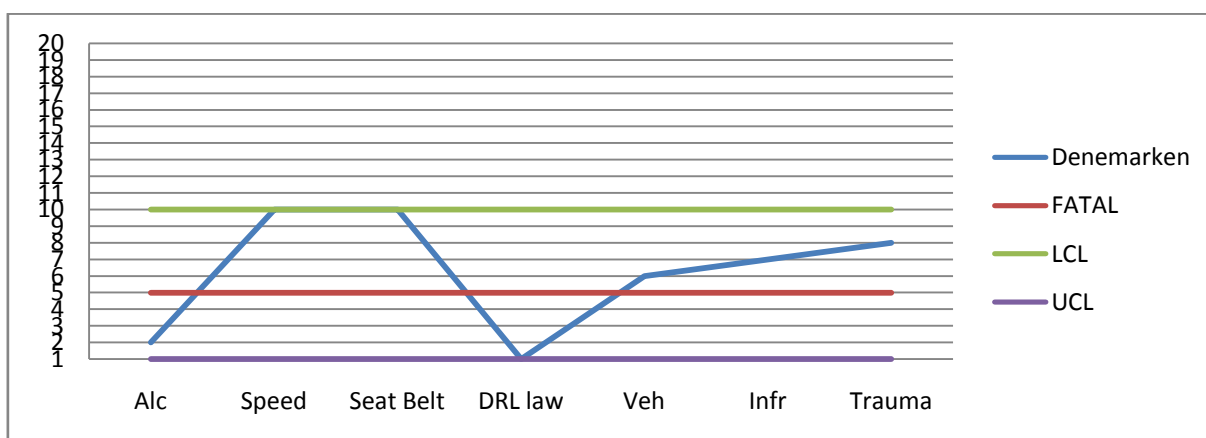
8.2.2 Analyse van de data



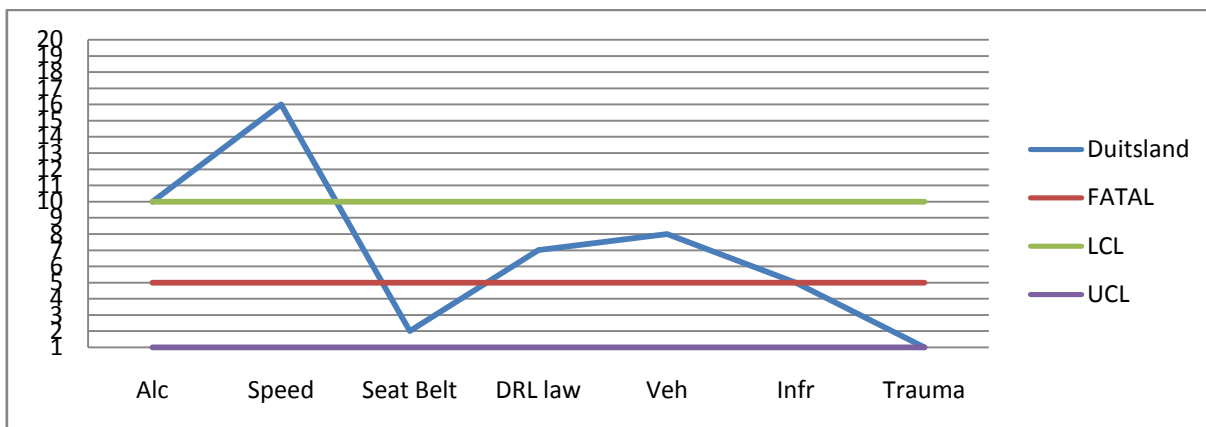
Figuur 3: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor België



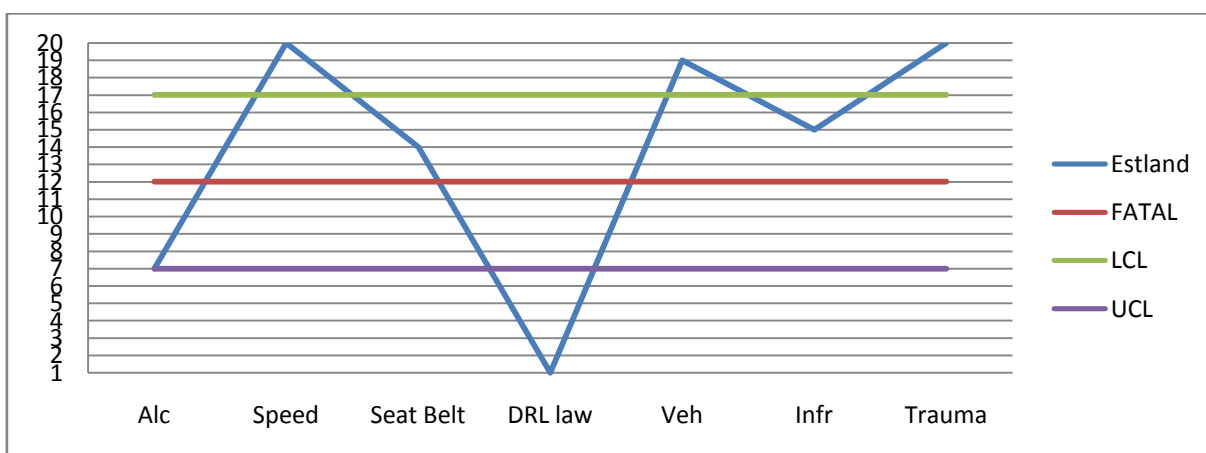
Figuur 4: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Cyprus



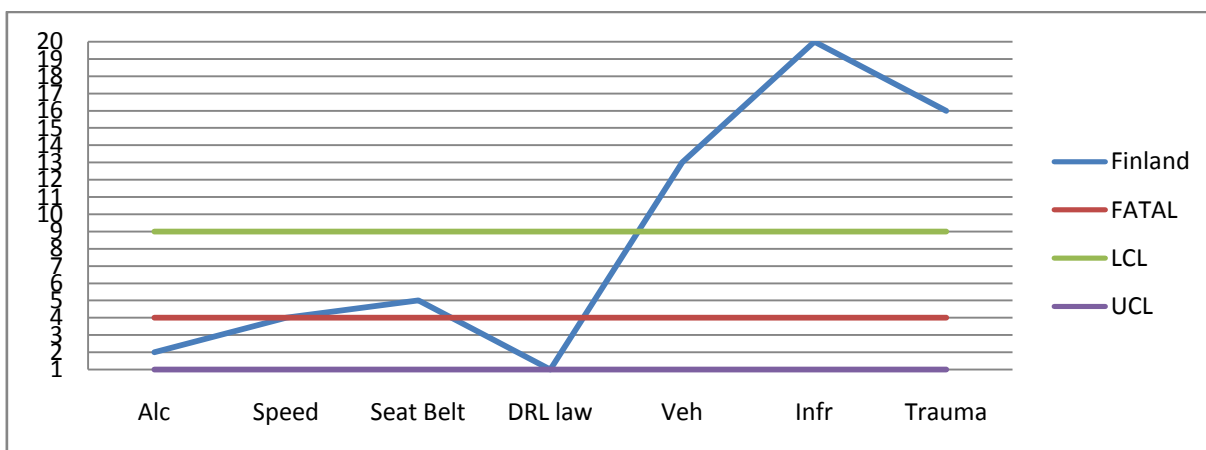
Figuur 5: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Denemarken



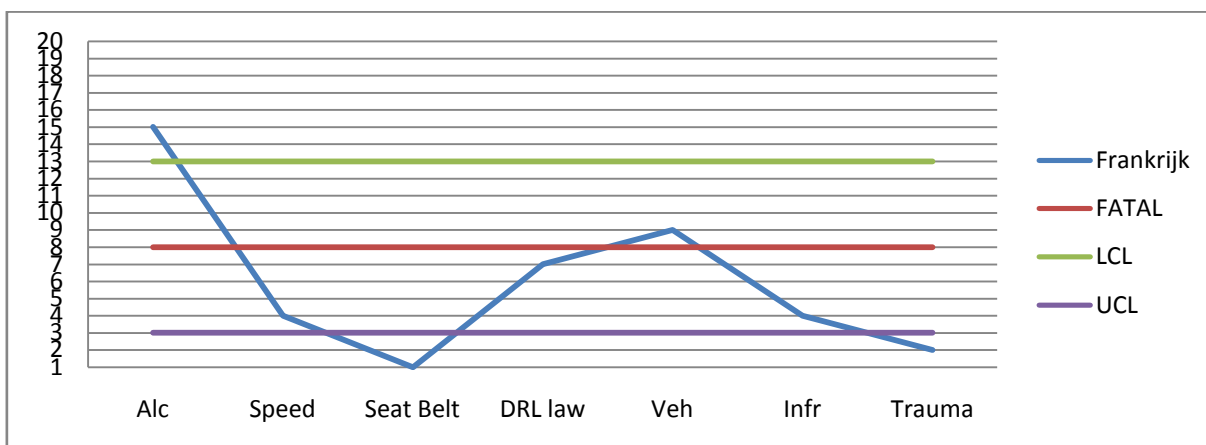
Figuur 6: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Duitsland



