
Methodiek voor correspondentietesten bij lokale besturen

29 APRIL 2022

Bert Leysen, Prof. dr. Pieter-Paul Verhaeghe
Vakgroep Sociologie - Vrije Universiteit Brussel



Inhoudstafel

1. Introductie	3
2. Doel van de opdracht	4
3. Kadering	6
3.1 Definities van discriminatie	6
3.2 Instrumenten voor nulmeting	7
4. Methodiek correspondentietesten	13
4.1 Testen bij publieke en private sector	13
4.2 Operationalisatie van de discriminatiegronden	16
4.3 Steekproefgrootte	22
4.4 Ongepaarde, duo- of triotesten	24
5. Stappenplan	26
Fase 1: Voorbereiding	26
Fase 2: Uitvoering	39
Fase 3: Analyse	44
6. Sensibilisering	48
7. Samengevat	52
Referenties	53

1. Introductie

Het meten van **discriminatie op de arbeidsmarkt** is niet voor de hand liggend. Traditionele methoden zoals interviews, enquêtes, of statistieken van officiële klachten hebben telkens hun beperkingen. Discriminatie is niet enkel bij wet verboden, maar er heerst ook een sterke sociale norm dat discriminatie verkeerd is. Hierdoor gaan respondenten vaak sociaal wenselijk antwoorden bij vragen over hun mogelijk discriminerende gedrag. Het gevolg is dat interviews of enquêtes doorgaans tot een serieuze onderschatting van discriminatie leiden. Langs de kant van de slachtoffers van discriminatie is het ook niet makkelijk om een accuraat beeld te krijgen. Indien men bijvoorbeeld niet wordt uitgenodigd voor de volgende ronde bij een sollicitatie, tasten kandidaten vaak in het duister of dit komt door discriminatie dan wel dat andere kandidaten een beter profiel voor de job hebben. Enerzijds kunnen kandidaten denken dat ze niet worden uitgenodigd omwille van discriminatie, terwijl dit niet het geval is. Anderzijds kunnen ze denken dat ze omwille van professionele redenen niet worden uitgenodigd terwijl er eigenlijk wél sprake is van discriminatie. We hebben met andere woorden nood aan een objectief meetinstrument.

Correspondentietesten worden door onderzoekers naar voor geschoven als “de gouden standaard” voor het objectief meten van discriminerend gedrag. Bij dit soort testen solliciteren paren van fictieve kandidaten via e-mail voor vacatures. De kandidaten hebben kwalitatief gelijkaardige profielen, gaande van gelijkaardige opleidingen aan instituten met gelijkaardige reputaties, tot gelijklopende carrière trajecten, vaardigheden, en woonplaatsen. Enkel de **discriminatiegrond**, namelijk het te testen kenmerk (bv. etnische afkomst, geslacht, leeftijd, etc.) verschilt tussen de kandidaten. Nadien worden de verschillen in reacties van de werkgevers tussen de testgroep en de controlegroep bijgehouden. Significante ongelijkheid tussen groepen wordt vervolgens toegewezen aan discriminatie.

In dit rapport wordt dieper ingegaan op de verschillende aspecten van correspondentietesten, specifiek bij het opzetten van **nulmetingen naar discriminatie bij lokale besturen als werkgever**. Na een korte toelichting bij het doel van de opdracht wordt er stilgestaan bij de bredere kadering waar het concept discriminatie en correspondentietesten centraal staan. Vervolgens wordt de methodiek uitgediept en het stappenplan overlopen voor de voorbereiding, uitvoering, en analyse van de correspondentietesten. Ter afsluiting staan we stil bij hoe de testen sensibiliserend ingezet kunnen worden.

2. Doel van de opdracht

De Vrije Universiteit Brussel (VUB) heeft de opdracht gekregen van de Dienst Diversiteitsbeleid van de Vlaamse overheid (VO) om een **methodiek** te ontwikkelen voor **correspondentietesten om discriminatie te onderzoeken bij de eerste fase van aanwerving door de lokale overheid als werkgever**.

De methodiek kwam tot stand na een uitgebreide consultatie van medewerkers rekrutering, selectie en personeelszaken, beleidsmedewerkers gelijke kansen en diversiteit, en diensthoofden van de lokale besturen Antwerpen, Genk, Gent, Kortrijk, Leuven, en Mechelen. Ook is de inbreng van de Vereniging van Vlaamse Steden en Gemeenten (VVSG), Diverscity en de Vlaamse overheid opgenomen in dit rapport. De specifieke moeilijkheden waarmee lokale besturen worden geconfronteerd bij de uitvoering van correspondentietesten worden in dit rapport in kaart gebracht.

Concreet baseren we ons in grote lijnen op voorgaande expertise opgebouwd in het Centraal Expertencomité tegen discriminatie¹ om de volgende vragen te beantwoorden:

V1) Hoe kunnen correspondentietesten concreet worden uitgevoerd bij lokale besturen en wat zijn de randvoorwaarden?

V2) Hoe kunnen verschillende discriminatiegronden worden geoperationaliseerd?

V3) Hoe kan de minimumsteekproefgrootte worden bepaald in functie van de hoogte van de non-response, verwachte hoogte van discriminatiegraad, en de gewenste statistische zekerheid?

V4) Rekening houdend met de grootte van de gemeentes en het aantal jaarlijkse aanwervingen, in welke mate kunnen gemeentes samenwerken bij de uitvoering van correspondentietesten? (zie excel bestand)

V5) Welke methoden kunnen lokale besturen als sensibiliserend instrument inzetten?

Het doel van deze methodiek is dat er op een onafhankelijke, effectieve en wetenschappelijke wijze gemonitord kan worden wat de **aard** en het **volume** van de discriminatie is bij de eerste fase van de aanwerving door lokale besturen. De adviezen inzake methodiek in dit rapport zijn gebaseerd op

¹ Möbius (2020). *Eindrapport van het Centraal Expertencomité tegen discriminatie*. Brussel: Möbius.

enerzijds voorgaand wetenschappelijk onderzoek alsook de onderlinge besprekingen tijdens de reflectiemomenten georganiseerd in de periode januari-april 2022.

Zolang de wetenschappelijke methode wordt gevrijwaard, kunnen ook niet-academische instanties en personen aan de slag met dit rapport. Hoewel ongetwijfeld niet alle elementen in wat volgt van toepassing zijn op de specifieke procedures van ieder lokaal bestuur, is een overzicht van de voornaamste hordes toch cruciaal. Dit overzicht kan de toekomstige onderzoekers en uitvoerders aanmoedigen tot reflectie en kan een hulpmiddel zijn om specifieke moeilijkheden verder in kaart te brengen. We hopen dan ook dat dit rapport een eerste hulpmiddel vormt bij het uitvoeren van correspondentietesten bij lokale besturen als werkgever.

3. Kadering

3.1 Definities van discriminatie

Alvorens in detail de uitwerking van correspondentietesten te bekijken, kaderen we kort de centrale problematiek van discriminatie. Discriminatie verwijst naar de ongelijke, nadelige en niet te rechtvaardigen behandeling van personen op basis van een door de wet beschermde discriminatiegrond. Er zijn 19 ‘beschermde criteria’ opgenomen in de antidiscriminatiewetgeving van 2007.² De meest voorkomende vormen van discriminatie binnen het domein arbeid en werkgelegenheid zijn discriminatie op basis van geslacht, ‘raciale criteria’, handicap en leeftijd.³

In het Vlaamse **Gelijkekansendecreet** (art.16§1 en art.16§2) wordt er verder een onderscheid gemaakt tussen **directe** en **indirecte discriminatie**. **Directe** discriminatie stelt zich wanneer er in een vergelijkbare situatie iemand minder gunstig wordt behandeld (of zou worden behandeld) dan iemand anders op basis van wat bij wet “beschermde kenmerken” zijn. Deze kenmerken hoeven niet objectief aanwezig te zijn, maar kunnen ook toegeschreven zijn door derden. Zolang er geen legitiem gerechtvaardigde reden is voor deze verschillende behandeling is er dus sprake van discriminatie. Er is sprake van **indirecte** discriminatie als er in een schijnbare neutrale situatie (of door een schijnbaar neutrale (be)handeling) iemand met werkelijke of vermeende bij wet beschermde kenmerken in vergelijking met andere personen benadeeld wordt. Deze tweede vorm is dus ook discriminatie tenzij er een legitiem doel is en de middelen proportioneel passend en nodig zijn om dit doel te bereiken. Met betrekking tot de grond van handicap specifiek moet er vervolgens nog worden aangetoond dat (verdere) redelijke aanpassingen niet mogelijk zijn.⁴

² De volledige lijst bestaat uit zogenaamd ‘ras’, huidskleur, nationaliteit, afkomst, nationale of etnische afstamming (de vijf zogenaamde ‘raciale criteria’), handicap, geloof of levensbeschouwing, geslacht, seksuele geaardheid, leeftijd, vermogen, burgerlijke staat, politieke overtuiging, syndicale overtuiging, gezondheidstoestand, fysieke of genetische eigenschap, geboorte, sociale afkomst, en samenstelling van het gezin.

