UNIVERSITEIT LEIDEN
FACULTEIT DER WISKUNDE EN NATUURWETENSCHAPPEN

Mathematisch Instituut

27 januari 2026

[redacted]

[redacted]

Geachte Raadsleden en wethouders sociaal domein,

Ik ben [redacted] een Engels-Nederlands wiskundige en statisticus, verbonden aan de Universiteit Leiden, faculteit der wiskunde en natuurwetenschappen, nu als Professor emeritus. Sinds het begin van de eenentwintigste eeuw verdiep ik mij in statistische argumentaties die bij rechtszaken gebruikt worden. Daarmee ben ik begonnen in 2006 na de spraakmakende veroordeling in de zaak-Lucia de Berk, die mede dankzij mijn onderzoek en bevindingen is vrijgesproken. Ook na die vrijspraak verdiep ik mij in statistische argumentaties, in rechtszaken maar ook daarbuiten.

Ik doe u deze bevindingen toekomen, omdat ik meen dat de audit niet voldoet aan de minimale eisen die kunnen worden gesteld aan een steekproefsgewijs onderzoek dat, zoals hier, leidt tot (her)certificering van een organisatie die ernstig kan ingrijpen in het gezinsleven.

U zult hierna lezen dat ik tot de slotsom kom dat de audit, gelet op de wijze waarop het onderzoek is uitgevoerd en de inhoud van het rapport, leidt tot conclusies die niet kunnen worden gedragen door de onderzoeksbevindingen, als wordt uitgegaan van de minimale standaarden die gelden bij steekproefsgewijs onderzoek. Daardoor komt de vraag op of wel op gronden die dat kunnen dragen is beslist tot (her)certificering, met alle veiligheidsrisico's voor ouders en kinderen die daarvan het gevolg kunnen zijn. Ik noem mijn belangrijkste bevindingen.

Hoeveel dossiers voor een representatief beeld?

Wat "representatief" is, hangt af van de gewenste betrouwbaarheid, de toegestane foutmarge in het geschatte foutpercentage (bijv. ± 5 procent), en de verwachte foutkans per dossier.

Leiden University, Mathematical Institute, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Einsteinweg 55, 2333 CA Leiden - Tel.: +31-(0)71-5277111

In audits en kwaliteitscontroles zie je vaak dat bij een onbekende foutkans, wordt uitgegaan van een conservatief scenario, en dat steekproeven van grofweg enkele honderd of meer dossiers worden getrokken bij populaties van enkele honderden tot duizenden dossiers; exact berekenen gebeurt normaal met statistische formules en tabellen. Tegenwoordig kan ChatGPT en dergelijke AI systemen de taak volledig overnemen mits men de juiste vraag weet te stellen. Hier is toch wel professioneel statistisch kennis voor nodig.

Een steekproef mag best heel klein zijn. Maar hoe kleiner, hoe minder men komt te weten over de populatie.

Een rekenvoorbeeld en uitwerking daarvan.

Laat ik bij wijze van voorbeeld uitgaan van een populatie van 950 clientendossiers die de GI behandelt in een bepaalde periode en een aselecte steekproef, zonder teruglegging, van 11 clientendossiers,. Het aantal 11 is vermeld op bladzijde 7 van het rapport van het Keurmerkinstituut. De populatieomvang staat echter niet in het rapport; het getal 950 is een raming gebaseerd op het rapport "Stand van zaken herstelplan Jeugdbescherming Noord", 10 september 2025. Daar wordt gesproken over een totaal van 944 "maatregelen (jeugdreclassering en ots)". Het rapport is overigens voornamelijk toegespitst op aantallen werknemers.

Je schat dan een percentage in met een zeer grote onzekerheidsmarge.

In de praktijk zou men voor een representatief beeld meestal een substantieel groter aantal kiezen, afhankelijk van de eisen aan betrouwbaarheid en nauwkeurigheid.

Wat is acceptabel?

Stel een onderzoek vindt 4 tekortkomingen in 11 dossiers.

In de steekproef zijn 4 van de 11 dossiers niet goed: dat is ongeveer $4/11 \approx 36\%$.

Dus de (puntschatting) op basis van deze steekproef is dat er in een totale populatie van 950 dossiers heel grofweg 350 dossiers niet zullen voldoen.

Belangrijk:

Dit is geen zekerheid maar een schatting met aanzienlijke onzekerheid omdat de steekproef heel klein is, en bovendien is de methodiek van selectie niet vermeld.

Ik berekende een zogenaamde betrouwbaarheidsinterval. Het resultaat is dat met een zekerheid van 95%, het foutpercentage tussen 11.05% en 69.05% ligt. In ronde getallen dus, tussen 10% en 70%. Deze constatering veronderstelt wel een aselecte steekproef; de eenvoudigste manier om representativiteit te garanderen.

Leiden University, Mathematical Institute, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Einsteinweg 55, 2333 CA Leiden - Tel.: +31-(0)71-5277111

Kort samengevat: 4 fouten in 11 dossiers (bij een professioneel uitgevoerd en geëvalueerd aselecte steekproef) wijst op een mogelijk heel hoog foutpercentage in de populatie. Tussen 1 op de 10 en 7 op de 10 (!!) van de dossiers kunnen wellicht dezelfde tekortkomingen vertonen die in de wel geselecteerde dossiers zijn vastgesteld. En wat is de garantie dat dit de ernstigste tekortkomingen zijn in het gehele populatie van dossiers?

Zorgen over het onderzoek

Het is voor mij niet te begrijpen dat is gekozen voor een steekproefgewijs onderzoek, maar slechts aan dat onderzoek uitvoering is gegeven door enige bevindingen samen te vatten zonder dat die zijn uitgewerkt op de wijze zoals hoort en ik hiervoor heb gedaan. Daardoor ontstaat een onjuist beeld over de kans dat in grote aantallen zaken sprake is van de ernstige tekortkomingen die in de wel onderzochte dossiers zijn gebleken.

Ik wil daarbij ook nog aantekenen dat de dossiers die onderzocht zijn zijn geselecteerd door de te certificeren instelling waarbij volstrekt onduidelijk is hoe de steekproef vorm is gegeven en op welke gronden deze dossiers zijn geselecteerd.

Als deskundige kom ik tot de slotsom dat de wijze waarop het onderzoek is uitgevoerd de lichtste toets der kritiek niet kan doorstaan en dat de uitwerking van het onderzoek in de audit vragen oproept, zowel ten aanzien van de dossier selectie als ten aanzien van de keuze voor het geringe aantal en het ontbreken van een statistische uitwerking van de onderzoeksbevindingen.



Leiden University, Mathematical Institute, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Einsteinweg 55, 2333 CA Leiden - Tel.: +31-(0)71-5277111