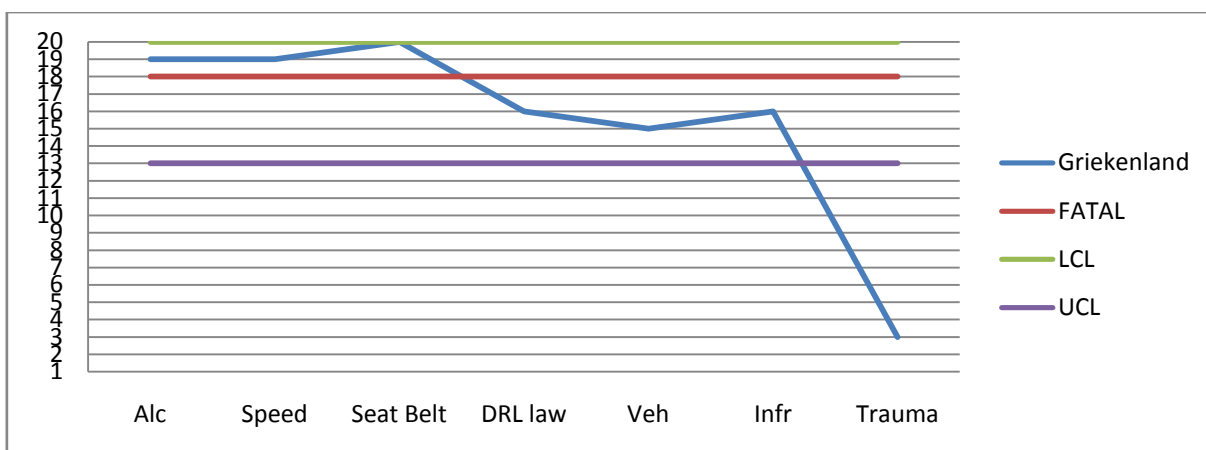
Figuur 7: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Estland



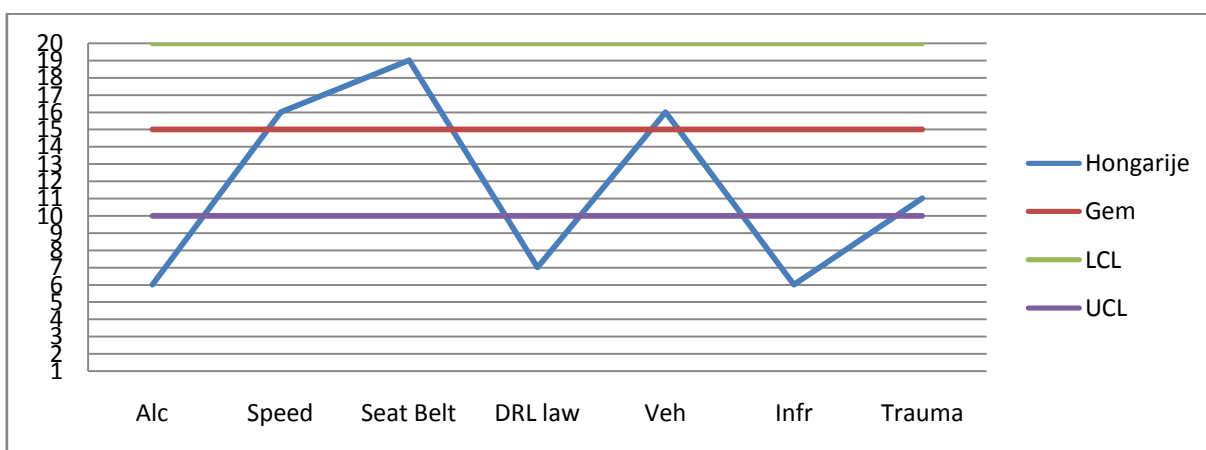
Figuur 8: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Finland



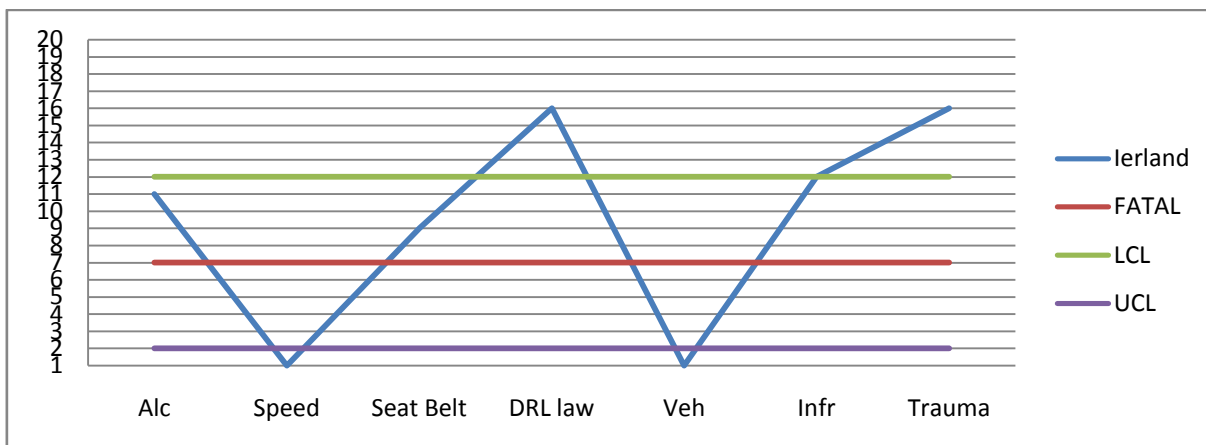
Figuur 9: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Frankrijk



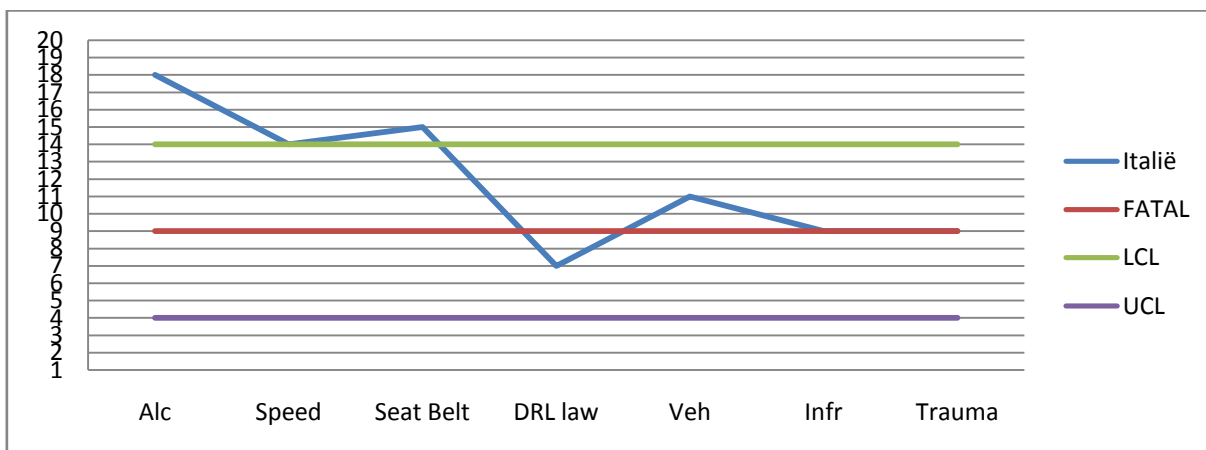
Figuur 10: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Griekenland