³ <https://www.unia.be/nl/publicaties-statistieken/cijfers-van-unia> (geraadpleegd 12 maart, 2022)

⁴ <https://codex.vlaanderen.be/portals/codex/documenten/1017082.html> (geraadpleegd 29 januari, 2022)

3.2 Instrumenten voor nulmeting

3.2.1 Praktijktesten en correspondentietesten

In de wetenschappelijke literatuur vallen correspondentietesten onder de bredere noemer van **praktijktesten**.⁵ Verhaeghe en Van de Bracht⁶ scheppen duidelijkheid in de verscheidene concepten door een typologie op te stellen gebaseerd op **drie criteria**. Elk criteria heeft zo zijn voor- en nadelen.

1. Wordt er een discriminerende vraag gesteld?
 2. Is er een test- en controleprofiel?
 3. Is er persoonlijk contact?
-
- 1) Indien er een **discriminerende vraag** gesteld wordt, kan dit potentieel worden bekeken als een vorm van uitlokking. Ook is deze techniek slechts beperkt inzetbaar. Er moet immers steeds een partij zijn aan wie men deze vraag kan stellen. Voorts kan een discriminerend antwoord op een vraag alleen als *intentie* tot discrimineren beschouwd worden. Er is namelijk geen sprake van een handeling, dus kan er geen discriminerend *gedrag* worden vastgesteld.
 - 2) Een **test- en controleprofiel** is belangrijk om te achterhalen op basis van *welk kenmerk* er wordt gediscrimineerd. Discriminerend gedrag kan uiteraard zonder een controleprofiel plaatsvinden. Om structureel benadeelde groepen vast te stellen is het echter nodig zorgvuldig voor de verschillen te controleren.
 - 3) **Persoonlijk contact** is verkiesbaar in situaties waar onpersoonlijk contact minder toepasbaar is. Zo zijn er bepaalde sectoren (bv. horeca) of fasen van een aanwervingsproces (bv. een sollicitatiegesprek na positieve evaluatie van het cv) waar onpersoonlijk contact moeilijk is. Echter zijn er ook verscheidene nadelen. Afhankelijk van het onderzoeksdesign kan dit een zeer tijdsintensieve methode zijn. Ook heeft de onderzoeker minder controle over de testomstandigheden en is het moeilijk de verschillende kenmerken te controleren.

⁵ Voor een omvattende inleiding tot o.a. correspondentietesten, zie ook Gaddis, M. (2018). An Introduction to Audit Studies in the Social Sciences. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3–44). Springer.

⁶ Verhaeghe, P.-P., & Van der Bracht, K. (2017). Praktijktesten: Van onderzoeksmethode naar beleidsinstrument tegen discriminatie? *Sociologos*, 38(1), 19.

Afhankelijk van het antwoord en de combinatie van deze drie criteria, maken Verhaeghe en Van de Bracht onderstaande typologie.

		Discriminerende vraag?	
		JA	NEE
Persoonlijk contact?	JA	Mystery calls & visits	Situatietesten
	NEE	Mystery mails	Correspondentietesten
		NEE	JA
		Test- en controleprofiel?	

Een eerste verschil situeert zich tussen de *mystery calls, visits*, en *mails* (ook gezamenlijk *mystery shopping* genaamd) enerzijds, en de situatie- en correspondentietesten anderzijds. Bij de eerste groep wordt er een discriminerende vraag gesteld, en is er geen test- en controleprofiel. Bij de tweede groep waar ook correspondentietesten deel van uitmaken, is er wel sprake van een controleprofiel. Discriminerende vragen worden dan weer niet gesteld. Het derde criterium van persoonlijk contact zorgt voor onderscheid binnen deze twee groepen. Binnen de groep van *mystery shopping* wordt er gesproken van *mystery calls* en *visits* indien dit persoonlijk contact aanwezig is. Is het afwezig, dan betreft het *mystery mails*. Hetzelfde onderscheid bestaat er tussen situatie- en correspondentietesten waarbij deze laatste direct persoonlijk contact vermijdt en dus een superieure controle bewaart over de te testen en niet te testen variabelen. Uit deze vier discriminatietesten spitst deze methodiek zich toe op correspondentietesten als meest geschikte methode voor een nulmeting. Correspondentietesten kunnen wel in een later stadium van het aanwervingsproces aangevuld worden met andere vormen van praktijktesten om discriminatie op te sporen (bv. situatietesten voor de fase van een sollicitatiegesprek).

In de wetenschappelijke literatuur worden **correspondentietesten** wel eens als de **gouden standaard** bestempeld om aanwervingsdiscriminatie tijdens het eerste stadium van het proces op te sporen. Als wetenschappelijke methode hebben ze vier voordelen. Ten eerste meten ze discriminatie rechtstreeks in het 'veld' van de aanwerving door lokale besturen en genieten ze dus van een hoge mate van **externe validiteit**. Ook **interne validiteit** is een sterkte omwille van de hoge mate van controle die behouden blijft over de variabelen tijdens het experiment. Dit laat causale interpretaties toe met betrekking tot

de variabele die verschilt tussen de testpersoon en de controle. Een derde voordeel is dat er **discriminair gedrag** wordt gemeten in tegenstelling tot bijvoorbeeld discriminerende intenties of stereotypen. Het laatste voordeel bestaat erin dat, in tegenstelling tot surveyonderzoek of interviews, correspondentietesten **geen last hebben van sociale wenselijkheid** aangezien de subjecten van het experiment niet weten dat ze getest worden. Deze laatste twee punten tonen dan ook dat **inhoudsvaliditeit of meetgeldigheid** een sterkte is van deze methode.

3.2.2 Beperkingen

Elke wetenschappelijke methode heeft zijn voor- en nadelen. Zo zijn er ook een aantal **beperkingen van correspondentietesten** waarbij wordt stilgestaan. Een eerste beperking is dat er slechts tijdens **de eerste fase van het aanwervingsproces** kan worden getest op discriminatie. Het unieke aan correspondentietesten is dat het om 'papieren' kandidaten gaat met het oog op het zo goed mogelijk controleren van het experiment. Indien een fictieve sollicitant bijgevolg een uitnodiging krijgt voor een volgende ronde (bv. bijkomende testen of interviews), wordt deze afgewezen. Dat impliceert dat het mogelijk voorkomen van discriminatie tijdens de vervolgfases van het aanwervingsproces buiten beschouwing wordt gelaten.

Een volgende beperking is dat de **oorzaak van discriminatie**, met andere woorden de **mechanismen** die discriminatie in de hand werken niet eenduidig kunnen worden gemeten via correspondentietesten.⁷ Bepaalde onderzoeksontwerpen kunnen wel voorzichtig specifieke aspecten belichten. Een recente meta-analyse van Lippens et al. (2020) vat samen welke studies bepaalde mechanismen van discriminatie ondersteunen of weerleggen.⁸ Echter dient een methode steeds het doel van het onderzoek te volgen, en het blootleggen van discriminatiemechanismen is eenduidiger te achterhalen met kwalitatieve methoden zoals vignette studies of diepte-interviews. Met andere woorden, correspondentietesten laten ons toe vast te stellen of er al dan niet discriminatie is tussen bepaalde groepen voor bepaalde vacatures. Echter is het niet mogelijk te bepalen **waarom** die er is.

⁷ Voor een overzichtelijke synthese van dit onderwerp, zie Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34(1), 181–209.

⁸ Lippens, Louis, et al. (2020). Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence. *Journal of Ethnic and Migration Studies*, 1-34.

3.2.3 Ethische reflecties en GDPR

Eerst moeten we benadrukken dat dit gedeelte **geen juridisch advies** vormt. Het blijft de eindverantwoordelijkheid van de lokale besturen om het onderzoek *AVG-compliant* op te stellen en hier advies over in te winnen. Wel schetsen we een aantal aandachtspunten die een leidraad kunnen vormen om mee aan de slag te gaan.

Bij het opzetten van correspondentietesten is het ten alle tijden noodzakelijk de lokale (of Vlaamse) DPO te informeren en **verder advies in te winnen** met het oog op compliance. Zo dient er een specifiek doeleinde gespecificeerd te worden (bv. “nulmeting van discriminatie”), en een rechtmatigheidsgrond geïdentificeerd te worden (bv. de rechtsgrond “vervulling van een taak van algemeen belang (art. 6.1.e AVG)” o.b.v. het Vlaamse gelijkekansen- en gelijkebehandelingsbeleid of met verwijzing naar de federale materie in de vorm van de antidiscriminatiewet, de antiracismewet, en de wet op de gelijkheid van vrouwen en mannen).

