



Ministerie van Infrastructuur  
en Waterstaat

# Beoordeling internationale benchmark spoor

Kennisinstituut voor Mobiliteitsbeleid | KiM



# Inhoud

**Samenvatting en conclusies 5**

**1 Inleiding 7**

**2 Uitvoering van de benchmark: de algemene aanpak 9**

**3 Uitvoering van de benchmark: geselecteerde onderwerpen 12**

**4 Bruikbaarheid van de resultaten 17**

**Literatuur 19**

**Colofon 20**



# Samenvatting en conclusies

## Aanleiding en doel van deze notitie

In opdracht van het ministerie van Infrastructuur en Waterstaat (IenW) hebben ProRail en NS een internationale benchmark uitgevoerd, zoals voorgeschreven is in respectievelijk de beheer- en vervoerconcessie. De benchmark vergelijkt een breed scala aan indicatoren en karakteristieken van het spoorstelsel voor de periode 2011-2015. De directie Openbaar Vervoer en Spoor (OVS) van het Ministerie van IenW heeft het Kennisinstituut voor Mobiliteitsbeleid (KiM) gevraagd, een oordeel over de benchmarkrapportage van ProRail en NS te geven. Daarover gaat deze notitie.

De beoordeling heeft drie doelen:

1. Nagaan of de benchmark op een verantwoorde manier is uitgevoerd.
2. Inzicht geven in de bruikbaarheid van de resultaten voor het trekken van lessen en voor verdere beleidsvorming.
3. Aanbevelingen geven voor toekomstige benchmarks.

## Focus op vier onderwerpen

In eerste instantie heeft het KiM een algemene analyse gedaan. Daarnaast hebben wij vanwege de omvang van de benchmarkrapportage in samenspraak met de opdrachtgever vier onderwerpen geselecteerd die wij (meer) in detail hebben onderzocht: punctualiteit, vervoersprestatie, veiligheid en financiën.

## Benchmarkrapportage is een belangrijke eerste stap

Benchmarken is zeker bij een internationale vergelijking niet eenvoudig. Het is lastig om vergelijkbare landen of spoorvervoerders te vinden die bereid zijn mee te werken. Iedere organisatie hanteert eigen definities en methoden van dataverzameling. Desondanks hebben ProRail en NS een grote hoeveelheid informatie verzameld en daar veel werk voor verzet. Zij hebben veel aandacht besteed aan een correcte feitenverzameling en verwerking van de data. Onze indruk is dat de benchmark in grote lijnen op verantwoorde wijze is uitgevoerd. De gepresenteerde scores geven een eerste inzicht in hoe de Nederlandse spoorsector presteert in vergelijking met een selectie andere landen of vervoerbedrijven. Daarmee is de benchmarkrapportage een belangrijke eerste stap. Vervolgstappen zijn echter nodig om de bruikbaarheid van de benchmark te vergroten.

## Benchmarken is een proces: de volgende stap is 'verklaren'

De literatuur over benchmarking benadrukt dat benchmarking nuttig is als er een continu proces van gemaakt wordt. Na het vaststellen van het doel, het vinden van partners en het verzamelen van data komt de analysefase. Daarop volgt een aanpassingsfase, waarbij de lessen geïmplementeerd worden. Hierna volgt een volgende ronde van benchmarking, om te kunnen nagaan of de aanpassingen gewerkt hebben en of er weer nieuwe lessen te leren zijn.

De huidige benchmarkrapportage heeft de eerste fasen doorlopen en geeft een eerste analyse. In overeenstemming met de concessies aan ProRail en NS gaat deze analyse niet verder dan het presenteren van de data. Het verklaren van verschillen gebeurt bijna niet. Om echt lessen te kunnen trekken uit een benchmark is een zoektocht naar verklaringen voor geconstateerde prestatieverschillen een noodzakelijke volgende stap. ProRail en NS zijn zich blijkens de gesprekken met het KiM terdege van deze situatie bewust en hebben de ambitie hun analyses verder te verdiepen.

### Voorzichtig zijn met trekken van inhoudelijke conclusies

De gepresenteerde scores geven een eerste inzicht in hoe de Nederlandse spoorsector presteert in vergelijking met enkele andere landen. Deze conclusie kan niet worden veralgemeniseerd naar de positie van de Nederlandse spoorsector in bijvoorbeeld Europa of de westerse wereld om de volgende redenen:

- De selectie van landen bepaalt mede hoe Nederland scoort. Frankrijk, Denemarken, het Verenigd Koninkrijk en Zweden zijn niet zo goed te vergelijken met Nederland vanwege verschillen in omvang, ruimtelijke structuur, structuur van het spoornet en/of de vervoerspatronen (zoals de verdeling tussen lange en korte reisafstanden). Vergelijking met deze landen is daarom alleen van nut voorzover structuurkenmerken geen rol spelen.
- Keuzes omtrent indicatoren en de operationalisering daarvan kunnen in grote mate de conclusie beïnvloeden.
- Uitschieters in de data waarvan de reden niet bekend is beïnvloeden in sommige gevallen sterk de resultaten.

### Aanbevelingen aan ProRail en NS voor toekomstige benchmarks

- Voorkom de verwarring die kan ontstaan doordat de analyses van NS en ProRail in één rapportage zijn samengevoegd. Dit speelt in het bijzonder bij onderwerpen waarover beide partijen rapporteren maar waarbij zij verschillende definities hebben gehanteerd.
- De vergelijkingslanden in de benchmark zijn verschillend voor vervoerders en infraproviders. Idealiter zouden ze gelijk moeten zijn. De vergelijkbaarheid wordt hierdoor vergroot, bijvoorbeeld omdat kruisvergelijkingen van de output van infraproviders met die van vervoerders mogelijk worden. Dat vergroot ook het leereffect.
- Een andere optie om vergelijkbaarheid te vergroten is om niet te kijken naar hele landen maar naar landsdelen. Zo lijken het Duitse Noordrijn-Westfalen en het Japanse eiland Kyushu qua inwonertal, ruimtelijke structuur, verstedelijkingsgraad en netwerkstructuur veel meer op Nederland dan menig ander, compleet land. Wij realiseren ons dat dit wel de nodige dataproblemen met zich meebrengt.
- De logische volgende stap is om meer de diepte in te gaan. Het is waarschijnlijk te arbeidsintensief en ook niet zinvol dit te doen voor alle onderwerpen. Kies dan een beperkt aantal onderwerpen vanuit actuele vraagstukken en werk die verder uit.
- In veel diagrammen van ProRail zit een vergelijkingspartner (land C) die zich in ongeveer dezelfde positie bevindt, maar vaak net iets beter scoort. Het kan interessant zijn om in een vervolgfase verder op deze partner te focussen en in verdiepende gesprekken te verkennen wat deze anders/beter doet.