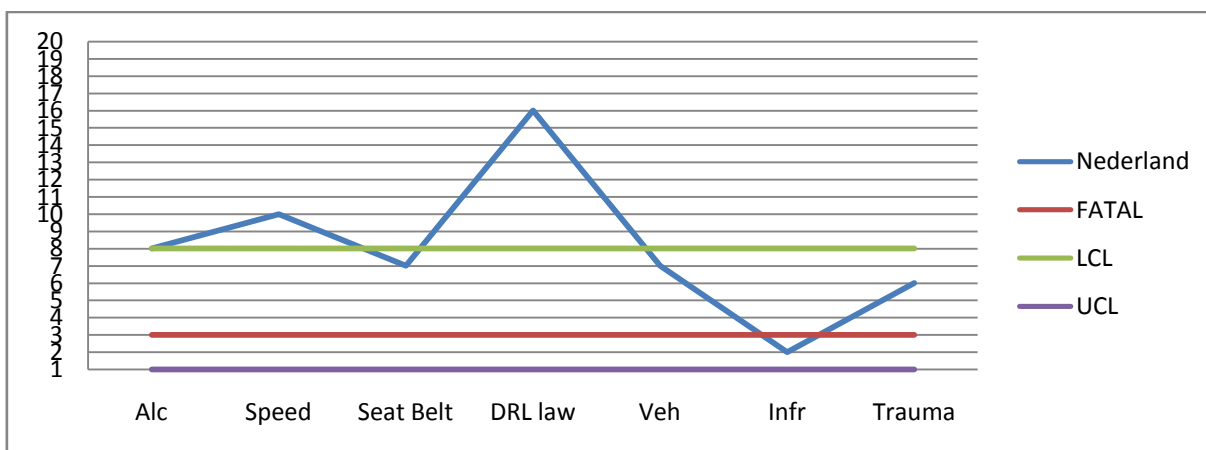
Figuur 11: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Hongarije



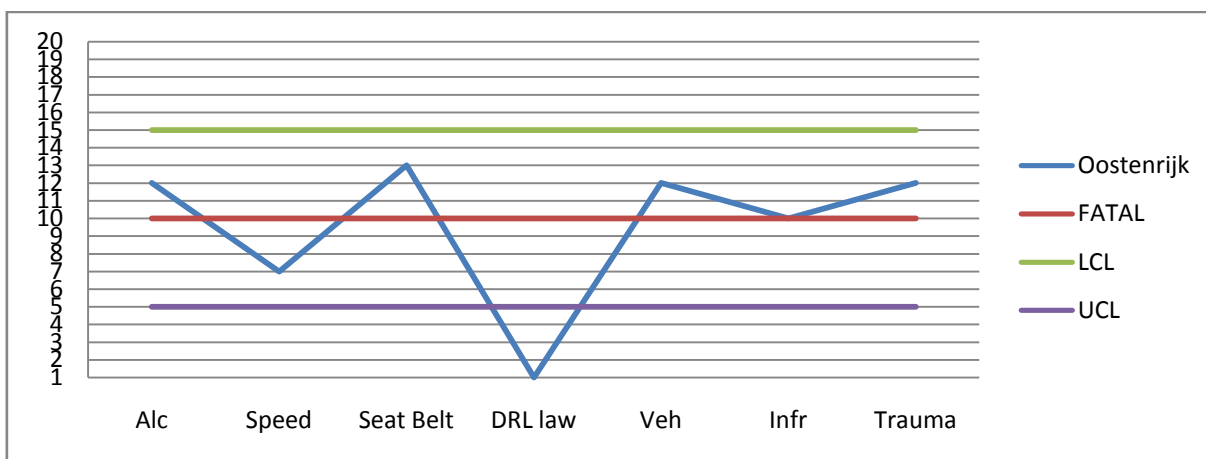
Figuur 12: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Ierland



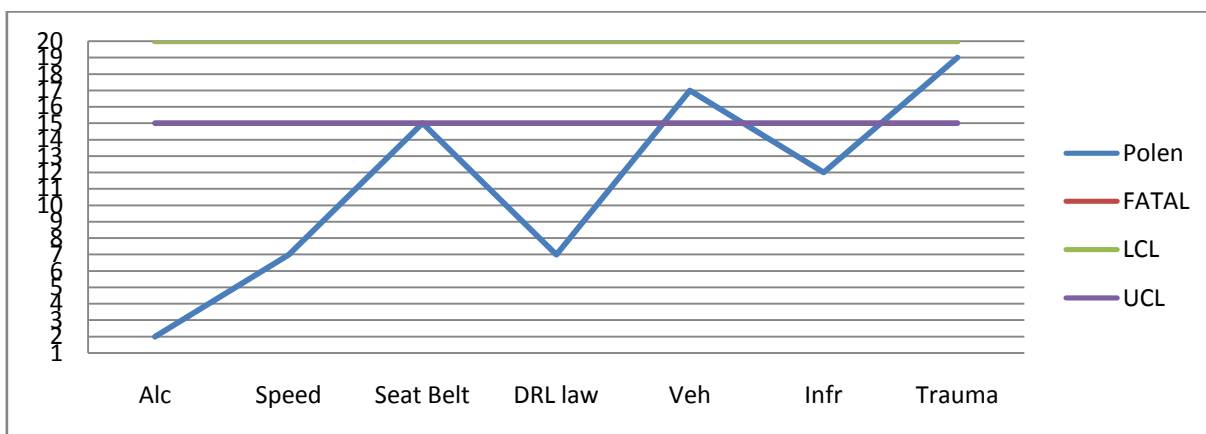
Figuur 13: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Italië



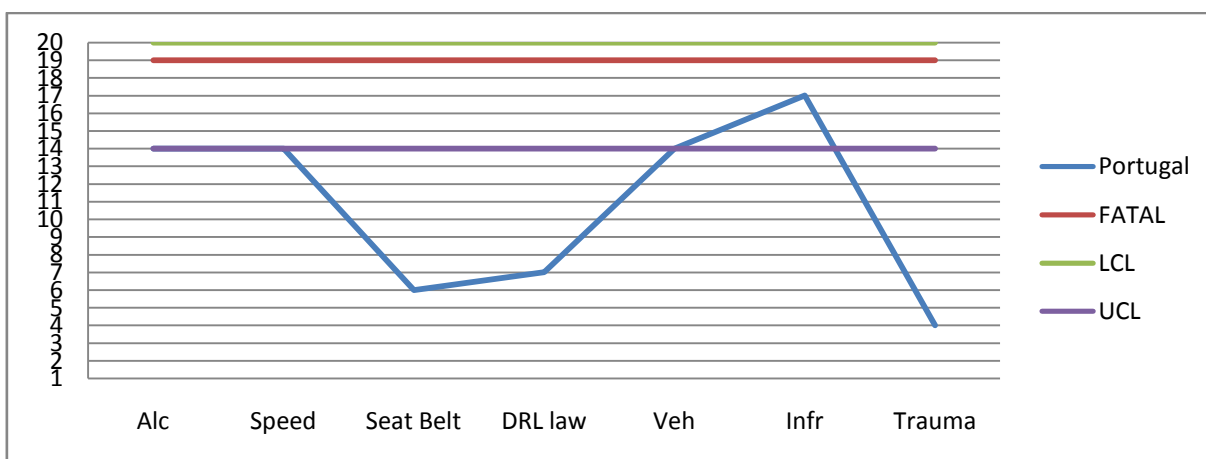
Figuur 14: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Nederland



