



UvA-DARE (Digital Academic Repository)

Benchmarking in research evaluations

we can do without it

van Vree, F.; van Leeuwen, T.

DOI

[10.59350/h4f1k-00y10](https://doi.org/10.59350/h4f1k-00y10)

[10.59350/crbvf-qkw52](https://doi.org/10.59350/crbvf-qkw52)

Publication date

2025

Document Version

Final published version

License

CC BY

[Link to publication](#)

Citation for published version (APA):

van Vree, F., & van Leeuwen, T. (2025). Benchmarking in research evaluations: we can do without it. Web publication or website, Centre for Science and Technology Studies (CWTS). <https://doi.org/10.59350/h4f1k-00y10>, <https://doi.org/10.59350/crbvf-qkw52>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Benchmarking bij onderzoeks-evaluaties: we kunnen zonder

Frank van Vree and Thed van Leeuwen

Published 30 januari 2025

Citation

van Vree, F., & van Leeuwen, T. (2025, January 30). Benchmarking bij onderzoeks-evaluaties: we kunnen zonder. *Leiden madtrics*. <https://doi.org/10.59350/crbvf-qkw52>

Copyright

Copyright © Frank van Vree et al. 2025. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Leiden Madtrics

([To read the English version, click here](#))

Bij iedere evaluatie van wetenschappelijk onderzoek ligt de vraag voor de hand: hoe presteert een onderzoekseenheid in vergelijking met anderen, nationaal en internationaal? Die vraag is makkelijker gesteld dan beantwoord, want een antwoord laat zich niet gemakkelijk onderbouwen. In het [Strategy Evaluation Protocol \(SEP\) 2021-2027](#), dat richtlijnen geeft voor de verplichte periodieke evaluatie van het onderzoek van Nederlandse universiteiten en onderzoeksinstituten van de KNAW en het nationale onderzoeksfonds NWO, wordt weliswaar gewezen op *benchmarking* als een mogelijke manier om robuuste data te genereren, maar daarbij wordt niet uitgelegd wat precies onder deze techniek/methodologie moet worden verstaan en hoe die precies is te operationaliseren. En dat is niet zonder gevaar, want er kleven nogal wat bezwaren tegen *benchmarking* in de strikte betekenis van *een op kwantitatieve criteria (tijd, kosten, opbrengsten) gebaseerde vergelijking van prestaties*. Met het oog op de lopende voorbereiding van het nieuwe SEP voor de periode 2027-2033 is het daarom verstandig *benchmarking* als methode kritisch tegen het licht te houden, zonder de idee van vergelijking als zodanig volledig in de ban te willen doen.

Een problematisch instrument

Op pagina 19 van het SEP wordt *benchmarking* geïntroduceerd als een mogelijke manier om robuuste data te genereren: 'Other sources of robust data may include benchmarking against peer research units [...]'. Anders gezegd: naast de door de eenheid geselecteerde indicatoren en casestudies om de kwaliteit en impact van het onderzoek te tonen, kan *benchmarking* worden ingezet ter onderbouwing van de narratieve argumenten in de zelfevaluatie. Wat precies onder *benchmarking* moet worden verstaan, wordt evenwel niet duidelijk, maar gesuggereerd wordt dat *benchmarking* en kwantitatieve indicatoren, in een soort mengvorm, een doeltreffende onderbouwing kunnen bieden.

Die aanname is, op z'n zachtst gezegd, problematisch. Het begint er al mee dat hier een techniek wordt voorgesteld zonder aan te geven hoe deze te operationaliseren. Dat probleem wordt knellender wanneer wordt uitgegaan van de gangbare definitie voor *benchmarking*: het vergelijken van bedrijfsprocessen en prestatiestatistieken van een instelling of bedrijf met *best practices* van andere bedrijven en instellingen in dezelfde of vergelijkbare branche, waarbij doorgaans gemeten wordt op de dimensies kwaliteit, tijd en kosten. De zuiver kwantitatieve criteria waarop een dergelijke vergelijking van prestaties doorgaans is gestoeld, staan evenwel op gespannen voet met de geest en het doel van het SEP. Daarin ligt de nadruk immers op een toetsing van de kwaliteit en impact vanuit het perspectief van de eigen missie en de daaraan verbonden strategie, waarvoor het gebruik van ogenschijnlijk 'objectieve' metrische criteria (JIF, [h-index](#)) als ongeschikt worden gekwalificeerd.

Maar dat is zeker niet het enige probleem. Het ontbreken van een heldere definiëring en operationalisering wekt de suggestie dat een dergelijke operatie op grond van kwantitatieve gegevens relatief makkelijk is, terwijl zij in feite zowel op inhoudelijke als ethische gronden buitengewoon problematisch en derhalve af te raden is.

Ongelijke data

Het ontbreken van een heldere definiëring en operationalisering voor benchmarking in het kader van onderzoeksevaluaties impliceert dat er ook geen criteria zijn om te bepalen waaraan het vergelijkingsmateriaal aan moet voldoen. Daarmee wordt de deur opengezet voor toeval, willekeur en opportunisme. Dat geldt, om te beginnen, voor de keuze van de onderzoekseenheden die in de benchmarking worden meegenomen. Een weinig ambitieuze keuze voor benchmark-eenheden kan een goede uitkomst opleveren in het evaluatieproces, maar hoeft niet per se de beste uitkomst te zijn op de langere termijn; omgekeerd kan een te ambitieuze keuze leiden tot een negatieve uitkomst waar dit niet nodig had hoeven zijn.