### Aanbevelingen aan OVS voor toekomstige benchmarks

- Verdiep in de toekomst de opdracht aan ProRail en NS door niet langer uitsluitend primaire prestatiegegevens te vragen, maar ook verklaringen voor gevonden verschillen die rekening houden met achtergronden en omstandigheden in te vergelijken landen. Dit vergroot de potentiële leereffecten aanzienlijk.
- Om verwarring te voorkomen bevelen wij aan om het rapportagedeel van ProRail en dat van NS beter op elkaar af te laten stemmen voor zover zij over dezelfde indicatoren rapporteren. Lukt dit niet, dan is het een optie om terug te vallen op twee afzonderlijke benchmarkrapportages.
- Kies te onderzoeken onderwerpen die beleidsrelevant zijn en ook beleidsmatig beïnvloed kunnen worden. Dit kan gaan om onderwerpen als financiële sturing, punctualiteit en veiligheid. Laat het benchmarken van de meer operationele aspecten aan ProRail en NS zelf. Op deze manier krijgt de analyse meer focus en kan meer tijd besteed worden aan het verklaren van verschillen en de leereffecten.

# 1

# Inleiding

## Aanleiding

In opdracht van het ministerie van Infrastructuur en Waterstaat (IenW) hebben ProRail en NS een internationale benchmark uitgevoerd, zoals voorgeschreven is in respectievelijk de beheer- en vervoerconcessie (Ministerie van IenM, 2014a en 2014b). De benchmark vergelijkt een breed scala aan indicatoren en karakteristieken over de periode 2011-2015, zoals de betrouwbaarheid, duurzaamheid, capaciteit en financiële prestaties (ProRail en NS, 2017).

De directie Openbaar Vervoer en Spoor (OVS) van het Ministerie van Infrastructuur en Waterstaat heeft het Kennisinstituut voor Mobiliteitsbeleid (KiM) gevraagd, een oordeel over deze benchmark te geven.

## Doel van de beoordeling

De beoordeling heeft drie doelen:

1. Nagaan of de benchmark op een verantwoorde manier is uitgevoerd.
2. Inzicht geven in de bruikbaarheid van de resultaten voor het trekken van lessen en verdere beleidsvorming.
3. Aanbevelingen geven voor toekomstige benchmarks.

## Onderzoeksvragen

1. Is de benchmark op een verantwoorde wijze uitgevoerd?
  - a. Hebben NS en ProRail de juiste data uitgevraagd en ontvangen?
  - b. Hebben NS en ProRail de ontvangen data op verantwoorde wijze verwerkt in de rapportage?
  - c. Zijn bewerkingen van de oorspronkelijke data op een (wetenschappelijk) verantwoorde manier gedaan?
2. Wat is de bruikbaarheid van de resultaten?
  - a. Welke type conclusies zijn wél en welke niet te trekken uit de resultaten van de internationale benchmark?
  - b. Welke aandachtspunten zijn er bij de interpretatie van de resultaten?

## Aanpak KiM

NS en ProRail hebben het eindconcept van hun benchmarkrapportage aan het KiM ter beschikking gesteld. ProRail heeft het KiM de onderliggende data vertrouwelijk aangeleverd voor de beoordeling. NS heeft het KiM niet de dataset geleverd, maar heeft ter plekke inzicht gegeven in de wijze van verzameling en verwerking van de data, in overeenstemming met geheimhoudingsafspraken met de vergelijkingspartners. Dit heeft tot gevolg dat het KiM de analyses en resultaten van NS heeft kunnen beoordelen voorzover dit na mondelinge toelichting door onze gesprekspartner mogelijk was.

Wat de eerste onderzoeksvraag betreft, is de benchmarkrapportage te omvangrijk om binnen de beschikbare (doorloop)tijd te verifiëren of over alle onderwerpen op verantwoorde wijze gerapporteerd is. Het KiM heeft daarom in samenspraak met OVS een beperkt aantal onderwerpen geselecteerd dat (meer) in detail is onderzocht. Belangrijkste criterium voor deze selectie is de beleidsmatige relevantie die vooral is gebaseerd op de doelen uit de Lange Termijn Spooragenda (Ministerie van IenM, 2013).

De selectie betreft de volgende onderwerpen:

1. Punctualiteit
2. Vervoerprestatie
3. Veiligheid
4. Financiële sturing

Het KiM heeft de uitwerking van de onderwerpen van de benchmark met beide partijen besproken, bij ProRail in samenhang met de aangeleverde achtergronddata.

Wij hebben de onderzoeksvragen beantwoord aan de hand van algemene bevindingen en bevindingen uit de verdiepingsslag bij de geselecteerde onderwerpen. Wij hebben een conceptversie van deze beoordeling voorgelegd aan ProRail en NS voor een toets op feitelijke onjuistheden. Naar aanleiding van onze voorlopige beoordeling hebben zij de definitieve benchmarkrapportage opgesteld. Vervolgens hebben wij de beoordeling definitief gemaakt. We hebben alleen gebruik gemaakt van de door NS en ProRail aangeleverde informatie en mondelinge toelichtingen.