Verder dienen er ook afspraken gemaakt te worden met leveranciers (DPA) en dient er eventueel een *Data Protection Impact Assessment* (DPIA) te worden uitgevoerd. Deze DPIA dient een schatting te maken van het risico van de verwerking en vereiste maatregelen met aandacht voor a) de inperking van de rechten van de betrokkenen (onder meer de afwezigheid van informed consent), b) de kans dat de gegevens niet in ieder stadium van het onderzoek van eigenlijke anonimisatie of pseudonomisatie kunnen worden voorzien, en c) de schaal van het onderzoek (temporeel, datapunten, betrokken partijen, etc.) en de gekozen technologieën die aan bod komen bij iedere fase. Overige AVG-verplichtingen dienen ook in acht te worden genomen, waaronder het aanvullen van het verwerkingsregister, transparantie van het onderzoek, de uitvoering van het uitoefenen van de rechten van de betrokkenen, en meer.

Academisch onderzoek is verder ook sterk gereguleerd. Het is daarom ook gebruikelijk dat er vragen worden gesteld vanuit ethische commissies om de methodologische keuzes te motiveren. Eva Zschirnt⁹ vat de ethische bezorgdheden samen in drie pijlers:

⁹ Zschirnt, E. (2019). Research Ethics in Correspondence Testing: An Update. *Research Ethics*, 15(2), 1–21.

1. Afwezigheid van *informed consent* en het schenden van vrijwillige participatie:

Participanten aan wetenschappelijk onderzoek dienen doorgaans enkel deel te nemen wanneer ze goed geïnformeerd worden over de studie en hun participatie, en wanneer ze hiervoor expliciet hun toestemming verlenen (*informed consent*). Het informeren van participanten en hun toestemming vragen is evenwel onmogelijk bij correspondentietesten. Werkgevers of selectieverantwoordelijken op de hoogte brengen zou de testen minder betrouwbaar maken omwille van sociaal wenselijk gedrag, vrees voor sancties, het omzeilen van de testen en andere gedragseffecten. Omdat werkgevers of selectieverantwoordelijken niet op de hoogte zijn, hebben ze dus ook niet ingestemd met deelname aan het experiment. Hoewel dit een kritiek punt is in wetenschappelijk onderzoek zijn dit soort experimenten echter de enige manier om op betrouwbare wijze discriminatie aan te tonen. Ethische commissies volgen daarom doorgaans de redenering dat hier een uitzondering gegeven kan worden op het principe van *informed consent*. Wanneer lokale besturen deze correspondentietesten uitvoeren, zou dezelfde ethische afweging gemaakt kunnen worden. Wat wel mogelijk is, is dat lokale besturen de betrokken participanten (hun selectieverantwoordelijken) na het onderzoek duidelijk informeren en debriefen.

2. Misleiding van de participanten:

Bij het verzenden van cv's wordt er aan misleiding gedaan omdat de kandidaten fictief zijn en bijgevolg niet echt de intentie hebben om een job uit te oefenen. Echter heeft dit geen impact op het al dan niet discrimineren van de werkgever of selectieverantwoordelijke. Deze wordt niet uitgelokt om discriminerend te handelen. Het zijn normale selectieprocedures die worden bestudeerd. Het is dus een milde vorm van misleiding die wordt overschaduwd door het maatschappelijk nut dat correspondentietesten kunnen hebben in het bestrijden van discriminatie.¹⁰

¹⁰ Zie ook Riach, P. A., & Rich, J. (2004). Deceptive Field Experiments of Discrimination: Are they Ethical? *Kyklos*, 57(3), 457–470.

3. Potentiële negatieve gevolgen voor de participanten en/of werkgevers:

Hier vallen twee aspecten onder. Zo is er ten eerste de **extra last (tijd of kost)** die correspondentietesten veroorzaken bij de werkgever. Elke fictieve cv wordt immers ook bekeken. Dit is echter een berekening die de lokale besturen als opdrachtgever van het onderzoek moet maken in functie van het onderzoeksontwerp. Ten alle tijden wordt dit geminimaliseerd door de onderzoekers door een uitnodiging tot gesprek of het afnemen van testen meteen te weigeren (of de test niet af te nemen) en zich terug te trekken uit de selectieprocedure. Vervolgens is er het aspect van **privacy**. Bepaalde selectieverantwoordelijken kunnen zich geïrriteerd voelen indien data op individueel niveau wordt verzameld. Hier is het aan de onderzoekers om data voldoende te anonimiseren en slechts het totaalplaatje te rapporteren aan de opdrachtgevers. Cruciaal is ook dat er geen negatieve gevolgen mogen zijn voor de selectieverantwoordelijken zelf. De resultaten van het experiment mogen in de eerste plaats niet gebruikt worden voor berisping of sancties.¹¹

Mits voldoende motivering en zorg bij het opstellen van het onderzoeksontwerp in functie van de reële marktsituatie zijn er in het algemeen weinig ethische bezwaren. Verscheidene bezwaren zijn namelijk niet accuraat. En voor de aspecten waar er wel legitieme bezorgdheden zijn, luidt dat het een kleine extra kost of last is in verhouding met de negatieve maatschappelijke impact van discriminatie als deze onbestreden blijft.

Op basis van het bovenstaande kan worden besloten dat het **afnemen van correspondentietesten de meest aangewezen methode** is voor zowel nulmetingen als opvolgmetingen om discriminatie tijdens het aanwervingsproces bij lokale overheden vast te stellen.

¹¹ Correspondentietesten kunnen ook specifiek gebruikt worden om discriminatie bij individuele actoren op te sporen. Dit is echter niet het doel van deze methodiek die zich toespitst op een nulmeting. Voor een voorbeeld van het gebruik van correspondentietesten op deze manier, zie Verhaeghe, P.-P., & Van der Bracht, K. (2020). *How many correspondence tests are enough to detect discrimination among single agents? A longitudinal study on the Belgian real estate market* (Discussion Paper No. 678; GLO Discussion Paper, p. 23). GLO.

4. Methodiek correspondentietesten

4.1 Testen bij publieke en private sector

Correspondentietesten op de arbeidsmarkt worden zo goed als altijd uitgevoerd in de private sector. Deels wordt dit enerzijds toegeschreven aan een aantal **praktische aspecten (1)** die ervoor zorgen dat de uitvoering bij de publieke sector moeilijker of meer tijdsintensief is. Zo werken sollicitaties bij lokale overheden dikwijls via specifieke sollicitatieportalen (bv. via de SAP-service *successfactors*) waar unieke profielen dienen te worden aangemaakt om te solliciteren. Daarentegen wordt er bij klassieke correspondentietesten in de private sector voornamelijk gesolliciteerd via e-mails wat makkelijker te automatiseren is.¹²

Anderzijds situeert een belangrijk aspect zich langs de **kant van de werkgever (2)**. Bij correspondentietesten in de private sector test men voornamelijk voor bepaalde job categorieën (zoals administratie, sales, of boekhouder) of binnen bepaalde sectoren (bv. de transportsector, de schoonmaaksector). Deze vrijheid is er niet in de publieke sector omdat er meestal niet genoeg vacatures zijn binnen de looptijd van het experiment om zo specifiek de functies af te bakenen. Er worden dus zoveel mogelijk vacatures getest, en bijgevolg is de variatie in vacatures veel groter. Met andere woorden worden er dus profielen met masterdiploma's zowel als profielen zonder diploma voorbereid.

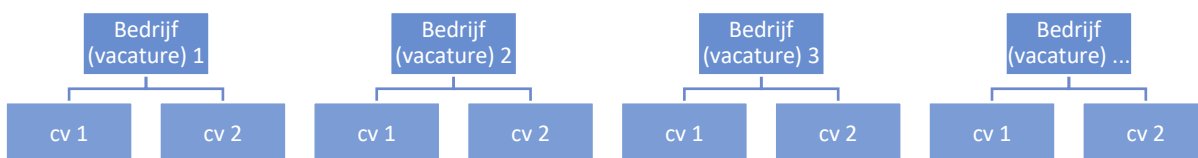
Hier sluit bij aan dat bij testen in de private sector werkgevers zelden **meerdere malen getest worden** tenzij daar een goede reden voor is (bv. vermoeden van discriminatie of werkgevers die doorheen het jaar meerdere vacatures hebben openstaan). Men selecteert dus telkens gelijkaardige vacatures maar bij andere bedrijven. Echter, bij testen in de publieke sector is de waaier aan 'werkgevers' veel beperkter (groendienst, IT-agentschappen, ruimtelijke ordening etc.) en vindt er dikwijls een voorselectie plaats via een **beperkte groep van verantwoordelijken** voor ieder lokaal bestuur. De taak van deze personen verschilt van stad tot stad, en kan variëren van de profielen communiceren naar de werkgever tot zelf

¹² Voor meer technische details over het gebruik van e-mails en het automatiseren, zie Crabtree, C. (2018) An introduction to conducting email audit studies. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer, pp.103–117.