Minstens zo problematisch bij kwantitatieve benchmarking is de ongelijkheid in de beschikbaarheid en de aard van de onderliggende data. Ten aanzien van de eigen onderzoekseenheid heeft men vaak wel een goed zicht op het verzamelde materiaal, afkomstig uit lokale informatiesystemen (zoals [Pure](#) of [Metis](#), maar ook [Converis](#)) en opgesteld volgens bekende criteria, maar die informatie heeft men zelden als het gaat om de data van de eenheden die ter vergelijking in de benchmark figureren. Dat geldt a fortiori voor instellingen in het buitenland, waar andere systemen en normen dominant zijn. Het gebrek aan inzicht in de wijze waarop data zijn verzameld en welke waarde ze vertegenwoordigen doet ernstig afbreuk aan de validiteit van de uiteindelijke vergelijking, vaak zonder dat men zich hiervan bewust is.

Bij een vergelijking op basis van kwantitatieve indicatoren wordt ervan uit gegaan dat deze indicatoren de eenheden op een gelijke wijze representeren. Dat is evenwel maar in zeer beperkte mate het geval. Dit probleem laat zich treffend illustreren in de praktijk van veldnormalisatie, die in bibliometrische studies wordt ingezet om de verschillen in referentiegedrag in uiteenlopende publicatie- en referentieculturen per vakgebied te compenseren. Zolang het gaat om het vergelijken van twee verwante specialismen, zoals cardiologie en oncologie, is een dergelijke normalisering zinvol om tot een vergelijking te komen. Maar dat wordt anders wanneer in een studie uiteenlopende vakgebieden - of zelfs hele universiteiten - worden geanalyseerd en vergeleken. Theoretisch gezien zou men de aantallen citaties na veldnormalisatie kunnen vergelijken, maar daarbij wordt dan wel voorbijgegaan aan de soms enorme verschillen in publicatiecultuur. Voor natuurkundigen – bijvoorbeeld - zou je kunnen vaststellen dat tussen de 80-85% van alle publicaties wel in internationale tijdschriften zijn gepubliceerd, waarmee men dus een redelijk goed beeld heeft van de output van die eenheid. Voor veel domeinen binnen de geesteswetenschappen, zoals de historische en literaire wetenschappen, liggen de verhoudingen evenwel heel anders, waarbij een substantieel deel van de resultaten worden gepubliceerd in de vorm van boeken en boekhoofdstukken, dikwijls ook in andere talen dan het Engels die niet in de systemen zoals Scopus en Web of Science voorkomen. Dergelijke verschillen bestaan zelfs binnen vakgebieden, waarbij meer toepassingsgerichte specialismen heel andere communicatiekanalen inzetten. Kortom, het 'vergelijkbaar' maken van kwantitatieve gegevens kunnen allicht leiden tot een eenzijdig en vertekend beeld van de kwaliteit en impact van onderzoekseenheden en de door hen voorgestane strategische keuzes.

Ethische bezwaren

Een laatste punt betreft de ethiek van dit proces. De data van de eigen eenheid zijn – als het goed is – gecontroleerd en gevalideerd als valide materiaal voor evaluatiedoeleinden. Maar dat geldt niet voor het materiaal dat gebruikt wordt voor de benchmark-eenheden: die kunnen op een andere manier of met een heel ander doel zijn verzameld. Kwaliteitscontroles op dat materiaal zou om die reden een eerste vereiste zijn, maar dat vraagt uiteraard om toestemming van de eigenaar van deze data. Los daarvan is het gebruik van data zonder toestemming ethisch onverantwoord, vooral in het licht van het gegeven dat in Nederland de resultaten van de evaluatie openbaar gemaakt dienen te worden, waarmee de reputatie en het imago van de benchmark-eenheden zouden kunnen worden beschadigd.

Tot slot

Op grond van het voorgaande kan niet anders dan worden geconcludeerd dat benchmarking in de betekenis van *een op kwantitatieve criteria (tijd, kosten, opbrengsten) gebaseerde vergelijking van prestaties* op zowel theoretische als ethische gronden in de meeste gevallen ongeschikt is voor evaluaties in de zin van het SEP. Dat wil evenwel niet zeggen dat een vergelijking, of, zo men wil: *benchmarking* in een minder strikte betekenis van de term, zinvol kan zijn. De eigen positie, missie en strategie zouden daarbij als vertrekpunt moeten fungeren, als een manier om de onderzoekseenheid, vanuit een meer kwalitatieve perspectief, scherper neer te zetten. Zo'n spiegel, bijvoorbeeld in de vorm van een casestudy, kan buitengewoon zinvol.

Header image: Estúdio Bloom on [Unsplash](#).

DOI: 10.59350/crbvf-qkw52 ([export/download/cite this blog post](#))