### **Leeswijzer**

Hoofdstuk 2 beschrijft en beoordeelt de algemene aanpak van de benchmark. Hoofdstuk 3 beschrijft en beoordeelt de aanpak en resultaten voor de vier geselecteerde onderwerpen. Hoofdstuk 4 tot slot gaat over de bruikbaarheid van de resultaten voor verdere beleidsontwikkeling.



# 2

## Uitvoering van de benchmark: de algemene aanpak

Dit hoofdstuk beschrijft allereerst de algemene aanpak door ProRail en NS van de benchmark: de keuze en diepgang van onderwerpen, de keuze van de vergeleken landen/regio's, de gevolgde werkwijze en de weergave van de resultaten. In de laatste paragraaf geeft het KiM een beoordeling van deze aanpak.

### Onderwerpen van de benchmark

De benchmarkrapportage gaat over een groot aantal karakteristieken van het spoorstelsel. Deze betreffen de kenmerken van het spoorstelsel en de stations, de infrastructuur, veiligheid, de aantrekkelijkheid van het spoorproduct, betrouwbaarheid, duurzaamheid, de capaciteit, de productiviteit en de financiële prestaties. ProRail en NS hebben in samenspraak met het ministerie van IenW een breed pakket aan te benchmarken onderwerpen vastgesteld. Vervolgens zijn al deze onderwerpen vertaald in één of meerdere indicatoren waarmee de prestaties zijn gemeten en vergeleken met die in andere landen. Daarbij is onderscheid gemaakt tussen de kenmerken van de infrastructuur en het vervoer over die infrastructuur.

Als basis hebben ProRail en NS gekozen voor beschikbare, internationaal gebruikte indicatoren. Aanvullend zijn de indicatoren gekozen die aansluiten bij in Nederland gebruikelijke prestatie-indicatoren.

### Keuze van de vergeleken landen/regio's

ProRail en NS hebben meerdere criteria gehanteerd voor de keuze van landen en bedrijven waar zij hun prestaties mee hebben vergeleken (de zogeheten 'peers'). De benchmarkrapportage noemt de vergelijkbaarheid qua netwerkstructuur, de verwachte leereffecten, de beschikbare contacten, de bereidheid tot medewerking, de deelname aan de eerdere benchmarkstudie en de verwachte beschikbaarheid van data. De gekozen peers zijn grotendeels dezelfde als bij de eerdere benchmark over 2011 (ProRail) en 2013 (NS). ProRail heeft vergeleken met infrastructuurmanagers in België, Denemarken, Zweden, Frankrijk, Groot-Brittannië en Zwitserland. Dit zijn allemaal landen in het (noord)westen van Europa, vanuit de gedachte dat die het best vergelijkbaar zijn. Duitsland wilde niet meewerken.

NS heeft haar prestaties vergeleken met vervoerders in vier van de zes hiervoor genoemde landen (Frankrijk en Zweden ontbreken). Binnen Groot-Brittannië heeft NS vergeleken met vier vervoerbedrijven, zodat er in totaal zeven vergelijkingspartners zijn: NMBS (België), DSB (Denemarken), SBB (Zwitserland) en vier Britse bedrijven (Greater Anglia, ScotRail, Northern Rail en Merseyrail). De keuzevrijheid voor NS is beperkt omdat dochterbedrijven van NS soms (potentiële) concurrenten zijn bij biedingen op concessies. Daarom ontbreken bijvoorbeeld Duitsland en een Zweedse partij.

## Werkwijze ProRail en NS

Voor het vullen van de indicatoren uit de benchmark hebben ProRail en NS gebruik gemaakt van openbare bronnen, informatie uit platforms waarin zij participeren en rechtstreekse informatie van infrastructuurbeheerders en vervoerders uit de genoemde landen. Hoewel de resultaten zijn samengevoegd in één rapportage, zijn de benchmarks van NS en ProRail grotendeels onafhankelijk van elkaar uitgevoerd. Ze zijn aan het einde van het proces samengevoegd in één rapportage. Aan de helderheid van het weergegeven logo rechtsboven op elke pagina is te zien welke organisatie de desbetreffende pagina heeft ingevuld.

### Werkwijze ProRail

ProRail heeft de indicatoren gedefinieerd en uitgevraagd aan de *peers*. Deze infrastructuurbeheerders hebben de informatie aangeleverd. Vervolgens heeft ProRail deze eerste data met de *peers* (afgezien van Groot-Brittannië) besproken, waarna de *peers* eventuele onduidelijkheden in de data aangepast hebben. Een concept met resultaten is ter validatie aan de *peers* gestuurd, waarna de scores van sommige indicatoren nogmaals aangepast zijn. Op deze manier zijn de indicatoren in drie ronden vastgesteld.

### Werkwijze NS

NS heeft in eerste instantie gezocht naar indicatoren die beschikbaar zijn in openbare bronnen. Vervolgens is de *peers* een vragenlijst gestuurd met de indicatoren. Omdat deze met name bij punctualiteit en klanttevredenheid minder gestandaardiseerd zijn aangeleverd dan bij ProRail, heeft NS eerst een bewerkingsslag uitgevoerd (zie appendix B van de benchmarkrapportage).

Vervolgens is een plausibiliteitscheck gedaan, per vervoerder op een andere wijze. Over de Britse vervoerders zijn veel openbare data beschikbaar. DSB zit met NS in een gezamenlijke benchmarkgroep, waardoor de check met bij NS beschikbare gegevens goed mogelijk was. SBB en NMBS zijn afzonderlijk benaderd met specifieke vragen. NS heeft de totale resultaten zoals gepresenteerd in de benchmarkrapportage niet ter validatie voorgelegd aan de *peers*.

## Weergave van de resultaten

ProRail en NS hebben conform hun afspraken met de *peers* de gegevens voor vrijwel alle indicatoren geanonimiseerd, alleen de score van ProRail c.q. NS zelf is niet anoniem weergegeven.