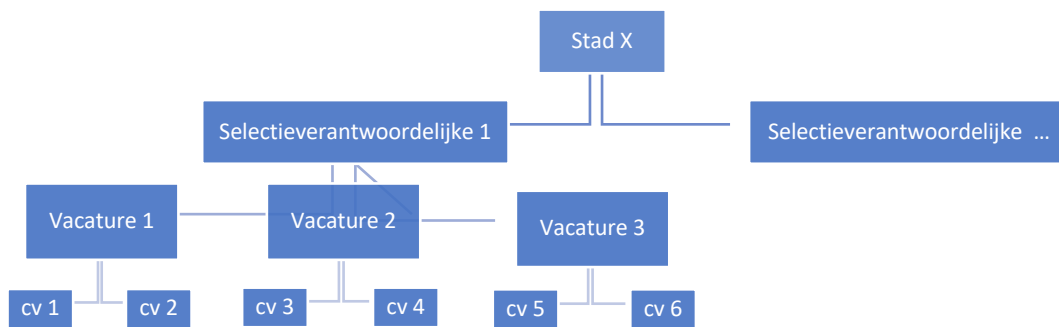
een eerste selectie maken. In de volgende pagina's worden deze personen 'selectieverantwoordelijken' genoemd hoewel de inhoud van de functie kan verschillen. Hoe dan ook, het gevolg is dat er een 'beperkter aantal ogen' al de fictieve kandidaten bekijkt. Het veelvuldig gebruik maken van fictieve sollicitanten bij een beperkt aantal selectieverantwoordelijken verhoogt de kans op detectie aanzienlijk. De voornaamste uitdaging is dus voldoende variatie creëren onder de sollicitanten om argwaan te vermijden. Mits voldoende voorbereiding zijn er een aantal aspecten die te automatiseren zijn (meer hierover in het volgende deel), maar de invulling van werkervaring gebaseerd op het gewenste profiel van de vacature blijft een werk van manuele aard.

Deze twee kwesties kunnen op volgende wijze worden samengevat: Bij correspondentietesten in de private sector is er een **lagere variabiliteit** van te testen vacatures en een **groot aantal ogen** dat de fictieve profielen (bv. cv's en motivatiebrieven) bekijkt. Bij testen in de publieke sector is er langs de andere kant een **hoge variabiliteit** en een **kleiner aantal ogen** dat de profielen bekijkt. Dit verhoogt de kans op **detectie** en noodzaakt een **groter aantal unieke cv's**. De verschillen worden schematisch samengevat in figuren 1 en 2.

Figuur 1: Schematisch overzicht correspondentietesten in de private sector



Figuur 2: Schematisch overzicht correspondentietesten in de publieke sector



Hoewel deze twee punten correspondentietesten bij de (lokale) publieke sector kunnen bemoeilijken, zijn er **ook gunstige aspecten** die de uitvoering efficiënter kunnen laten verlopen. Het voornaamste voordeel van testen in de publieke sector is dat de sollicitatieprocedure **structureel verankerd** is. Elke sollicitatie wordt volgens gelijkwaardige criteria en diepgang geanalyseerd. Als het profiel aan bepaalde vooropgestelde criteria voldoet gaat de kandidaat door naar de volgende ronde. Met andere woorden kan de onderzoeker voor de aanvang van de testen te weten komen welke details van het cv op welke manier worden bekeken. Concreet zijn er bijvoorbeeld de telefoonnummers van de sollicitanten. Als deze tijdens de eerste fase van de sollicitatie verplicht zijn in te geven in het online profiel maar praktisch niet worden gebruikt, dan kan hier tijd worden gewonnen door fictieve nummers te gebruiken. Op deze manier moeten er geen profiel-specifieke voicemails worden opgezet om antwoorden van de werkgever te ontvangen en kunnen kosten bespaard worden. Of als e-mailadressen alleen intern in een systeem zichtbaar zijn maar niet voor de selectieverantwoordelijke, dan is er minder nood aan realistische e-mailadressen. Als e-mailadressen niet zichtbaar zijn kunnen meerdere aliassen gebaseerd op één e-mailadres gebruikt worden. Aliassen zijn sneller op te zetten dan nieuwe emailadressen. Ook als het bestaan van bepaalde opleidingen bij specifieke scholen niet wordt nagekeken is dit een aspect dat kan worden uitbesteed aan randomisatie. Op deze manier kan een groot deel van de benodigde details worden uitgeklaard tijdens de voorbereidende fase van het experiment. Samenvattend kan dus de kennis van de selectieprocedures in de openbare sector onderzoekers een unieke kans geven om via geautomatiseerde randomisatie op erg korte tijd een breed scala aan details te genereren wat het vermijden van detectie ten goede komt.

We voegen hier ook kort toe dat er variatie kan zijn op de voorgaande tweedeling. Lokale besturen kunnen ook hun selectieprocedures geheel of gedeeltelijk uitbesteden. Hierdoor kan het element van een beperkt aantal selectieverantwoordelijken (en een beperkt aantal ogen) vermeden worden of niet van toepassing zijn. In dit geval zijn de voornaamste methodologische beperkingen dan ook niet van toepassing omdat detectie een kleiner risico vormt. In wat volgt gaan we uit van een *worst-case* scenario waarin het beperkte aantal selectieverantwoordelijken een primair obstakel vormt.

In de volgende hoofdstukken worden frequent correspondentietesten in de private sector aangehaald als *benchmark* voor de testen in de publieke sector. De wetenschappelijke literatuur heeft doorheen de jaren erg veel ervaring opgebouwd in de private sector,¹³ maar het is nodig deze bij te sturen om correspondentietesten af te nemen in de publieke sector.

4.2 Operationalisatie van de discriminatiegronden

Het vertalen van discriminatiegronden is bij correspondentietesten een belangrijke stap. Als het verschil in de discriminatiegrond (zoals etnische afkomst of geslacht) tussen de kandidaten niet duidelijk naar voor komt uit de cv's, dan mislukt het experiment. Gelukkig is er in de wetenschappelijke literatuur voldoende evidentie om deze gronden betrouwbaar en valide te vertalen naar signalen die kunnen worden meegenomen bij het solliciteren.¹⁴ Een hogere mate van validiteit kan bekomen worden door de operationalisatie eerste te verifiëren door middel van bijvoorbeeld een enquête waarin wordt nagegaan of de signalen (bv. gender of etnische afkomst gesignaleerd door namen) effectief zo worden geïnterpreteerd. In het onderstaande worden deze gronden die typisch aanleiding geven tot discriminatie bij het betreden van de arbeidsmarkt geoperationaliseerd.

- a. **Geslacht** wordt geoperationaliseerd via de namen van de fictieve kandidaten. Het is belangrijk dat er gekozen wordt voor namen die niet genderneutraal zijn en duidelijk

¹³ Quillian en Midtbøen vatten de internationale resultaten van meer dan 140 veldexperimenten (inclusief correspondentietesten) samen in een up-to-date overzicht (Quillian, L., & Midtbøen, A. H. (2021). *Comparative Perspectives on Racial Discrimination in Hiring: The Rise of Field Experiments*. *Annual Review of Sociology*, 14(7), 25).

¹⁴ Voor een recent overzicht inzake operationalisatie, zie Lippens, L., Vermeiren, S., & Baert, S. (2021). *The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments* (Discussion Paper No. 972; GLO Discussion Paper). GLO.

verwijzen naar mannen of vrouwen. Zo kunnen namen zoals “Kim” of “Kris” dikwijls zowel naar mannen of naar vrouwen verwijzen. Het operationaliseren van gender op deze manier is relatief eenvoudig, maar vereist dus wel goede kennis van de genderconnotatie van namen (iets wat misschien minder evident is bij niet-Belgische namen).

- b. **Leeftijd** is makkelijk te operationaliseren door de geboortedatum op het cv te zetten. Het geboortjaar maakt in één oogopslag duidelijk aan de selectieverantwoordelijke tot welke leeftijdsgroep een kandidaat behoort. Ook kan men de leeftijd vermelden als onderdeel van een korte introductie bij deze vacatures waar een begeleidende motivatiebrief nodig is. Deze worden soms niet al te nauwgezet gelezen, dus het geboortjaar vermelden draagt de voorkeur weg. In tegenstelling tot de private sector, is het ook courant om tijdens de sollicitatie het rijksregisternummer in te geven. Later wordt er dieper ingegaan op de technische aspecten, maar hier wordt reeds meegegeven dat de geboortedatum wel dient overeen te komen met de eerste zes cijfers van dit nummer.