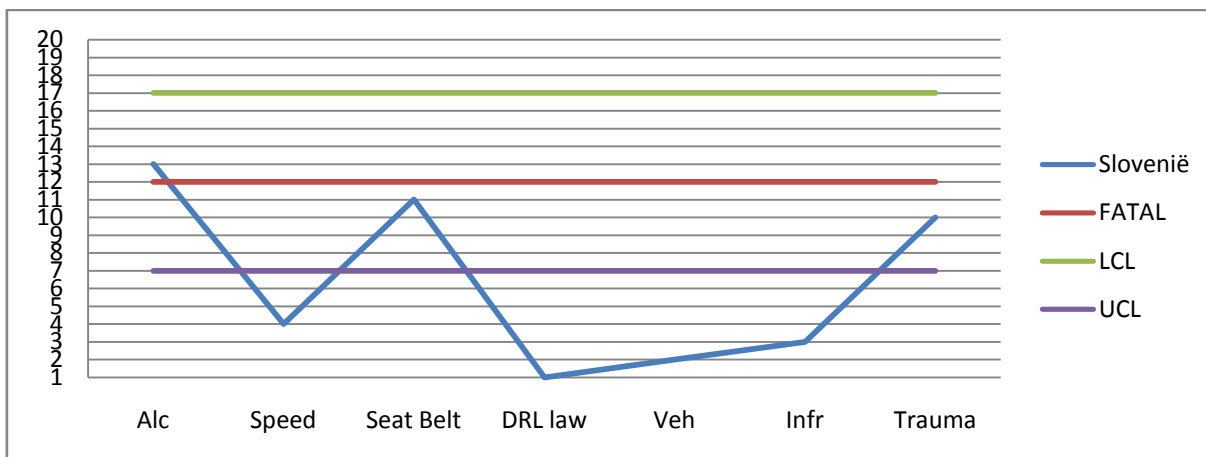
Figuur 15: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Oostenrijk



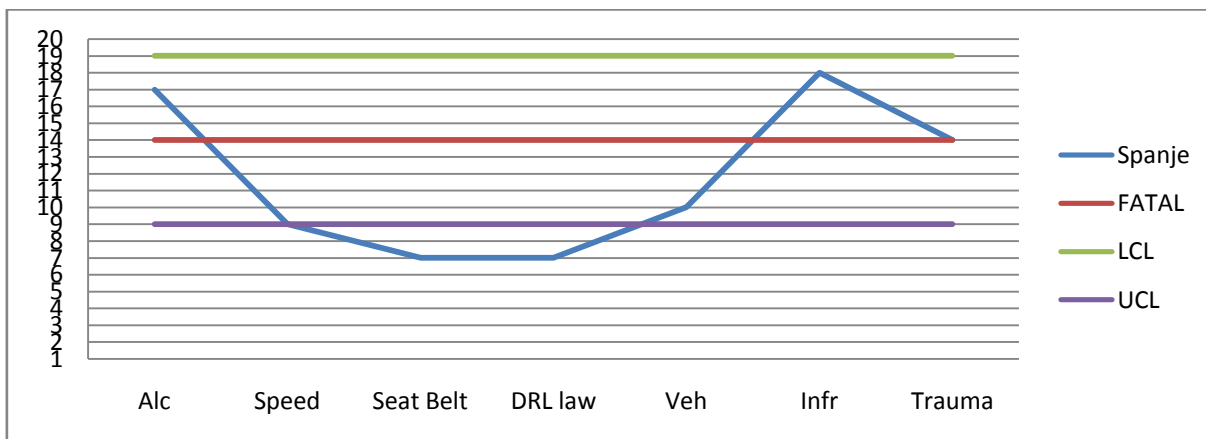
Figuur 16: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Polen



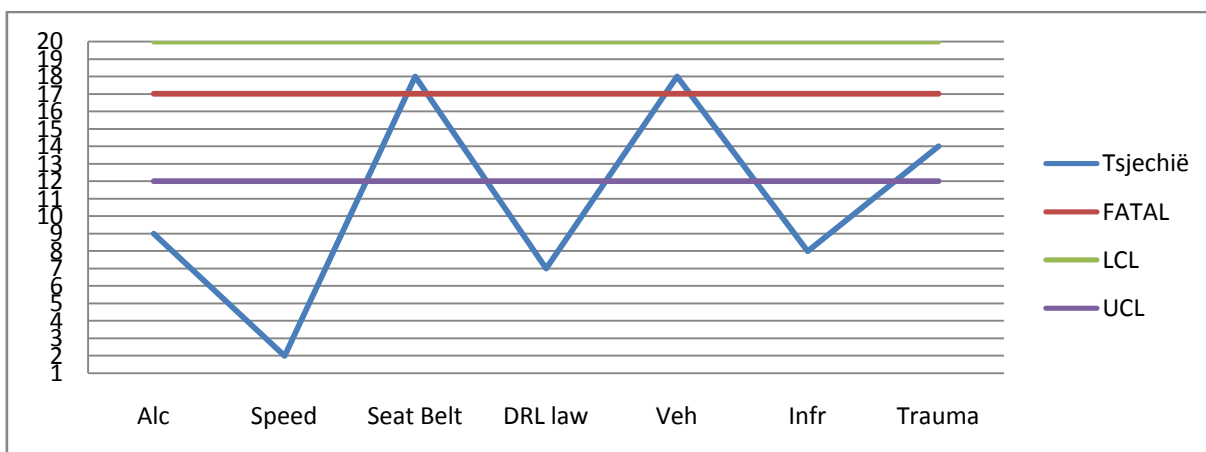
Figuur 17: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Portugal



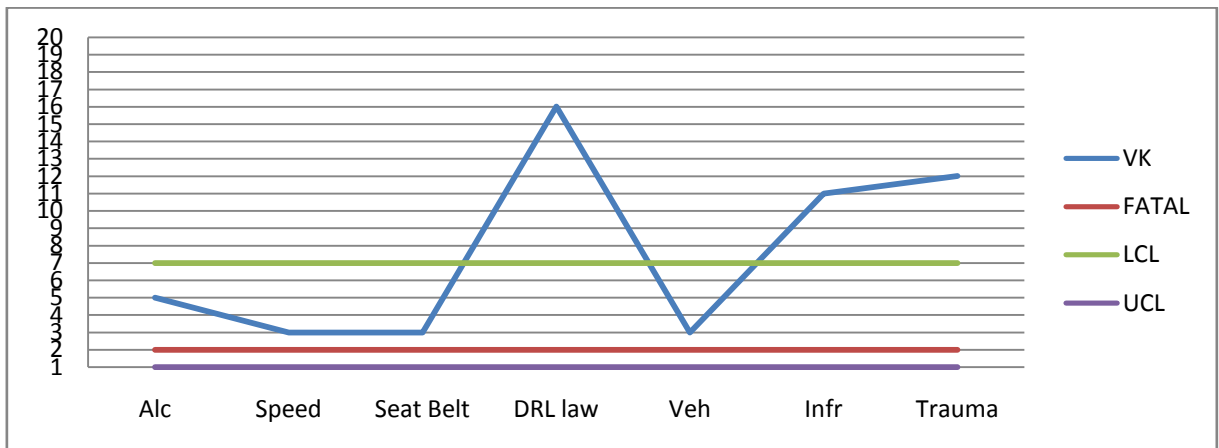
Figuur 18: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Slovenië



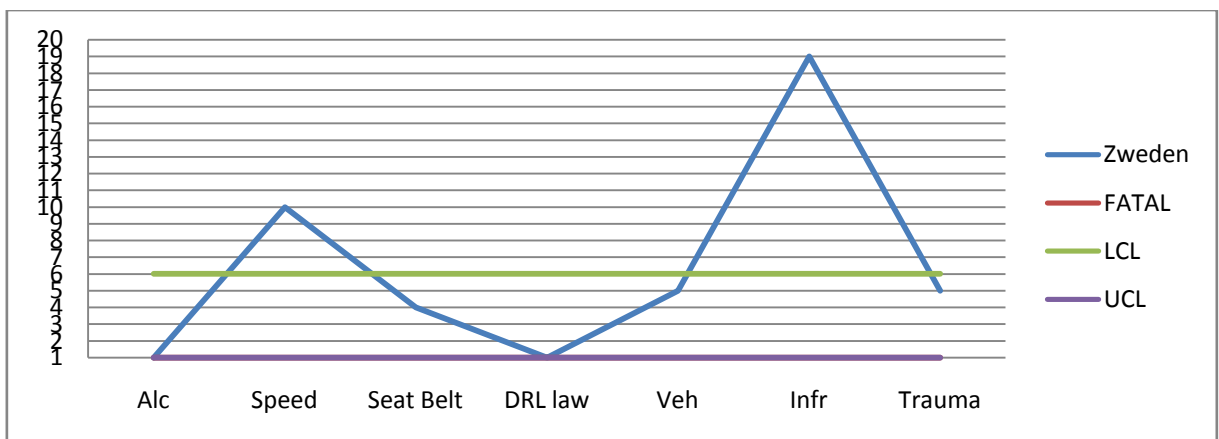
Figuur 19: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Spanje