## KiM-beoordeling van de algemene aanpak

- ProRail en NS hebben veel aandacht besteed aan een correcte feitenverzameling en verwerking van de data. In grote lijnen is de benchmark op verantwoorde wijze uitgevoerd. De gepresenteerde scores geven een eerste inzicht in hoe de Nederlandse spoorsector presteert in vergelijking met enkele andere landen.
- Conform de concessievoorwaarden en in samenspraak met het ministerie van IenW hebben ProRail en NS gekozen voor een breed scala aan onderwerpen. Vervolgens beperkt de benchmarkrapportage zich tot het weergeven van resultaten en verschillen met de andere landen. Specifieke onderzoeksvragen en diepergaande analyses om tot verklaringen voor geconstateerde verschillen te komen, ontbreken. Het leereffect van de rapportage is daardoor beperkt. Zie ook hoofdstuk 4 van deze notitie.
- De keuze van de *peers* is grotendeels gedaan op pragmatische gronden. Van alle betrokken landen hebben alleen België en Zwitserland een netwerkstructuur die enigszins met die van Nederland is te vergelijken. Geconstateerde verschillen in prestaties kunnen dus vanuit die omstandigheid te verklaren zijn. Dit is op zich niet erg, maar het betekent wel dat terughoudendheid geboden is met ranglijstjes of conclusies over hoe Nederland 'scoort'.
- De omvang van de netwerken wijkt sterk af van NS: alleen SBB heeft een vergelijkbare grootte, de Britse bedrijven zijn veel kleiner.
- Het is verwarrend dat de vervoerders en de infrastructuurbeheerders een andere peer groep hebben, dit temeer daar de resultaten van het ProRail- en NS-deel door elkaar heen staan.
- Het benchmarkproces van het ProRail-deel is voldoende zorgvuldig vormgegeven voor een 'brede' analyse. Door in drie rondes de indicatoren vast te stellen, valt te verwachten dat de aangeleverde gegevens volgens dezelfde definities vastgesteld zijn. In hoofdstuk 3 hebben wij dit nader onderzocht.

- Het NS-deel is minder gestructureerd vormgegeven. De verschillende eigen bewerkingen van de door de *peers* aangeleverde data (nodig om data te harmoniseren) lijken zorgvuldig te zijn uitgevoerd. Maar mede doordat resultaten niet integraal ter validatie zijn voorgelegd aan alle *peers*, is de betrouwbaarheid van de getoonde gegevens mogelijk lager dan bij de door ProRail gevolgde aanpak.
- Zoals afgesproken met de *peers* zijn de gepresenteerde scores van de indicatoren geanonimiseerd. Zo is wél te zien hoe ProRail en NS scoren ten opzichte van hun *peers*, maar het is niet duidelijk wie de anderen zijn.
- In alle ProRail-diagrammen hebben de landen een vaste volgorde, waardoor een anonieme vergelijking tussen de diagrammen mogelijk is. Bij de NS-indicatoren is dat moeilijker, omdat de volgorde van de *peers* per grafiek verschilt. Dit is een gevolg van de geheimhoudingsafspraken tussen NS en de *peers*.

# 3

## Uitvoering van de benchmark: geselecteerde onderwerpen

Het was niet mogelijk om binnen de beschikbare (doorloop)tijd te verifiëren of over alle onderwerpen uit de benchmarkrapportage op verantwoorde wijze gerapporteerd is. Daarvoor is de rapportage te omvangrijk. Het KIM heeft daarom in samenspraak met de directie OVS vier onderwerpen geselecteerd die wij (meer) in detail hebben onderzocht: punctualiteit, vervoerprestatie, veiligheid en financiën. Per onderwerp beschrijven wij wat ProRail en NS hebben gedaan en geven wij daar commentaar op. Aan het eind van het hoofdstuk geven wij een overall-beoordeling over de vier geselecteerde onderwerpen aan de hand van twee thema's: de dataverzameling en de presentatie van de resultaten.

### Punctualiteit

Er zijn twee typen punctualiteit:

1. De treinpunctualiteit: het percentage treinen dat op tijd rijdt. In Nederland wordt hierbij gekeken naar het aantal treinen met maximaal 3 minuten vertraging. Eén van de met ProRail in de beheerconcessie afgesproken normen voor de periode 2015-2019 is dat minimaal 87% van de treinen op tijd moet rijden. Uitgevallen treinen tellen niet mee (Bron: ProRail.nl). Andere indicatoren in de benchmark zijn de uitval van treinen en het aantal 'zwarte dagen' waarin de punctualiteit lager dan 85% is.
2. De reizigerspunctualiteit: dit is het percentage reizigers dat op tijd op zijn bestemming aankomt. Behalve met vertraagde treinen wordt ook rekening gehouden met uitgevallen treinen en (gemiste) overstappen. De met NS en ProRail afgesproken bodemwaarde voor 2017 is dat 89,2 % van de reizigers over het hoofdrailnet (HRN) met maximaal 5 minuten vertraging op zijn bestemming aankomt.

#### *Treinpunctualiteit (ProRail)*

De definitie van het aantal punctuele treinen in de benchmark komt niet overeen met de Nederlandse norm. De data van ProRail presenteren een norm van 5.29 minuten (5 minuten + afronding). Dit is de internationale indicator zoals die bij UIC (*International Union of Railways*, internationale organisatie van spoorwegbedrijven) is afgesproken. ProRail heeft deze norm gekozen, omdat de gegevens eenvoudig door de *peers* aan te leveren zijn.

NS presenteert ook de treinpunctualiteit – maar dan zonder afronding oftewel binnen de 4.59 minuten. De cijfers die NS presenteert hebben betrekking op *vervoerders*, met andere netwerken dan de *infrastructuurbeheerders*. Daarom komen deze data niet altijd overeen met die van ProRail. De richting van de verschillen tussen de jaren komt wel overeen.

In Nederland wordt de treinuitval blijkens de website van ProRail vastgesteld op 48 stations. In de benchmark is een andere definitie gekozen (alle treinen): dit is logisch omdat het anders internationaal niet vergelijkbaar te maken is.