De impact die leeftijd heeft op andere aspecten van het cv zijn echter ingewikkelder. Zo hebben kandidaten uit een hogere leeftijdsgroep meer **werkervaring** wat zich uit in een complexer of uitgebreider cv. Als het onderzoeksopzet is om leeftijdsdiscriminatie te meten, dan is dit iets om in het achterhoofd te houden. Wordt leeftijd echt getest? Of zijn het verschillen in werkervaring?¹⁵ Als leeftijd getest dient te worden dan is het belangrijk om de **relevante** ervaring zo gelijk mogelijk te houden tussen twee kandidaten, en het verschil in jaren werkervaring (leeftijd) op te vullen met ervaring die niet van pas zou komen bij deze job.

- c. **Etniciteit** wordt in onderzoek op dezelfde manier geoperationaliseerd als gender, namelijk via de namen van de respondenten. Recent onderzoek van Martiniello en Verhaeghe (2021)¹⁶ heeft echter aangetoond dat etnische verschillen tussen niet-Belgische namen moeilijker

¹⁵ Voor een onderzoek toegespitst op discriminatie op grond van leeftijd, zie bijvoorbeeld Baert, S., Norga, J., Thuy, Y., & Hecke, M. V. (2016). Getting Grey Hairs in the Labour Market: An Alternative Experiment on Age Discrimination. *Journal of Economic Psychology*, 57, 86–101.

¹⁶ Martiniello, B., & Verhaeghe, P.-P. (2021). *Signaling Ethnicity and Race Through Names? The Perception of Names from an Intersectional Perspective*. SSRN Electronic Journal.

correct worden geïnterpreteerd (bv. het verschil tussen Turkse en Marokkaanse namen). Het toewijzen van concrete etnische origines aan namen blijkt niet voor iedereen even gemakkelijk te zijn. Hetzelfde geldt voor gemengde namen, zoals een Marokkaanse achternaam met een Belgische voornaam. We raden bijgevolg aan het onderzoeksdesign zo eenduidig mogelijk te maken om de validiteit van de resultaten te beschermen. Gebaseerd op dit onderzoek is het bijvoorbeeld geen goed idee om binnen hetzelfde experiment verschillen in discriminatie te onderzoeken tussen twee Oost-Europese landen of tussen twee niet-Europese landen van dezelfde regio. En als het meten van zulke verschillen toch een vereiste is voor de opdrachtgever, dan is een pilootstudie om de validiteit van de namen te meten aangewezen.¹⁷

Etniciteit kan ook op andere manieren worden gesignaleerd, bijvoorbeeld door lidmaatschap aan bepaalde culturele organisaties of vrijwilligerswerk.¹⁸ Deze signalen zijn goed te gebruiken in combinatie met namen, maar zijn op zichzelf misschien minder betrouwbaar. De onderzoeker weet namelijk niet of iemand die dit cv leest effectief de link legt met de etnische herkomst van de kandidaat zelf. Ook is het plausibel dat dit onderdeel van het cv niet altijd gelezen wordt. Het bevindt zich dikwijls onderaan en is van minder belang dan relevante werkervaring. Verder is deze manier van operationalisatie minder evident bij correspondentietesten in de openbare sector waar een selectieverantwoordelijke tijdens de looptijd van het experiment misschien wel tientallen cv's te zien krijgt. Meerdere cv's met, bijvoorbeeld, interesses in Pools theater zou de kans tot detectie aanzienlijk verhogen in vergelijking met een Poolse naam. De naam staat altijd op een cv. Interesse in Pools theater komt minder frequent voor.

- d. **Handicap of chronische ziekte** operationaliseren kan op verschillende manieren. Een sterk voordeel van correspondentietesten in de openbare sector is dat er een beleid van actieve werving kan zijn, gericht op deze groep. Het is dus mogelijk dat er op het online

¹⁷ Bijvoorbeeld Edo (2019) maakt gebruik van een pilootstudie voor het testen van namen (Edo, A., Jacquemet, N., & Yannelis, C. (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household*, 17(1), 349–376).

¹⁸ Fossati, F., Liechti, F., & Auer, D. (2020). Can signaling assimilation mitigate hiring discrimination? Evidence from a survey experiment. *Research in Social Stratification and Mobility*, 65, 100462.

sollicitatieformulier een onderdeel is waar dit kan worden aangegeven. Soms wordt er ook plaats voorzien waar kan worden aangegeven welke aanpassingen aan de werkplek de kandidaat nodig heeft om de job (of verdere selectietesten) te kunnen uitvoeren (bv. speciale software, gemakkelijke toegang tot het kantoor voor mensen met locomotorische handicap).

Deze optie is echter niet altijd beschikbaar, dus raden we ook aan om op andere manieren de handicap of chronische ziekte te signaleren indien nodig. Hier is enige voorzichtigheid geboden. Het vermelden van een handicap of chronische ziekte op het cv wordt namelijk soms afgeraden, vooral indien deze geen impact heeft op de job inhoud.¹⁹ De redenering is dat de werkgever zo niet kan discrimineren gebaseerd op handicap of chronische ziekte en zich moet focussen op de productiviteitsaspecten van het cv. We weten niet in welke mate dit advies door de gemiddelde sollicitant wordt opgevolgd, maar het feit dat dit een punt van discussie is maakt het afnemen van correspondentietesten met deze discriminatiegrond uiteraard uitdagender. Vooral in de publieke sector waar detectie een delicaat probleem vormt, is het belangrijk dat onze profielen niet te sterk afwijken van een gemiddeld cv.

Het signaleren kan gebeuren op verschillende manieren: a) **vermelden (beschrijvend)** dat men gebruik maakt van een rolstoel/slecht ziet/auditieve beperkingen heeft. In dit geval werkt men **best via een motivatiebrief** om dit op een positieve manier te kaderen.²⁰ Ook b) kan men vermelden dat men **recht heeft op bepaalde bijstand** (bv. Vlaamse Ondersteuningspremie). c) Een **opleiding aan buitengewoon onderwijs** is een derde manier van signaleren. Hier moet echter worden opgelet dat het onderwijsniveau tussen de kandidaten gelijk wordt gehouden.²¹

Ook dienen de vacatures voor het testen van deze discriminatiegrond zorgvuldiger te worden gekozen. Een bepaalde handicap kan de productiviteit van een potentiële werknemer

¹⁹ Dit advies komt voor op verscheidene populair cv-websites zoals cvster.nl of cvwizard.be. Andere blogs raden aan om dit juist wel te vermelden.

²⁰ Zie bijvoorbeeld Baert (2016) voor een operationalisatie van handicap via cv en begeleidend schrijven (Baert, S. (2016). Wage subsidies and hiring chances for the disabled: Some causal evidence. *The European Journal of Health Economics*, 17(1), 71–86).

²¹ Een creatieve manier van operationalisatie van handicap vinden we bij Bellemare et al. die experimenteren met video-cv's bij rolstoelgebruikers (Bellemare, C., Goussé, M., Lacroix, G., & Marchand, S. (2020). Video Resumes and Job Search Outcomes: Evidence from a Field Experiment. *SSRN Electronic Journal*).

beïnvloeden afhankelijk van de invulling van de job (bv. een vacature voor redder in een kustgemeente en iemand met een locomotorische handicap is moeilijker te combineren). Het niet geselecteerd worden is voor deze functies niet noodzakelijk een vorm van directe discriminatie. Dit zorgt niet alleen voor problemen met de validiteit van het experiment, maar heeft ook weer een impact op mogelijke detectie. Als een stad of gemeente normaal gezien bijvoorbeeld gemiddeld 50 sollicitaties per jaar krijgt van iemand met een handicap of chronische ziekte, dan zou het erg verdacht zijn als hier 200 fictieve sollicitaties aan worden toegevoegd. Met andere woorden dienen de vacatures voor het testen van een handicap of chronische ziekte met meer omzichtigheid te worden gekozen.

- e. Het valide operationaliseren van **seksuele oriëntatie** kan op verschillende manieren. Recent onderzoek door Acquisti en Fong (2021) maakt bijvoorbeeld gebruik van sociale media. Ze vertrekken van de hypothese dat werkgevers courant sociale media gebruiken om extra informatie over sollicitanten op te zoeken. Zodoende maken ze fictieve profielen aan op de meest voorname kanalen, en signaleren hun seksuele oriëntatie hierop.²² Hoewel deze operationalisatie misschien minder van toepassing is in de publieke sector waar het nagaan van sociale media minder courant is, toont dit onderzoek aan hoe op creatieve manieren toch moeilijkere aspecten kunnen worden geoperationaliseerd.