Figuur 20: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Tsjechië



Figuur 21: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor het Verenigd Koninkrijk



Figuur 22: Samenhang tussen SPI en verkeersdoden per miljoen inwoners voor Zweden

8.3 Aggregatiemethodes: Landenrangschikkingen

Tabel 9: Landenrangschikking volgens MathMean (wegingmethodes meet consensus)

MATH. MEAN	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,69	2	8	+6
Zweden	0,66	3	1	-2
Ierland	0,66	4	7	+3
Duitsland	0,66	5	5	0
Denemarken	0,65	6	5	-1
VK	0,64	7	2	+5
België	0,61	8	11	+3
Nederland	0,60	9	3	-6
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,54	12	19	-7
Spanje	0,52	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,46	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,40	17	15	-2
Estland	0,35	18	12	+6
Griekenland	0,26	19	18	-1
Cyprus	0,23	20	16	-4

Tabel 11: Landenrangschikking volgens MathMean (wegingmethodes zonder consensus)

MATH. MEAN	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,66	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Oostenrijk	0,62	7	10	+3
Finland	0,62	8	4	-4
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,52	13	19	+6
Spanje	0,50	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,43	17	15	-2
Estland	0,39	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

Tabel 10: Landenrangschikking volgens GeoMean (wegingmethodes met consensus)

GEO MEAN	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Frankrijk	0,66	2	8	+6
Zweden	0,66	3	1	-2
Denemarken	0,66	4	5	+1
Duitsland	0,64	5	5	0
Ierland	0,63	6	7	+1
België	0,62	7	11	+4
VK	0,61	8	2	-6
Oostenrijk	0,61	9	10	+1
Finland	0,59	10	4	-6
Nederland	0,58	11	3	-8
Portugal	0,52	12	19	+7
Spanje	0,50	13	14	+1
Italië	0,48	14	9	-5
Tsjechië	0,48	15	17	+2
Polen	0,44	16	20	+4
Hongarije	0,42	17	15	-2
Estland	0,37	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,20	20	16	-4

Tabel 12: Landenrangschikking volgens GeoMean (wegingmethodes zonder consensus)

GEO MEAN	Score	Rank	Fatal	Diff
Ierland	0,74	1	7	+6
VK	0,71	2	2	0
Frankrijk	0,68	3	8	+5
Slovenië	0,67	4	12	+8
Nederland	0,66	5	3	-2
Duitsland	0,63	6	5	-1
België	0,63	7	11	+4
Denemarken	0,61	8	5	-3
Zweden	0,61	9	1	-8
Oostenrijk	0,57	10	10	0
Finland	0,56	11	4	-7
Tsjechië	0,53	12	17	+5
Portugal	0,53	13	19	+6
Spanje	0,50	14	14	0
Polen	0,48	15	20	-5
Italië	0,48	16	9	-7
Hongarije	0,42	17	15	-2
Estland	0,32	18	12	-6
Griekenland	0,30	19	18	-1
Cyprus	0,21	20	16	-4

**Tabel 13: Landenrangschikking volgens ADM
(wegingmethodes met consensus)**

ADM	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,69	2	8	+6
Ierland	0,67	3	7	+4
Zweden	0,67	4	1	-3
Duitsland	0,66	5	5	0
Denemarken	0,65	6	5	-1
VK	0,64	7	2	-5
België	0,61	8	11	+3
Nederland	0,61	9	3	-6
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,54	12	19	+7
Spanje	0,52	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,45	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,39	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,26	19	18	-1
Cyprus	0,24	20	16	-4

**Tabel 14: Landenrangschikking volgens SDM
(wegingmethodes met consensus)**

SDM	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,69	2	8	+6
Zweden	0,67	3	1	-2
Ierland	0,67	4	7	+3
Duitsland	0,66	5	5	0
Denemarken	0,65	6	5	-1
VK	0,64	7	2	-5
België	0,61	8	11	+3
Nederland	0,60	9	3	-6
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,54	12	19	+7
Spanje	0,52	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,45	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,40	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,26	19	18	-1
Cyprus	0,24	20	16	-4

**Tabel 15: Landenrangschikking volgens ADM
(wegingmethodes zonder consensus)**

ADM	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,66	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Oostenrijk	0,62	7	10	+3
Finland	0,62	8	4	-4
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,52	13	19	+6
Spanje	0,50	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,43	17	15	-2
Estland	0,39	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

**Tabel 16: Landenrangschikking volgens SDM
(wegingmethodes zonder consensus)**

SDM	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,66	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Oostenrijk	0,62	7	10	-3
Finland	0,62	8	4	+4
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,52	13	19	+6
Spanje	0,50	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,43	17	15	-2
Estland	0,39	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

Tabel 17: Landenrangschikking volgens MinScore (wegingmethodes met consensus)

MinScore	Score	Rank	Fatal	Diff
Slovenië	0,72	1	12	+11
Frankrijk	0,70	2	8	+6
Ierland	0,70	3	7	+4
Zweden	0,69	4	1	-3
Duitsland	0,68	5	5	0
VK	0,67	6	2	-4
Denemarken	0,66	7	5	-2
Nederland	0,61	8	3	-5
Finland	0,60	9	4	-5
Oostenrijk	0,60	10	10	0
België	0,60	11	11	0
Portugal	0,55	12	19	+7
Spanje	0,54	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,44	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,39	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,25	19	18	-1
Cyprus	0,24	20	16	-4

Tabel 18: Landenrangschikking volgens MinRank (wegingmethodes met consensus)

MinRank	Score	Rank	Fatal	Diff
Slovenië	0,74	1	12	+11
Zweden	0,71	2	1	-1
Ierland	0,71	3	7	+4
Frankrijk	0,70	4	8	+4
Duitsland	0,68	5	5	0
Denemarken	0,68	6	5	-1
VK	0,67	7	2	-5
Finland	0,62	8	4	-4
Oostenrijk	0,61	9	10	+1
Nederland	0,61	10	3	-7
België	0,59	11	11	0
Portugal	0,56	12	19	+7
Spanje	0,55	13	14	+1
Italië	0,50	14	9	-5
Polen	0,43	15	20	+5
Tsjechië	0,43	16	17	+1
Hongarije	0,40	17	15	-2
Estland	0,37	18	12	-6
Griekenland	0,25	19	18	-1
Cyprus	0,23	20	16	-4

Tabel 19: Landenrangschikking volgens MinScore (wegingmethodes zonder consensus)

MinScore	Score	Rank	Fatal	Diff
Slovenië	0,73	1	12	+11
Zweden	0,68	2	1	-1
Denemarken	0,68	3	5	+2
Frankrijk	0,65	4	8	+4
Oostenrijk	0,63	5	10	+5
Duitsland	0,62	6	5	-1
Finland	0,61	7	4	-3
België	0,61	8	11	+3
Ierland	0,60	9	7	-2
VK	0,57	10	2	-8
Nederland	0,54	11	3	-8
Portugal	0,52	12	19	+7
Spanje	0,50	13	14	+1
Italië	0,48	14	9	-5
Tsjechië	0,48	15	17	+2
Polen	0,45	16	20	+4
Hongarije	0,42	17	15	-2
Estland	0,41	18	12	-6
Griekenland	0,25	19	18	-1
Cyprus	0,18	20	16	-4

Tabel 20: Landenrangschikking volgens MinRank (wegingmethodes zonder consensus)

MinRank	Score	Rank	Fatal	Diff
Ierland	0,74	1	7	+6
VK	0,71	2	2	0
Slovenië	0,70	3	12	+9
Frankrijk	0,69	4	8	+4
Zweden	0,66	5	1	-4
Denemarken	0,65	6	5	-1
Finland	0,65	7	4	-3
Nederland	0,63	8	3	-5
Oostenrijk	0,63	9	10	+1
Duitsland	0,62	10	5	-5
België	0,60	11	11	0
Tsjechië	0,59	12	17	+5
Polen	0,55	13	20	+7
Portugal	0,54	14	19	+5
Spanje	0,53	15	14	-1
Italië	0,48	16	9	-7
Hongarije	0,45	17	15	-2
Estland	0,41	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,20	20	16	-4