Het aantal zwarte dagen is vastgesteld op het aantal dagen dat de treinpunctualiteit beneden de 85% is. Uitval telt hierbij dus niet mee. Logischerwijs kennen landen die een relatief lage punctualiteit hebben (onder de 90%) een groot aantal zwarte dagen. Uit de benchmarkrapportage blijkt dat de treinpunctualiteit in Nederland hoog is (en dus het aantal ‘zwarte dagen’ laag). Wel is sprake van een hoge uitval van treinen ten opzichte van de *peers*.

Het valt op dat infrastructuurbeheerder E een uitschieter is in de zin dat deze een relatief groot aantal zwarte dagen combineert met een relatief lage treinpunctualiteit. Infrastructuurbeheerder C combineert een hoge treinpunctualiteit met een lage uitval van treinen.

#### *Reizigerspunctualiteit (NS)*

De *peers* hebben gegevens over de reizigerspunctualiteit bij diverse normen aangeleverd aan NS. NS heeft een statistische analyse op haar eigen data uitgevoerd om de relatie te schatten tussen de norm (maximaal aantal minuten vertraging) en de mate waarin de punctualiteit af- of toeneemt bij een verschuiving van de norm (zie ook appendix B van de benchmarkrapportage). Met dit resultaat zijn de aangeleverde data van de *peers* bewerkt om zodoende tot een eenduidige indicator te komen. Hoewel dit op basis van de appendix en de mondelinge toelichting van NS zorgvuldig lijkt te zijn gedaan, betekent een dergelijke analyse per definitie een extra onzekerheid over de hoogte van de indicatoren.

Doordat de reizigerspunctualiteit per land anders gemeten wordt (in Nederland tegenwoordig op basis van chipkaartdata, in andere landen met andere data) is de vergelijkbaarheid waarschijnlijk beperkt. Ook zijn de resultaten maar voor een beperkt aantal *peers* beschikbaar (vier). De verschillen tussen de *peers* zijn bovendien beperkt. Het vergelijken van absolute niveaus is daardoor van weinig waarde. Het onderling vergelijken van de *ontwikkeling* door de jaren heen binnen landen is daarentegen wél zinvol.

## Vervoerprestatie

#### *Treinkilometers (ProRail)*

Het aantal treinkilometers is door de *peers* aangeleverd vanuit hun eigen databestanden (behalve die van Network Rail die afkomstig zijn van de UIC). Hetzelfde geldt voor de tonkilometers; hierbij is het gewicht van treinen, reizigers en vracht bij elkaar opgeteld.

Hier en daar wordt gemeld dat er meerdere inframanagers in het land zijn. De data hebben dan niet zozeer betrekking op het land, maar op de dominante inframanager. Bij Zwitserland gaat het om ongeveer 60% van het totale netwerk, bij de andere landen om het overgrote deel van het netwerk.

#### *Reizigerskilometers (NS)*

Het meten van reizigerskilometers vindt in diverse landen op verschillende wijze plaats (mede afhankelijk van kaartsoorten, abonnementen e.d.). De data worden internationaal bijgehouden, waarbij de definities en meetmethoden zo goed mogelijk vergelijkbaar zijn gemaakt. NS heeft deze data overgenomen. Omdat het openbare data zijn, zijn deze niet geanonimiseerd.

De gekozen definitie voor spits versus niet-spits verschilt per land qua precieze tijdstippen. Uit de mondelinge toelichting is gebleken dat de in de benchmarkrapportage gepresenteerde prestatieverschillen goed zijn te verklaren. Het gegeven dat de spitsstijden per land verschillend zijn gedefinieerd, beïnvloedt echter onvermijdelijk het getoonde beeld, zodat men voorzichtig moet zijn met het trekken van conclusies (zoals het rapport ook aangeeft). Hetzelfde geldt voor de gemiddelde ritafstand en de mate waarin materieel gebruikt wordt (passagierskilometers per treinkilometer).

De reizigerskilometers in een land hangen wél in sterke mate af van specifieke kenmerken van het land en de vervoerstromen. Eenduidige conclusies kunnen hier zonder verdere analyse niet uit getrokken worden. Zo is het bepalend voor de exploitatiemogelijkheden of een spoornetwerk een polycentrische structuur heeft met ongeveer gelijke stromen in beide richtingen van een vervoersrelatie of een structuur met één grote stad met in de spits al het verkeer in één richting. Ook de omvang van de aandelen forensenverkeer, vrijetijdsverkeer en lange-afstandsverkeer zijn structuurkenmerken met impact op de exploitatieresultaten.

Daarnaast hebben - zoals de benchmarkrapportage aangeeft - factoren als investeringen, beleid en economische ontwikkelingen invloed op het aantal reizigerskilometers. De vervoerder kan deze factoren maar beperkt beïnvloeden. Ook kunnen afzonderlijke jaren niet zomaar met elkaar vergeleken worden. Zo wordt het jaar-op-jaar-volume bij NS beïnvloed doordat treindiensten zijn overgegaan naar regionale vervoerders. Dergelijke omstandigheden kunnen ook bij andere *peers* spelen.

## Veiligheid

De veiligheidsindicatoren zijn zowel voor NS als voor ProRail door de *peers* aangeleverd. ProRail heeft naast enkele infrastructuurkenmerken de volgende indicatoren in de benchmark opgenomen:

1. Aantal keer door rood sein gereden per treinkilometer en per 1.000 passeerbewegingen;
2. Aantal zwaar gewonden en doden per treinkilometer;
3. Aantal zwaargewonde en omgekomen werknemers;
4. Aantal suicides per treinkilometer.

Het aantal slachtoffers is dusdanig laag dat incidenten sterk de scores bepalen. Dit betekent dat een kwantitatieve benchmark die jaartotalen vergelijkt zoals in het ProRail-deel gebeurt, minder zinvol is. In het NS-deel wordt gebruik gemaakt van vijfjarige voortschrijdende gemiddelden. Dit vermindert het effect van de incidenten op de scores. Bij dodelijke slachtoffers blijft echter gelden dat incidenten sterk bepalend kunnen zijn voor het ontstane beeld.