Signalen kunnen ook op andere manieren gegeven worden, bijvoorbeeld door het beschrijven van de gezinssituatie in een begeleidend schrijven of motivatiebrief.²³ Aldus kan via de naam van de fictieve partner diens geslacht worden gesignaleerd. In combinatie met de naam van de sollicitant kan zodoende in beperkte mate seksuele oriëntatie worden weergegeven.²⁴ Eerder onderzoek heeft seksuele oriëntatie ook geoperationaliseerd door middel van vrijwilligerswerk of vrijetijdsbesteding. Zo zijn bijvoorbeeld een aantal kandidaten in een onderzoek van Weichselbaumer (2015) lid van de grootste LGBT-organisatie in Duitsland.²⁵ Echter is ook deze operationalisatie niet perfect. Uit een overzichtsstudie van

²² Acquisti, A., & Fong, C. (2020). An experiment in hiring discrimination via online social networks. *Management Science*, 66(3), 1005–1024.

²³ Baert, S. (2013). *Career Lesbians: Getting Hired for Not Having Kids?* (Discussion Paper No. 7767; IZA Discussion Paper).

²⁴ Dit werkt natuurlijk niet om specifiek biseksualiteit te signaleren.

²⁵ Weichselbaumer, D. (2015). Testing for Discrimination against Lesbians of Different Marital Status: A Field Experiment. *Industrial Relations: A Journal of Economy and Society*, 54(1), 131–161.

Lippens et al. (2021) blijkt dat door lid te zijn van zulke organisaties sollicitanten ook als militant kunnen worden beschouwd en bijgevolg minder worden uitgenodigd (p.20).²⁶ Hier bestaat dus het risico dat er discriminatie optreedt maar dat we deze niet zuiver kunnen toeschrijven aan seksuele oriëntatie.

Deze manieren van operationaliseren zijn ook minder evident bij het afnemen van correspondentietesten bij lokale besturen waar een beperkt aantal selectieverantwoordelijken een relatief groot aantal fictieve cv's te zien krijgt. Het is namelijk niet zo dat elke echte sollicitant seksuele oriëntatie wil benadrukken aangezien dit aspect tot de privésfeer behoort. Daarom zou een groot aantal sollicitaties waarin dit wél wordt benadrukt erg verdacht zijn en de kans op detectie aanzienlijk verhogen. Een mogelijkheid is om verschillende soorten signalen te gebruiken (bv. test kandidaat 1: naam van partner vermelden, test kandidaat 2: vermelden van vrijwilligersorganisatie, test kandidaat 3: via sociale media) om de variabiliteit hoog genoeg te houden.

- f. **Genderidentiteit** is tot op heden minder frequent bestudeerd. In een recent onderzoek van Baert et al. (2021) voor de Stad Gent werd gekozen om **transgender** identiteit te operationaliseren via de naam.²⁷ De auteurs wisselen mannelijke en vrouwelijke paren af, waarbij er respectievelijk een transgender man of vrouw als 'minderheidskandidaat' wordt geselecteerd. Voor de helft van de profielen werden in het cv zowel de geboortenaam als huidige naam opgelijst. De andere helft vermeldde een gelijkaardige wijziging maar dan via een motivatietekst waarin werd beschreven dat sommige oudere documenten misschien nog naar een oude naam verwijzen.

Indien men discriminatie van **non-binaire personen** wil testen, dan moet er worden gebruik gemaakt van een andere operationalisatie. Een mogelijkheid is om te kiezen voor het specificeren van voornaamwoorden zoals 'die/hen/hun' of 'die/die/hun'. Dit is natuurlijk niet exact omdat eender wie die genderneutraal wil zijn deze voornaamwoorden kan gebruiken. Verdere operationalisatie alleen via het cv is moeilijk. Zo kan lidmaatschap tot bepaalde

²⁶ Lippens, L., Vermeiren, S., & Baert, S. (2021). *The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments* (Discussion Paper No. 972; GLO Discussion Paper). GLO.

²⁷ Baert, S., Dalle, A., Louis, L., & Malfait, L. (2021). *Discriminatie op de Gentse arbeidsmarkt anno 2021: Rapport*.

groepen of vrijwilligerswerk bij organisaties die gericht zijn op het ondersteunen van non-binaire personen worden toegevoegd. Het is echter maar de vraag of deze vermeldingen duidelijk genoeg zijn om de link te leggen met non-binaire personen, en of de kandidaat dan effectief als een non-binair persoon wordt aanzien door de werkgever. Beter is om hier te opteren voor een motivatiebrief of motivatietekstje dat de situatie van de persoon beschrijft met de wens om niet via de tweedeling man/vrouw te worden geïdentificeerd.

Hier moet ook weer opgemerkt worden dat het **aantal te testen vacatures** moet worden gebalanceerd met de kans op detectie. Zoals bij handicap of chronische ziekte is het cruciaal een beeld te krijgen van hoeveel cv's de stad of gemeente gemiddeld per jaar ontvangen die kunnen worden gekoppeld aan transgender of non-binaire personen. Zorgt het experiment ervoor dat dit getal met een veelvoud toeneemt, dan moet het aantal sollicitaties worden beperkt om detectie te vermijden.

4.3 Steekproefgrootte

Om statistisch significante conclusies te trekken is het noodzakelijk dat de steekproefomvang voldoende hoog is. We volgen dan ook de aanbeveling van het Centraal Expertencomité tegen discriminatie²⁸ waarin de vuistregel van **minimum 200 gepaarde correspondentietesten** (of 400 niet-gepaarde testen) is opgenomen. Hierbij moet worden opgemerkt dat dit geen absolute cijfers zijn. Een kleiner aantal testen kan bijvoorbeeld nog steeds significante resultaten produceren. Veel hangt echter af van a) de **hoogte van de non-respons**, en b) de **hoogte van de discriminatiegraad**. Non-respons 'verbergt' discriminatie omdat zowel de test- als controlepersoon geen uitnodiging krijgen. Het is dus niet geweten of de werkgever zou discrimineren in de eerste fase van het proces. Respons betekent dus dat minstens één van onze fictieve kandidaten een positief antwoord krijgt. Bij een erg hoge non-respons kan het aangewezen zijn om de profielen kwalitatief te verbeteren om meer reactie van de werkgevers uit te lokken. Een **hogere mate van non-respons noodzaakt ons tot meer testen**. Als er meer discriminatie is, wordt het statistisch eenvoudiger om dit aan te tonen. Daarom dat bij een **hogere mate van discriminatie minder testen** nodig zijn.

²⁸ Möbius (2020). *Eindrapport van het Centraal Expertencomité tegen discriminatie*. Brussel: Möbius.

Natuurlijk is het moeilijk om nog voor een nulmeting een duidelijk zicht te hebben op deze twee aspecten. Inzicht verwerven is namelijk net een doelstelling van nulmetingen. Echter kan **voorgaand onderzoek** in gelijkaardige situaties een leidraad vormen. Ook, als de looptijd van de studie lang genoeg is, kan worden geopteerd voor een **pilootproject** om deze twee aspecten te bestuderen. Uit overleg met lokale besturen blijkt dat bij aanwervingen in de openbare sector er dikwijls getrapte selectieprocedures zijn waardoor de *non-respons eerder laag* ligt als de sollicitatie aan alle criteria voldoet, en de *hoogte van discriminatie laag* is om diezelfde redenen. Om deze redenen concluderen we om het minimumaantal testen niet te verlagen en het aantal van 200 als richtlijn te behouden voor gepaarde correspondentietesten.

Concreet betekent het voorgaande dat er nood is aan voldoende vacatures tijdens de periode van het experiment. Bij het uitvoeren van correspondentietesten op lokaal niveau is het dan ook maar de vraag of er voldoende vacatures ter beschikking komen. In de eerste plaats kan er gekeken worden naar de vacatures die in de te testen scope vallen van de voorgaande jaren. Indien deze telkens beduidend minder dan 200 zijn is het aangewezen om correspondentietesten uit te voeren **in samenwerking met verschillende lokale besturen**. Dit komt enerzijds de validiteit van de resultaten ten goede, maar heeft ook praktische en financiële voordelen. De uitvoering van correspondentietesten is erg tijdsintensief, en het kostenplaatje kan een hindernis vormen voor kleinere steden en gemeenten met een beperkter budget. Het nadeel is wel dat dit **gemeentespecifieke cijfers belet** en bijgevolg dus ook conclusies met betrekking tot discriminatie in een bepaalde gemeente.