**Tabel 21: Landenrangschikking volgens CorRank
(wegingmethodes met consensus)**

CorRank	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,69	2	8	+6
Ierland	0,67	3	7	+4
Zweden	0,67	4	1	-3
Duitsland	0,66	5	5	0
Denemarken	0,65	6	5	-1
VK	0,64	7	2	-5
België	0,61	8	11	+3
Nederland	0,61	9	3	-6
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,54	12	19	+7
Spanje	0,52	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,45	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,39	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,26	19	18	-1
Cyprus	0,24	20	16	-4

**Tabel 22: Landenrangschikking volgens CorRelPos
(wegingmethodes met consensus)**

CorRelPos	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,69	2	8	+6
Ierland	0,67	3	7	+4
Zweden	0,67	4	1	-3
Duitsland	0,67	5	5	0
Denemarken	0,65	6	5	-1
VK	0,65	7	2	-5
Nederland	0,61	8	3	-5
België	0,60	9	11	+2
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,55	12	19	+7
Spanje	0,53	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,45	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,39	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,26	19	18	-1
Cyprus	0,24	20	16	-4

**Tabel 23: Landenrangschikking volgens CorRank
(wegingmethodes zonder consensus)**

CorRank	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,66	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Oostenrijk	0,62	7	10	+3
Finland	0,62	8	4	-4
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,52	13	19	+6
Spanje	0,50	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,44	17	15	-2
Estland	0,39	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

**Tabel 24: Landenrangschikking volgens CorRelPos
(wegingmethodes zonder consensus)**

CorRelPos	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,67	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Finland	0,63	7	4	-3
Oostenrijk	0,62	8	10	+2
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,53	13	19	+6
Spanje	0,51	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,44	17	15	-2
Estland	0,40	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

**Tabel 25: Landenrangschikking volgens CPR
(wegingmethodes met consensus)**

CPS	Score	Rank	Fatal	Diff
Slovenië	0,70	1	12	+11
Frankrijk	0,70	2	8	+6
Ierland	0,67	3	7	+4
Zweden	0,67	4	1	-3
Duitsland	0,67	5	5	0
Denemarken	0,65	6	5	-1
VK	0,65	7	2	-5
Nederland	0,61	8	3	-5
België	0,60	9	11	+2
Oostenrijk	0,59	10	10	0
Finland	0,59	11	4	-7
Portugal	0,55	12	19	+7
Spanje	0,53	13	14	+1
Italië	0,49	14	9	-5
Tsjechië	0,45	15	17	+2
Polen	0,43	16	20	+4
Hongarije	0,39	17	15	-2
Estland	0,35	18	12	-6
Griekenland	0,25	19	18	-1
Cyprus	0,24	20	16	-4

**Tabel 26: Landenrangschikking volgens CPR
(wegingmethodes zonder consensus)**

CPS	Score	Rank	Fatal	Diff
Slovenië	0,71	1	12	+11
Ierland	0,67	2	7	+5
Frankrijk	0,67	3	8	+5
Denemarken	0,66	4	5	+1
Zweden	0,66	5	1	-4
VK	0,64	6	2	-4
Finland	0,63	7	4	-3
Oostenrijk	0,62	8	10	+2
Duitsland	0,61	9	5	-4
België	0,61	10	11	+1
Nederland	0,58	11	3	-8
Tsjechië	0,54	12	17	+5
Portugal	0,53	13	19	+6
Spanje	0,51	14	14	0
Polen	0,50	15	20	+5
Italië	0,48	16	9	-7
Hongarije	0,44	17	15	-2
Estland	0,40	18	12	-6
Griekenland	0,27	19	18	-1
Cyprus	0,18	20	16	-4

Tabel 27: Analyse van de landenrangschikkingen (wegingmethodes met consensus)

Land	FAT.	Math	Geo	ADM	SDM	MinS core	MinR ank	CorR el	CorR ank	CPS	+	-	=	>3	>5
BEL	11	8	7	8	8	11	11	9	8	9	7	0	2	1	0
CYP	16	20	20	20	20	20	20	20	20	20	0	9	0	9	0
DEN	5	6	4	6	6	7	6	6	6	6	1	8	0	0	0
DUI	5	5	5	5	5	5	5	5	5	5	0	0	9	0	0
EST	12	18	18	18	18	18	18	18	18	18	0	9	0	9	9
FIN	4	11	10	11	11	9	8	11	11	11	0	9	0	9	7
FRA	8	2	2	2	2	2	4	2	2	2	9	0	0	9	8
GRI	18	19	19	19	19	19	19	19	19	19	0	9	0	0	0
HON	15	17	17	17	17	17	17	17	17	17	0	9	0	0	0
IER	7	4	6	3	4	3	3	3	3	3	9	0	0	6	0
ITA	9	14	14	14	14	14	14	14	14	14	0	9	0	9	0
NED	3	9	11	9	9	8	10	8	9	8	0	9	0	9	6
OOS	10	10	9	10	10	10	9	10	10	10	2	0	7	0	0
POL	20	16	16	16	16	16	15	16	16	16	9	0	0	9	0
POR	19	12	12	12	12	12	12	12	12	12	9	0	0	9	9
SLO	12	1	1	1	1	1	1	1	1	1	9	0	0	9	9
SPA	14	13	13	13	13	13	13	13	13	13	9	0	0	0	0
TSJ	17	15	15	15	15	15	16	15	15	15	9	0	0	0	0
VK	2	7	8	7	7	6	7	7	7	7	0	9	0	9	1
ZWE	1	3	3	4	3	4	2	4	4	4	0	9	0	0	0

Tabel 28: Analyse van de landenrangschikkingen (wegingmethodes zonder consensus)

Land	FAT.	Math	Geo	ADM	SDM	MinS core	MinR ank	CorR el	CorR ank	CPS	+	-	=	>3	>5
BEL	11	10	7	10	10	8	11	10	10	10	8	0	1	1	0
CYP	16	20	20	20	20	20	20	20	20	20	0	9	0	9	0
DEN	5	4	8	4	4	3	6	4	4	4	7	2	0	0	0
DUI	5	9	6	9	9	6	10	9	9	9	0	9	0	7	0
EST	12	18	18	18	18	18	18	18	18	18	0	9	0	9	9
FIN	4	8	11	8	8	7	7	7	8	7	0	9	0	5	1
FRA	8	3	3	3	3	4	4	3	3	3	9	0	0	9	0
GRI	18	19	19	19	19	19	19	19	19	19	0	9	0	0	0
HON	15	17	17	17	17	17	17	17	17	17	0	9	0	0	0
IER	7	2	1	2	2	9	1	2	2	2	8	1	0	8	2
ITA	9	16	16	16	16	14	16	16	16	16	0	9	0	9	8
NED	3	11	5	11	11	11	8	11	11	11	0	9	0	8	7
OOS	10	7	10	7	7	5	9	8	7	8	8	0	1	1	0
POL	20	15	15	15	15	16	13	15	15	15	9	0	0	9	1
POR	19	13	13	13	13	12	14	13	13	13	9	0	0	9	8
SLO	12	1	4	1	1	1	3	1	1	1	9	0	0	9	9
SPA	14	14	14	14	14	13	15	14	14	14	1	1	7	0	0
TSJ	17	12	12	12	12	15	12	12	12	12	9	0	0	8	0
VK	2	6	2	6	6	10	2	6	6	6	0	7	2	7	1
ZWE	1	5	9	5	5	2	5	5	5	5	0	9	0	8	1

8.4 Resultaten van de wijzigingen

8.4.1 MinRank & MinScore: gewichten

Tabel 29: Overzicht geaggregeerde gewichten voor MinRank onder verschillende omstandigheden (wegingmethodes zonder consensus)

MINRANK	Original	Per km	Per car	Select	Detail
Alc	0,234	0,247	0,261	0,140	0,214
Speed	0,252	0,258	0,273	0,177	0,213
Seatbelt	0,178	0,192	0,184	0,157	0,189
DRL	0,053	0,043	0,018	0,171	0,083
Veh	0,083	0,074	0,073	0,137	0,095
Infr?	0,120	0,118	0,137	0,102	0,116
Trauma	0,080	0,068	0,055	0,117	0,089

Tabel 30: Overzicht geaggregeerde gewichten voor MinRank onder verschillende omstandigheden (wegingmethodes met consensus)