Het valt op dat Nederland relatief veel suicides per treinkilometer heeft. In de verschillen tussen landen spelen naast de wijze van registreren waarschijnlijk culturele aspecten (de wijze waarop mensen suicide plegen verschilt sterk per land) een belangrijke rol en niet zozeer (beleids)maatregelen.

De NS-indicatoren hebben betrekking op het aantal dodelijke en zwaargewonde slachtoffers onder treinpassagiers. Van slechts twee *peers* zijn deze data beschikbaar. Opvallend is dat NS een gunstige score heeft voor wat betreft het aantal dodelijke slachtoffers, maar niet wat betreft het aantal gewonden. Doordat het (zeker bij dodelijke slachtoffers) om incidenten gaat, is het niet verantwoord om harde conclusies te trekken. Het aantal gewonden hangt sterk af van de registratiemethode en gebruikte definities.

Ook hier geldt dat de data op zorgvuldige wijze verzameld zijn. Enkele gegevens roepen vragen op, maar zonder verdieping en verdere verklaring is moeilijk aan te geven of en waarom ProRail en NS relatief goed scoren op veiligheid (afgezien van gewonde passagiers) en relatief slecht op suicides.

## Financiën

De financiële gegevens zijn aangeleverd door de *peers*. De verdeling van verantwoordelijkheden tussen overheden, vervoerders en infrastructuurmanagers verschilt sterk per land. Dit heeft logischerwijs invloed op de financiële boekhouding.

Per kilometer spoor zijn de onderhoudskosten in Nederland bovengemiddeld en de investeringen ondergemiddeld. Mogelijk hangt dit met elkaar samen. Dit is bij uitstek een onderwerp waarover zonder verdere verdieping geen conclusies zijn te trekken. Zo kunnen bijvoorbeeld definitieverschillen van onderhoud en vernieuwing een rol spelen. Zoals de benchmarkrapportage ook aangeeft geldt dit ook voor toevallige pieken in de tijd in grote uitgaven voor bijvoorbeeld vervanging van kunstwerken.

In kosten per treinkilometer scoort ProRail goed. Dit is conform verwachting gezien het intensieve gebruik van het netwerk. Zonder verdere analyses zijn geen conclusies over de kosten te trekken.

Bij NS duiden lage opbrengsten per kilometer in combinatie met beperkte winstmarges en het ontbreken van exploitatiesubsidies eveneens op lage kosten. De structuur van het netwerk (polycentrisch) en verdeling van treindiensten over het totale netwerk is hierbij van belang. Op de meeste regionale lijnen in Nederland rijden namelijk andere vervoerders dan NS. Deze lijnen zijn wél gesubsidieerd.

Nadere analyses zijn nodig om hier eenduidige conclusies over te trekken. Een interessante vraag is bijvoorbeeld wat de invloed is van marketing en kortingen op het treingebruik in daluren en op reizen op de lange afstand. Zo ook wat de invloed is van een sterke concentratie van forensenstromen naar economische kerngebieden, met als gevolg een onevenwichtige bezetting van treinen.

## Overall-beoordeling van geselecteerde onderwerpen: de dataverzameling

### ProRail

- De indicatoren zijn veelal vastgesteld aan de hand van internationaal geldende definities, en ook uitgevraagd onder verwijzing naar dergelijke standaarden. Een groot aantal indicatoren wordt daardoor standaard door de *peers* bijgehouden.
- Diverse andere indicatoren zijn specifiek voor deze studie gedefinieerd. In de uitvraag zijn deze goed en eenduidig geformuleerd. Vervolgens is met de *peers* informatie uitgewisseld of een nadere toelichting gegeven. In de meeste gevallen zijn de data door de *peers* aangeleverd, in sommige gevallen wordt daarbij aangegeven dat andere definities zijn gebruikt, die in beperkte mate afwijken. Soms betreft het een schatting van de peer. Dit is gezien de diepgang van de benchmark acceptabel aangezien de *peers* dit het best kunnen bepalen.
- Alle gepresenteerde basisindicatoren zijn aangeleverd door de *peers*. Er heeft dus geen bewerking achteraf plaatsgevonden, afgezien bijvoorbeeld een deling door het aantal kilometers spoor. Dergelijke eenvoudige bewerkingen beïnvloeden de kwaliteit niet.
- De resultaten zijn aan de *peers* voorgelegd en de op- en aanmerkingen zijn correct verwerkt.

### NS

- Veel indicatoren die bij de benchmark voor de vervoerders zijn gebruikt, zijn gebaseerd op enquêtes of andere berekeningen. Dit geldt bijvoorbeeld voor de reizigerspunctualiteit. We zien dit overigens ook terug bij de klanttevredenheid. De methode (bijvoorbeeld interviews, online, wijze van berekenen) is erg bepalend voor de uitkomsten. De cijfers zijn daardoor moeilijk vergelijkbaar. NS vermeldt dit overigens regelmatig in de rapportage. Wij zien geen eenvoudige mogelijkheden om dit beter te doen, maar het betekent wel dat er geen harde conclusies verbonden mogen worden aan de absolute hoogte van deze indicatoren. Voorzover de methode per vervoerder hetzelfde is gebleven, kunnen wel conclusies getrokken worden uit de ontwikkeling van de indicator in de loop van de tijd.