Een laatste punt betreft het aantal te testen **discriminatiegronden**. Hoe meer discriminatiegronden men wil testen, hoe meer correspondentietesten er in totaal nodig zijn en hoe groter het gevaar voor detectie. Zo zijn gender, leeftijd, en etniciteit (afhankelijk van de specifieke etnische herkomsten) relatief veilige opties omdat de minderheidsgroepen relatief goed vertegenwoordigd zijn in onze samenleving. Gronden zoals handicap en chronische ziekten, en genderidentiteit zijn delicaat. Het is belangrijk om hier een idee te krijgen van het gemiddelde aantal sollicitaties dat voldoet aan de te testen kenmerken. Zo vallen 200 fictieve sollicitaties van iemand met een handicap op als een stad er zo maar gemiddeld 50 per jaar ziet. Er resten dan twee mogelijkheden om detectie te verlagen. Een eerste optie is om de **looptijd van het experiment te verlengen** (bv. twee jaar in plaats van een jaar). Een tweede optie is om te kiezen voor **een kleiner aantal cv's** met deze discriminatiegrond, en zo een exploratief

onderzoek te ondernemen. Het is van vitaal belang dat de resultaten hier op een correcte manier worden geïnterpreteerd. Bij dergelijke exploratieve testen is de steekproefgrootte vaak te klein om tot statistisch significante conclusies te komen. Concreet wil dit zeggen dat er niet kan worden besloten op een wetenschappelijke manier of er sprake is van discriminatie of niet.

4.4 Ongepaarde, duo- of triotesten

Correspondentietesten kunnen op verschillende manier afgenomen worden. Hier wordt het onderscheid gemaakt tussen ongepaarde testen, duo-, of triotesten. Specifiek gaat het over hoeveel fictieve sollicitaties er plaatsvinden per vacature: één, twee, of drie.

Er zijn enkele **praktische verschillen** die meteen opvallen. Zo is de extra **last** die een fictieve sollicitant veroorzaakt natuurlijk lager als er slechts één test is per vacature. Duo- en zeker triotesten **verhogen de 'last'** van het testen voor de sollicitatieverantwoordelijken. Ze hebben immers telkens twee of drie dossiers met fictieve kandidaten te verwerken. De extra kost die dit met zich meebrengt moet in de eerste plaats worden bekeken door de opdrachtgever, indien mogelijk in samenspraak met het selectiebureau. Als er bijvoorbeeld een chronisch tekort is aan medewerkers om dossiers te screenen dan is de additionele last van triotesten in vergelijking met duotesten moeilijker te rechtvaardigen. De situatie in verband met screening en de werkdruk is echter bij ieder lokaal bestuur verschillend, dus is het aan de lokale instanties om hier een passend evenwicht in te vinden.

Voorgaand onderzoek heeft aangetoond dat triotesten succesvol gebruikt kunnen worden om het aantal te testen discriminatiegronden te verhogen.²⁹ Echter zijn er ook redenen om triotesten achterwegen te laten en toch duotesten of zelfs ongepaarde testen te verkiezen. De voornaamste reden is de **grotere kans op detectie**.³⁰ In het bijzonder kan dit een probleem zijn voor vacatures waar slechts beperkt reactie op komt of erg specialistisch zijn. Zo is er natuurlijk ook meer variatie nodig bij het

²⁹ Verhaeghe, P.P., Martiniello, B., Ghekiere, A. (2020). *Discriminatie door makelaars op de huurwoningmarkt van Antwerpen*. Brussel: Vakgroep sociologie – Vrije Universiteit Brussel.

³⁰ Zie bv. Weichselbaumer, D. (2015). Testing for Discrimination against Lesbians of Different Marital Status: A Field Experiment. *Industrial Relations: A Journal of Economy and Society*, 54(1), 131–161.

voorbereiden van cv's voor de drie verschillende profielen. Gepaard hierbij gaat extra zorg bij het constant houden van de cv's en alleen te variëren in discriminatiegrond. Na de sollicitatie is er ook het aspect van de reacties op de sollicitaties. Correspondentietesten zijn namelijk in de eerste plaats geschikt om de eerste fase van het solliciteren te onderzoeken. De methode vereist bijgevolg dat een uitnodiging tot gesprek of volgende ronde steevast moet worden geweigerd op een gelijkwaardige manier, en dit moet dus ook in het drievoud gebeuren. Met andere woorden, ieder moment tijdens de sollicitatieprocedure waar de methode vereist dat profielen en werkwijzen constant worden gehouden, verhoogt de kans op detectie. En bij triotesten zijn zulke momenten nu eenmaal talrijker dan bij duotesten. Het is dus aangewezen om spaarzaam om te springen met triotesten en deze te reserveren voor vacatures waarbij er voldoende sollicitanten zijn.

In het geval van correspondentietesten bij lokale besturen speelt ook **het aantal beschikbare** vacatures mee. Dit aantal is normaal gezien beperkter dan in de private sector. Dit is dus een argument in het voordeel van duotesten of triotesten. Bij een lager aantal beschikbare vacatures kan er bijvoorbeeld gekozen worden voor triotesten om toch een groter aantal discriminatiegronden te kunnen testen. Zo solliciteren er drie fictieve kandidaten voor eenzelfde vacature. Een eerste testpersoon (bv. een vrouw van Poolse origine met een gemiddelde leeftijd), een tweede testpersoon (bv. een vrouw van Belgische origine met een hogere leeftijd), en een controlepersoon (bv. een vrouw van Belgische origine met een gemiddelde leeftijd). Op deze manier wordt er op één vacature zowel getest op 1) etnische origine, en 2) leeftijd.

Het is moeilijk om tot een conclusie te komen. Detectie is één van de meest reële problemen bij het uitvoeren van deze testen bij lokale overheden dus zijn ongepaarde testen aangewezen. Langs de andere kant is het lage aantal vacatures ook een heus struikelblok dus zijn triotesten ideaal. Hier kan wederom de interne kennis van de opdrachtgever een concrete leidraad vormen bij het inschatten van het gemiddelde aantal echte sollicitaties en of duo, al dan niet triotesten mogelijk zijn.

5. Stappenplan

In dit deel worden stap voor stap de drie fasen van een experiment met correspondentietesten behandeld. Zo wordt eerst de **voorbereiding**, vervolgens de **uitvoering**, en uiteindelijk de **analyse** besproken.³¹

Fase 1: Voorbereiding

De voorbereidende fase is misschien wel de belangrijkste van het experiment. Het is hier dat de meeste tijd- en efficiëntiewinsten kunnen worden geboekt. Het is dan ook cruciaal om minstens een paar vacatures te testen als onderdeel van deze fase. Aangezien dit gewoon een kleinschaligere versie is van het experiment waarna de voorbereidende stappen nogmaals worden overlopen, wordt er hier niet verder op ingegaan. Afhankelijk van het te testen platform is het mogelijk dat er nog andere voorbereidingen moeten worden getroffen. Het onderstaande is dus niet exhaustief maar behandelt wel de meeste cruciale aspecten.³²

Communicatie

Zeker in de publieke sector is het al dan niet communiceren van correspondentietesten geen makkelijke beslissing. Het **vooraf communiceren** zal tot gevolg hebben dat er geen zuiver experiment meer kan worden uitgevoerd omdat selectieverantwoordelijken zich anders kunnen gedragen als ze weten dat ze gecontroleerd kunnen worden. Hoewel dit vanuit een wetenschappelijke optiek dus af te raden is, kan het communiceren soms nodig zijn omwille van politieke redenen. In de mate van het mogelijke kan dit dan worden beperkt tot bepaalde sleutelfiguren die enerzijds voldoende aansprakelijkheid hebben om de testen te verantwoorden naar diegenen die worden getest, en anderzijds kunnen optreden als een soort informant met betrekking tot de systemen die getest gaan worden. Het kort voor de start van het experiment rechtstreeks benaderen van diegene die worden getest, wordt ten stelligste afgeraden. Ook

³¹ Een voorbeeld van voorgaand onderzoek dat de verschillende stappen duidelijk bespreekt, zie Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991–1013.

³² Voor een uitgebreide discussie over de technische aspecten en hoe deze weerslag hebben op externe validiteit, zie ook Lahey, J. en Beasley, (2018). Technical Aspects of Correspondence Studies. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81–102). Springer.

zit er liefst enige tijd tussen de aankondiging en de uitvoering zodat deze informatie niet vers meer in het hoofd van de betrokkenen zit.³³

Het **communiceren nadien** is een ander vraagstuk. Bij correspondentietesten in de private sectoren zijn er zowel voor- als tegenargumenten te vinden in de wetenschappelijke literatuur. In de publieke sector zijn we echter voorstander van een soort *debriefing*. Indien lokale besturen met selectieverantwoordelijken werken, dan krijgen deze dikwijls vele cv's te zien, inclusief veel fictieve cv's die worden gebruikt voor de correspondentietesten. Omdat succesvolle sollicitaties telkens moeten worden afgebroken, kan het gevoel ontstaan bij deze selectieverantwoordelijken dat bepaalde profielen (bv. met een bepaalde handicap of etniciteit) buitensporig afhaken. Zulke gevoelens kunnen leiden tot vooroordelen en discriminatie net in de hand werken. De bedoeling van het experiment is om discriminatie te meten en om het uiteindelijk de wereld uit te helpen. Daarom is het belangrijk om te duiden waarom er veel mensen met een bepaald profiel leken af te haken tijdens de selectieprocedure.