MINRANK	Original	Per km	Per car	Select	Detail
Alc	0,116	0,087	0,098	0,105	0,118
Speed	0,124	0,153	0,140	0,134	0,134
Seatbelt	0,240	0,291	0,275	0,261	0,203
DRL	0,096	0,042	0,062	0,075	0,112
Veh	0,226	0,165	0,194	0,207	0,192
Infr?	0,068	0,101	0,083	0,078	0,102
Trauma	0,130	0,161	0,147	0,140	0,139

Tabel 31: Overzicht geaggregeerde gewichten voor MinScore onder verschillende omstandigheden (wegingmethodes zonder consensus)

MINSCORE	Original	Per km	Per car	Select	Detail
Alc	0,129	0,180	0,131	0,124	0,163
Speed	0,150	0,183	0,150	0,152	0,171
Seatbelt	0,140	0,166	0,141	0,140	0,157
DRL	0,173	0,117	0,169	0,183	0,135
Veh	0,148	0,118	0,147	0,150	0,128
Infr?	0,123	0,125	0,126	0,118	0,126
Trauma	0,136	0,112	0,137	0,135	0,121

Tabel 32: Overzicht geaggregeerde gewichten voor MinScore onder verschillende omstandigheden (wegingmethodes met consensus)

MINSCORE	Original	Per km	Per car	Select	Detail
Alc	0,107	0,111	0,111	0,088	0,111
Speed	0,137	0,143	0,133	0,151	0,133
Seatbelt	0,246	0,209	0,238	0,290	0,238
DRL	0,083	0,102	0,091	0,044	0,091
Veh	0,196	0,171	0,203	0,169	0,203
Infr?	0,089	0,115	0,085	0,098	0,085
Trauma	0,143	0,148	0,139	0,159	0,139

8.4.2 MinRank & MinScore: Rangschikkingen

Tabel 33: Overzicht landenrangschikkingen voor MinScore onder verschillende omstandigheden (wegingmethodes met consensus)

MINSCORE	Origin	Per km	Per car	Select	Detail
België	11	8	11	9	11
Cyprus	20	20	20	19	20
Denemarken	7	5	6	7	6
Duitsland	5	6	5	5	5
Estland	18	18	18	18	18
Finland	9	11	9	10	9
Frankrijk	2	2	2	1	2
Griekenland	19	19	19	20	19
Hongarije	17	17	17	17	17
Ierland	3	4	3	2	3
Italië	14	14	14	14	14
Nederland	8	10	8	8	8
Oostenrijk	10	9	10	11	10
Polen	16	16	16	16	16
Portugal	12	12	12	12	12
Slovenië	1	1	1	3	1
Spanje	13	13	13	13	13
Tsjechië	15	15	15	15	15
VK	6	7	7	4	7
Zweden	4	3	4	6	4

Tabel 35: Overzicht landenrangschikkingen voor MinRank onder verschillende omstandigheden (wegingmethodes met consensus)

MINRANK	Origin	Per km	Per car	Select	Detail
België	11	9	11	11	9
Cyprus	20	19	19	20	20
Denemarken	6	7	7	7	4
Duitsland	5	5	4	5	6
Estland	18	18	18	18	18
Finland	8	10	9	9	10
Frankrijk	4	1	1	2	3
Griekenland	19	20	20	19	19
Hongarije	17	17	17	17	17
Ierland	3	2	2	3	5
Italië	14	14	14	14	14
Nederland	10	8	8	8	11
Oostenrijk	9	11	10	10	8
Polen	15	16	16	16	16
Portugal	12	12	12	12	12
Slovenië	1	4	3	1	1
Spanje	13	13	13	13	13
Tsjechië	16	15	15	15	15
VK	7	3	6	6	7
Zweden	2	6	5	4	2

Tabel 34: Overzicht landenrangschikkingen voor MinScore onder verschillende omstandigheden (wegingmethodes zonder consensus)

MINSCORE	Origin	Per km	Per car	Select	Detail
België	8	10	7	8	10
Cyprus	20	20	20	20	20
Denemarken	3	4	3	3	3
Duitsland	6	8	6	6	7
Estland	18	18	18	18	18
Finland	7	7	8	7	8
Frankrijk	4	3	4	4	4
Griekenland	19	19	19	19	19
Hongarije	17	17	17	17	17
Ierland	9	5	9	9	5
Italië	14	16	14	14	15
Nederland	11	11	11	11	11
Oostenrijk	5	9	5	5	6
Polen	16	15	16	16	16
Portugal	12	12	12	12	12
Slovenië	1	1	1	1	1
Spanje	13	14	13	13	14
Tsjechië	15	13	15	15	13
VK	10	6	10	10	9
Zweden	2	2	2	2	2

Tabel 36: Overzicht landenrangschikkingen voor MinRank onder verschillende omstandigheden (wegingmethodes zonder consensus)

MINRANK	Origin	Per km	Per car	Select	Detail
België	11	11	11	9	11
Cyprus	20	20	20	20	20
Denemarken	6	7	7	3	6
Duitsland	10	10	10	8	9
Estland	18	18	18	18	18
Finland	7	5	6	6	7
Frankrijk	4	4	3	4	3
Griekenland	19	19	19	19	19
Hongarije	17	17	17	17	17
Ierland	1	1	1	7	2
Italië	16	16	16	15	16
Nederland	8	8	5	11	10
Oostenrijk	9	9	9	5	8
Polen	13	13	13	16	14
Portugal	14	14	14	12	13
Slovenië	3	3	4	1	1
Spanje	15	15	15	13	15
Tsjechië	12	12	12	14	12
VK	2	2	2	10	4
Zweden	5	6	8	2	5

8.4.3 Overige aggregatiemethodes: Gewichten

Tabel 37: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes met consensus)

PER KM	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,106	0,134	0,105	0,105	0,104	0,105	0,104
Speed	0,148	0,149	0,147	0,148	0,147	0,147	0,147
Seatbelt	0,218	0,162	0,227	0,222	0,229	0,226	0,232
DRL Law	0,093	0,119	0,087	0,090	0,086	0,088	0,084
Vehicle	0,160	0,150	0,164	0,162	0,164	0,163	0,165
Infrastructure	0,121	0,136	0,116	0,119	0,116	0,117	0,114
Trauma Care	0,154	0,150	0,153	0,153	0,153	0,153	0,153

Tabel 38: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes zonder consensus)

PER KM	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,181	0,194	0,181	0,181	0,181	0,181	0,182
Speed	0,210	0,217	0,210	0,211	0,211	0,210	0,211
Seatbelt	0,144	0,168	0,145	0,144	0,145	0,145	0,146
DRL Law	0,111	0,000	0,111	0,111	0,110	0,111	0,110
Vehicle	0,110	0,132	0,110	0,111	0,110	0,110	0,110
Infrastructure	0,128	0,154	0,128	0,128	0,127	0,127	0,126
Trauma Care	0,116	0,135	0,116	0,116	0,115	0,116	0,115

Tabel 39: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes met consensus)

PER CAR	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,244	0,191	0,215	0,232	0,205	0,202	0,203
Speed	0,201	0,192	0,187	0,196	0,182	0,180	0,181
Seatbelt	0,150	0,170	0,169	0,157	0,175	0,168	0,178
DRL Law	0,079	0,000	0,079	0,079	0,080	0,086	0,079
Vehicle	0,119	0,154	0,136	0,126	0,142	0,139	0,144
Infrastructure	0,099	0,140	0,096	0,098	0,095	0,103	0,093
Trauma Care	0,108	0,153	0,118	0,112	0,121	0,122	0,122

Tabel 40: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes zonder consensus)

PER CAR	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,181	0,194	0,181	0,181	0,181	0,180	0,181
Speed	0,210	0,217	0,210	0,211	0,210	0,209	0,210
Seatbelt	0,144	0,168	0,145	0,144	0,145	0,145	0,145
DRL Law	0,111	0,000	0,111	0,111	0,111	0,112	0,111
Vehicle	0,110	0,132	0,110	0,111	0,110	0,111	0,110
Infrastructure	0,128	0,154	0,128	0,128	0,126	0,127	0,126
Trauma Care	0,116	0,135	0,116	0,116	0,116	0,116	0,116

Tabel 41: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes met consensus)