## Overall-beoordeling van geselecteerde onderwerpen: de presentatie

- Nadat de te benchmarken aspecten gekozen waren, hebben ProRail en NS zelf de indicatoren vastgesteld. Dit leidt in sommige gevallen tot dubbelingen, zoals bij de treinpunctualiteit, veiligheid en enkele financiële indicatoren. Voor treinpunctualiteit hanteert ProRail een norm van 5.29 minuten, NS van 4.59 minuten. Ook enkele veiligheidsindicatoren zijn verschillend gedefinieerd. Tevens is het type conclusies dat getrokken wordt verschillend. Een en ander is feitelijk juist, maar oogt verwarrend.
- In sommige gevallen is duidelijk sprake van een uitschieter (één sterk afwijkend resultaat). Voorbeelden zijn het aantal incidenten waar personeel bij betrokken is (bij land 6 zeer hoog) en het aantal vertragingen veroorzaakt door de infrastructuurbeheerder, waar land D opvallend slecht scoort. Dergelijke uitschieters hebben een grote invloed op het gemiddelde. ProRail en NS presenteren vaak

ook gemiddelden zonder de uitschieter. Omdat een toelichting veelal ontbreekt, rijst de vraag of de peer ook na de nodige checks wel de juiste vergelijkbare data aangeleverd heeft. NS merkt af en toe op dat een definitie is veranderd of dat een incident de mogelijke oorzaak is.

- In relevante gevallen worden twee variabelen tegen elkaar afgezet, waarbij de keuze voor deze combinatie niet wordt verantwoord. Zo wordt de klanttevredenheid wel gerelateerd aan de punctualiteit, maar niet aan de uitval van treinen.
- Bij diverse figuren wordt benadrukt dat de vergelijkbaarheid beperkt is door methodologische problemen of uitschieters, zoals bij het aantal gewonden per reizigerskilometer. Hier wordt bovendien aangegeven dat de slechte score van NS waarschijnlijk komt door definitieverschillen. Dit zal echter ook gelden voor de figuur daarboven waar NS juist goed scoort.
- Nogal wat NS-indicatoren zijn als een index gepresenteerd, waardoor het absolute niveau niet bekend is. Dit kan de presentatie sterk beïnvloeden. Als bijvoorbeeld het absolute niveau laag is, worden verschillen al gauw uitvergroot.
- De conclusies bij ProRail gaan meestal over de score van Nederland ten opzichte van andere landen. Onduidelijk is waarom bepaalde scores wél en andere niet worden uitgelicht.
- Bij NS zijn diverse indicatoren zoals klanttevredenheid, reizigerspunctualiteit en reizigerskilometers minder goed vergelijkbaar te maken doordat deze elders anders gemeten en gedefinieerd zijn.



# 4

## Bruikbaarheid van de resultaten

Benchmarken is zeker bij een internationale vergelijking niet eenvoudig. Het is lastig om vergelijkbare *peers* te vinden die bereid zijn mee te werken. Iedere organisatie hanteert eigen definities en methoden van dataverzameling. ProRail en NS hebben een grote hoeveelheid informatie verzameld en daar veel werk voor verzet.

### **Benchmarken is een proces: de volgende stap is ‘verklaren’**

De literatuur omtrent benchmarking benadrukt dat benchmarking nuttig is als er een continu proces van gemaakt wordt (Išoraitė, 2004). Na het vaststellen van het doel, het vinden van partners en het verzamelen van data komt de analysefase. Daarop volgt een aanpassingsfase, waarbij de lessen geïmplementeerd worden. Hierna volgt een volgende ronde van benchmarking, om te kunnen nagaan of de aanpassingen gewerkt hebben en of er weer nieuwe lessen te leren zijn. Dit proces is vergelijkbaar met de PDCA cyclus (Plan-Do-Check-Act) die veelal in beleidsprocessen wordt toegepast.

De huidige benchmarkrapportage heeft de eerste fasen doorlopen en geeft een eerste analyse. In overeenstemming met de concessies aan NS en ProRail gaat deze analyse niet verder dan het presenteren van de data. Het verklaren van verschillen gebeurt bijna niet. Om echt lessen te kunnen trekken uit een benchmark is een zoektocht naar verklaringen voor geconstateerde prestatieverschillen een noodzakelijke volgende stap. ProRail en NS zijn zich blijkens de gesprekken met het KiM terdege van deze situatie bewust en hebben de ambitie hun analyses verder te verdiepen.

In theorie zijn vier typen verklaringen te geven voor geconstateerde prestatieverschillen:

1. Verschillen in definities, of wijze van dataverzameling: binnen de beschikbare tijd en middelen zijn de data zoveel mogelijk eenduidig gemaakt. Desalniettemin is niet te voorkomen dat als er nader geanalyseerd wordt, blijkt dat dit nog niet geheel het geval is.
2. Structurele (niet-beïnvloedbare) factoren: denk aan specifieke kenmerken van het land (zoals bevolkingsdichtheid, geografische kenmerken). Deze zijn eigenlijk niet beïnvloedbaar door de sector en/of het beleid.
3. Lange termijn keuzes: dit gaat om zaken als beschikbaar materieel, keuzes met betrekking tot de infrastructuur (zoals bovenleidingspanning, beveiligingssysteem) en de marktordening van de sector. Deze kunnen alleen op lange termijn aangepast worden. Hiervoor zijn strategische keuzes en/of lange termijn investeringen nodig.
4. Te beïnvloeden factoren: dit zijn factoren die de organisatie zelf of beleidsmakers (binnen afzienbare termijn) kunnen aanpassen.

Voor zowel NS en ProRail als het ministerie van IenW zijn de derde en vierde categorie uiteindelijk het meest interessant: op basis daarvan kan actie ondernomen worden en/of gericht beleid gedefinieerd worden om de prestaties van de Nederlandse spoorsector te verbeteren. Daarvoor is het uiteraard nodig om ook de eerste twee categorieën mee te nemen in de verklaring.

### Voorzichtig zijn met trekken van inhoudelijke conclusies

De gepresenteerde scores geven een eerste inzicht in hoe de Nederlandse spoorsector presteert op de desbetreffende indicator, in vergelijking met een aantal andere landen en vervoerbedrijven. Deze conclusie kan niet worden veralgemeniseerd naar de positie van Nederland in bijvoorbeeld Europa of in de westerse wereld om de volgende redenen:

- De selectie van landen bepaalt mede hoe Nederland scoort. Frankrijk, Denemarken, het Verenigd Koninkrijk en Zweden zijn niet zo goed te vergelijken met Nederland vanwege verschillen in omvang, ruimtelijke structuur, structuur van het spoornet en/of de vervoerspatronen (zoals de verdeling tussen lange en korte reisafstanden). Vergelijking met deze landen is daarom alleen van nut voorzover structuurkenmerken geen rol spelen.
- Keuzes omtrent indicatoren en de operationalisering daarvan kunnen in grote mate de conclusie beïnvloeden.
- Uitschieters in de data waarvan de reden niet bekend is beïnvloeden in sommige gevallen sterk de resultaten.

### Aanbevelingen aan ProRail en NS voor toekomstige benchmarks

- Voorkom de verwarring die kan ontstaan doordat de analyses van NS en ProRail in één rapportage zijn samengevoegd. Dit speelt in het bijzonder bij onderwerpen waarover beide partijen rapporteren maar waarbij zij verschillende definities hebben gehanteerd.
- Idealiter heeft de benchmark voor vervoerders en infraproviders dezelfde vergelijkingslanden. De vergelijkbaarheid wordt hierdoor vergroot, bijvoorbeeld omdat kruisvergelijkingen van de output van infraproviders met die van vervoerders mogelijk worden. Dat vergroot ook het leereffect.
- Een andere optie om vergelijkbaarheid te vergroten is om niet te kijken naar hele landen maar naar landsdelen. Zo lijken het Duitse Noordrijn-Westfalen en het Japanse eiland Kyushu qua inwonertal, ruimtelijke structuur, verstedelijkingsgraad en netwerkstructuur veel meer op Nederland dan menig ander, compleet land. Wij realiseren ons dat dit wel de nodige dataproblemen met zich meebrengt.
- De logische volgende stap is om meer de diepte in te gaan. Het is waarschijnlijk te arbeidsintensief en ook niet zinvol dit te doen voor alle onderwerpen. Kies dan een beperkt aantal onderwerpen vanuit actuele vraagstukken en werk die verder uit.
- In veel diagrammen van ProRail zit een vergelijkingspartner (land C) die zich in ongeveer dezelfde positie bevindt, maar vaak net iets beter scoort. Het kan interessant zijn om in een vervolgfase verder op deze partner te focussen en in verdiepende gesprekken te verkennen wat deze anders/beter doet.

### Aanbevelingen aan OVS voor toekomstige benchmarks

- Verdiep in de toekomst de opdracht aan ProRail en NS door niet langer uitsluitend primaire prestatiegegevens te vragen, maar ook verklaringen voor gevonden verschillen die rekening houden met achtergronden en omstandigheden in te vergelijken landen. Dit vergroot de potentiële leereffecten aanzienlijk.
- Om verwarring te voorkomen bevelen wij aan om het rapportagedeel van ProRail en dat van NS beter op elkaar af te laten stemmen voor zover zij over dezelfde indicatoren rapporteren. Lukt dit niet, dan is het een optie om terug te vallen op twee afzonderlijke benchmarkrapportages.
- Kies te onderzoeken onderwerpen die beleidsrelevant zijn en ook beleidsmatig beïnvloed kunnen worden. Dit kan gaan om onderwerpen als financiële sturing, punctualiteit en veiligheid. Laat het benchmarken van de meer operationele aspecten aan ProRail en NS zelf. Op deze manier krijgt de analyse meer focus en kan meer tijd besteed worden aan het verklaren van verschillen en de leereffecten.

# Literatuur

Išoraite, M. (2004). *Benchmarking Methodology in A Transport Sector*, *Transport*, Vol. XIX-6, blz. 269-275.

Ministerie van IenM (2013). *Lange Termijn Spooragenda; Visie, Ambities, Doelen*.  
Den Haag: Ministerie van Infrastructuur en Milieu

Ministerie van IenM (2014a). *Beheerconcessie ProRail 2015-2015*.  
Den Haag: Ministerie van Infrastructuur en Milieu

Ministerie van IenM (2014b). *Concessie voor het hoofdrailnet 2015-2015*.  
Den Haag: Ministerie van Infrastructuur en Milieu

ProRail en NS, 2017, *International Benchmark 2011-2015*, Utrecht.

## Colofon

Dit is een uitgave van het Ministerie van Infrastructuur en Waterstaat  
November 2017

Kennisinstituut voor Mobiliteitsbeleid (KiM)

### Auteurs

Peter Bakker  
Fons Savelberg  
Sytze Rienstra

### Vormgeving en opmaak

VormVijf, Den Haag

Kennisinstituut voor Mobiliteitsbeleid (KiM)

Postbus 20901  
2500 EX Den Haag  
Telefoon: 070 456 19 65  
Fax: 070 456 75 76

Website: [www.kimnet.nl](http://www.kimnet.nl)

E-mail: [info@kimnet.nl](mailto:info@kimnet.nl)

Publicaties van het KiM zijn als PDF te downloaden van onze website [www.kimnet.nl](http://www.kimnet.nl).  
U kunt natuurlijk ook altijd contact opnemen met één van onze medewerkers.

*Delen uit deze publicatie mogen worden overgenomen onder vermelding van het KiM als bron.*



Het Kennisinstituut voor Mobiliteitsbeleid (KiM) maakt analyses van mobiliteit die doorwerken in het beleid. Als zelfstandig instituut binnen het ministerie van Infrastructuur en Waterstaat (IenW) maakt het KiM strategische verkenningen en beleidsanalyses. De inhoud van de publicaties van het KiM behoeft niet het standpunt van de minister en/ of de staatssecretaris van IenW weer te geven.



Kennisinstituut voor Mobiliteitsbeleid

Dit is een uitgave van het

**Ministerie van Infrastructuur en Waterstaat**

Postbus 20901 | 2500 EX Den Haag  
[www.rijksoverheid.nl/ienw](http://www.rijksoverheid.nl/ienw)

[www.kimnet.nl](http://www.kimnet.nl)

November 2017