SELECTION	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,106	0,134	0,105	0,106	0,105	0,105	0,105
Speed	0,148	0,149	0,147	0,148	0,146	0,147	0,145
Seatbelt	0,218	0,162	0,223	0,220	0,227	0,226	0,233
DRL Law	0,093	0,119	0,090	0,092	0,088	0,089	0,086
Vehicle	0,160	0,150	0,162	0,161	0,166	0,165	0,172
Infrastructure	0,121	0,136	0,119	0,120	0,115	0,116	0,109
Trauma Care	0,154	0,150	0,153	0,153	0,152	0,153	0,151

Tabel 42: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes zonder consensus)

SELECTION	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,181	0,194	0,179	0,179	0,177	0,177	0,175
Speed	0,210	0,217	0,208	0,209	0,207	0,207	0,206
Seatbelt	0,144	0,168	0,143	0,144	0,144	0,143	0,144
DRL Law	0,111	0,000	0,113	0,112	0,116	0,115	0,118
Vehicle	0,110	0,132	0,112	0,111	0,112	0,112	0,113
Infrastructure	0,128	0,154	0,128	0,128	0,126	0,127	0,125
Trauma Care	0,116	0,135	0,117	0,117	0,118	0,118	0,119

Tabel 43: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes met consensus)

DETAIL	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,106	0,134	0,104	0,105	0,104	0,105	0,104
Speed	0,148	0,149	0,148	0,148	0,147	0,147	0,146
Seatbelt	0,218	0,162	0,228	0,223	0,231	0,228	0,233
DRL Law	0,093	0,119	0,087	0,090	0,086	0,087	0,084
Vehicle	0,160	0,150	0,163	0,162	0,166	0,165	0,167
Infrastructure	0,121	0,136	0,117	0,119	0,114	0,116	0,113
Trauma Care	0,154	0,150	0,154	0,154	0,153	0,153	0,153

Tabel 44: Geaggregeerde gewichten voor de verschillende aggregatiemethodes (wegingmethodes zonder consensus)

DETAIL	MathMean	GeoMean	ADM	SDM	CorRelPos	CorRank	CPS
Alcohol	0,181	0,194	0,183	0,181	0,182	0,179	0,182
Speed	0,210	0,217	0,213	0,211	0,211	0,209	0,211
Seatbelt	0,144	0,168	0,145	0,144	0,145	0,144	0,145
DRL Law	0,111	0,000	0,108	0,110	0,109	0,112	0,109
Vehicle	0,110	0,132	0,109	0,110	0,111	0,112	0,111
Infrastructure	0,128	0,154	0,127	0,128	0,128	0,128	0,128
Trauma Care	0,116	0,135	0,114	0,116	0,114	0,116	0,115

8.5 Robuustheid

8.5.1 MinScore & MinRank

Tabel 45: Correlatie tussen gewichten - Per km (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	-0,036	0,893
Kendall	0,048	0,810

Tabel 46: Correlatie tussen gewichten - Per car (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,964	0,964
Kendall	0,905	0,905

Tabel 47: Correlatie tussen gewichten - Selection (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,964	0,964
Kendall	0,905	0,905

Tabel 48: Correlatie tussen gewichten - Detail (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,964	1,000
Kendall	0,905	1,000

Tabel 49: Correlatie tussen rankings - Per km (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,980	0,946
Kendall	0,916	0,832

Tabel 50: Correlatie tussen rankings - Per car (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,998	0,973
Kendall	0,989	0,884

Tabel 51: Correlatie tussen gewichten - Per km (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	-0,071	0,995
Kendall	-0,048	1,000

Tabel 52: Correlatie tussen gewichten - Per car (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	1,000	0,983
Kendall	1,000	1,000

Tabel 53: Correlatie tussen gewichten - Selection (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	1,000	0,321
Kendall	1,000	0,238

Tabel 54: Correlatie tussen gewichten - Detail (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,357	0,986
Kendall	0,238	0,905

Tabel 55: Correlatie tussen rankings - Per km (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,949	0,995
Kendall	0,842	0,979

Tabel 56: Correlatie tussen rankings - Per car (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,998	0,973
Kendall	0,989	0,884

Tabel 57: Correlatie tussen rankings - Selection (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,983	0,986
Kendall	0,916	0,926

Tabel 58: Correlatie tussen rankings – Detail (met consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,998	0,983
Kendall	0,989	0,926

Tabel 59: Correlatie tussen rankings - Selection (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,983	0,986
Kendall	0,916	0,926

Tabel 60: Correlatie tussen rankings – Detail (zonder consensus)

Correlatie coëfficiënt	MinScore	MinRank
Spearman	0,998	0,983
Kendall	0,989	0,926

8.5.2 Overige aggregatiemethodes (enkel correlaties tussen gewichtensets)

Tabel 61: Correlatie tussen gewichten - Per km (met consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	1,000	1,000	1,000
Kendall	1,000	1,000	1,000	1,000	1,000

Tabel 65: Correlatie tussen gewichten - Per km (zonder consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	1,000	1,000	1,000
Kendall	1,000	1,000	0,905	1,000	0,905

Tabel 62: Correlatie tussen gewichten - Per car (met consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	1,000	1,000	1,000
Kendall	1,000	1,000	1,000	1,000	1,000

Tabel 66: Correlatie tussen gewichten - Per car (zonder consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	0,998	1,000	1,000
Kendall	1,000	1,000	0,905	1,000	0,905

Tabel 63: Correlatie tussen gewichten - Selection (met consensus)

Correlatie coëfficiënt	ADM	SDM	CorRe IPos	Cor- Rank	CPS
Spearman	1,000	1,000	1,000	1,000	1,000
Kendall	1,000	1,000	1,000	1,000	1,000

Tabel 67: Correlatie tussen gewichten - Selection (zonder consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	0,964	1,000	0,964
Kendall	1,000	1,000	0,905	1,000	0,905

Tabel 64: Correlatie tussen gewichten - Detail (met consensus)

Correlatie coëfficiënt	ADM	SDM	CorRel Pos	Cor- Rank	CPS
Spearman	1,000	1,000	1,000	1,000	1,000
Kendall	1,000	1,000	1,000	1,000	1,000

Tabel 68: Correlatie tussen gewichten - Detail (zonder consensus)

Correlatie coëfficiënt	ADM	SDM	CorRe IPos	Cor- Rank	CPS
Spearman	0,998	1,000	0,998	1,000	0,998
Kendall	0,905	0,905	1,000	1,000	1,000

Tabel 69: Genormaliseerde waarden per SPI

Country	Alcohol	Speed	Seatbelt	DRL	Vehicle	Infr	Trauma
België	0,75	0,25	0,50	0,50	0,75	1,00	0,75
Cyprus	0,00	0,00	0,75	0,00	0,00	0,25	0,25
Denemarken	1,00	0,25	0,75	1,00	0,50	0,25	0,75
Duitsland	1,00	0,00	1,00	0,50	0,50	0,25	1,00
Estland	1,00	0,00	0,75	1,00	0,00	0,25	0,00
Finland	1,00	0,75	1,00	1,00	0,25	0,00	0,25
Frankrijk	0,75	0,75	1,00	0,50	0,50	0,25	0,75
Griekenland	0,75	0,00	0,00	0,00	0,25	0,00	0,75
Hongarije	1,00	0,00	0,25	0,50	0,25	0,25	0,50
Ierland	1,00	1,00	0,75	0,00	1,00	0,25	0,25
Italië	0,75	0,25	0,50	0,50	0,50	0,25	0,50
Nederland	1,00	0,25	1,00	0,00	0,50	0,75	0,75
Oostenrijk	1,00	0,50	0,75	1,00	0,50	0,25	0,50
Polen	1,00	0,50	0,50	0,50	0,00	0,25	0,00
Portugal	1,00	0,25	1,00	0,50	0,25	0,00	0,75
Slovenië	1,00	0,75	0,75	1,00	1,00	0,25	0,50
Spanje	0,75	0,50	1,00	0,50	0,50	0,00	0,50
Tsjechië	1,00	1,00	0,25	0,50	0,00	0,25	0,50
VK	1,00	0,75	1,00	0,00	0,75	0,25	0,50
Zweden	1,00	0,25	1,00	1,00	0,50	0,00	0,75