Proces van aanwerving

Bij correspondentietesten in de openbare sector is het belangrijk het **aanwervingsproces** goed te begrijpen. Veel van het voorbereidende werk moet namelijk aangepast worden aan de specifieke vereisten van het solliciteren bij de specifieke lokale overheid. Indien er omwille van steekproefgrootte of kosten wordt geopteerd om samen te werken met verschillende steden en gemeenten, dan is het cruciaal om de verschillpunten tussen deze processen goed in kaart te brengen.

De aanwervingen worden door een waaier van **platformen** in goede banen geleid. Enkele voorbeelden zijn *JobSolutions*, *SAP successfactors*, *Carerix*, *HR Organizer*, *OTYS Recruitment Technology*, of eenvoudige webformulieren. Deze brede waaier aan platformen vertaalt zich telkens in een (licht) verschillend aanwervingsproces. Zo moet er soms eerst een volledig profiel met cv worden aangemaakt. Pas nadat dit profiel is aangemaakt, kan de sollicitant zich kandidaat stellen voor specifieke vacatures. Ook zijn er systemen waar demografische data en contactgegevens tijdens de sollicitatie vereist zijn. Een eenvoudigere procedure bestaat uit slechts het ingeven van contactgegevens (naam, e-mailadres,

³³ Het zogenaamde "Hawthorne-effect" dat hierdoor gecreëerd wordt, blijkt namelijk tijdelijk te zijn (Purssell, E., Drey, N., Chudleigh, J., Creedon, S., & Gould, D. J. (2020). The Hawthorne effect on adherence to hand hygiene in patient care. *Journal of Hospital Infection*, 106(2), 311–317.).

telefoonnummer) en het opladen van het cv en eventueel een motivatiebrief. Soms is er zelfs geen platform, en moet er gewoon een cv worden verzonden naar een e-mailadres.

Het is onmogelijk om een exacte werkwijze te voorzien die overeenkomt met elk platform en elke manier van solliciteren. Daarvoor is er te veel variatie tussen de lokale besturen. Het blijft niettemin belangrijk om deze processen mee in rekening te nemen. Vooral bij lokale samenwerkingen waarbij er verschillende systemen worden gebruikt in één experiment is het cruciaal om eerst en vooral een duidelijk beeld te krijgen van de verschillende processen. Zo kan het bekijken van een aantal advertenties de onderzoeker al helpen: Het wordt duidelijk welke informatie er wordt gevraagd, welke gegevens verplicht en optioneel zijn, of er een motivatiebrief nodig is, en of er specifieke vragen moeten worden beantwoord (bv. *“Wat spreekt je aan in deze functie?”* of *“Hoe zie jij je dagelijkse taken in deze functie?”*).

Onderzoekers zijn echter alleen maar in staat het **front-end** van het aanwervingssysteem te bekijken. Een medewerker van de overheidsdiensten die de onderzoeker kan helpen met het **back-end** plaatje te verduidelijken is bijgevolg een grote meerwaarde. Concreet is het interessant te weten wat de selectieverantwoordelijke kan zien van de gegevens die worden ingevuld door een sollicitant. Zien ze het e-mailadres en de naam? Dan moeten realistische namen en adressen worden aangemaakt wat extra tijd kost. Zien ze dit niet en wordt dit op de achtergrond in een systeem bijgehouden, dan kan tijd worden bespaard door algemene e-mailadressen te maken met aliassen. Wordt het gsm-nummer van de sollicitant niet of nauwelijks gebruikt? Dan is dat weer een aspect waarop zowel tijd als kosten kunnen worden bespaard. Met andere woorden loont het dus om in de mate van het mogelijke samen doorheen het hele proces van een correspondentietest te gaan met iemand van de **back-end** om de pijnpunten op te lijsten. Deze zaken bijsturen tijdens het experiment zelf kan veel tijd kosten - tijd die beter kan worden besteed aan het verhogen van de kwaliteit van de cv's.

Diploma/certificaten check

Sommige systemen vereisen dat er een kopie van het diploma of andere certificaten worden opgeladen bij het solliciteren. Ook kunnen diploma's worden opgezocht in de LED-databank op basis van demografische gegevens van de sollicitant (bv. via het rijksregisternummer). Dit brengt natuurlijk moeilijkheden met zich mee. Want waar het creëren van een fictief persoon *an sich* geen probleem is, is het opstellen van valse certificaten en diploma's **in naam van bestaande instanties** dat wel. Valsheid

in geschrifte is namelijk een strafbaar feit en dus kan deze route alleen maar worden genomen mits uitdrukkelijk (en schriftelijk) akkoord van de onderwijsinstelling. Praktisch gezien is dit niet altijd een oplossing omdat men van verschillende instanties diploma's en certificaten nodig heeft. Dit telkens bespreken, onderhandelen en vastleggen zou erg veel tijd vragen. Hoewel dit dus niet onmogelijk is, is het praktischer om op een hoger niveau een overeenkomst te sluiten met een duidelijk afgelijnd juridisch kader (en dus niet telkens tussen één stad en de verstrekker van het getuigschrift).

Ervaring leert ons echter dat toelating krijgen moeilijk kan zijn. Is dit geen mogelijkheid, dan resten er ons nog een paar opties. Er is de mogelijkheid om het **proces van aanwerving te wijzigen**. Zo kan de opdrachtgever – het lokale bestuur – het opladen (en/of controleren) van het diploma bijvoorbeeld laten verschuiven naar een volgende ronde onder het mom van efficiëntie: indien alleen de diploma's gecontroleerd worden na de eerste selectieronde, dan bespaart dit tijd aan de kant van de selectieverantwoordelijken. Bij correspondentietesten breekt de onderzoeker het proces af vóór de sollicitant naar de volgende ronde doorgaat, dus zou dit het probleem oplossen. Wordt dit proces echter gewijzigd net wanneer het experiment start, dan neemt de kans op detectie natuurlijk toe. Medewerkers kunnen zich vragen beginnen te stellen, of het nieuwe proces negeren en toch de persoon opzoeken in een databank of om het diploma vragen. De foutmeldingen die de fictieve rijksregisternummers met zich meebrengen verhogen dan weer de argwaan. Belangrijk hier is dat er ook wordt gecontroleerd op deze wijziging zodat de onderzoeker zekerheid heeft dat deze op de werkvloer wordt geïmplementeerd. Liefst wordt deze wijziging dan ook minstens een paar maanden tot zelfs een jaar voor het experiment geïmplementeerd zodat de verantwoordelijken zich hiermee vertrouwd voelen.

Een derde mogelijkheid is om te werken met de **diploma's en certificaten van echte personen** (bv. vrijwilligers). Indien mensen toestemming zouden geven om hun informatie te gebruiken bij het maken van cv's voor correspondentietesten, dan kunnen hun diploma's ook gebruikt worden voor de diplomacheck. Ook dit is praktisch moeilijker en een werk van langere adem. Er is namelijk nood aan een relatief grote set van gegevens om detectie te vermijden. Daarnaast moeten er ook voldoende gelijkwaardige diploma's voor verschillende leeftijdsgroepen, functieniveaus, job profielen, etc. worden verzameld. Ook is er het aspect van gegevensbescherming waarbij het nodig is om goede afspraken te maken in verband met het doel, de termijn, en de opslag van deze echte gegevens.

Selectie van vacatures

In wetenschappelijk onderzoek moet er voor de selectie van de **steekproef** telkens eerst de **populatie** worden geselecteerd. De populatie is de groep eenheden waarover uitspraken worden gedaan terwijl de steekproef de specifiek groep is die wordt getest. Het is steeds cruciaal dat deze steekproef representatief is voor de populatie. Dit is vooral een moeilijkheid bij testen in de private sector waar er keuzes moeten worden gemaakt met betrekking tot hoe men vacatures zoekt, welke kanalen men wel of niet raadpleegt, enzovoort. Steden en gemeenten hebben dikwijls slechts één of een beperkt aantal kanalen waar vacatures staan opgesteld wat de steekproeftrekking vergemakkelijkt. De specifieke selectie van de vacatures (bijvoorbeeld extern, intern, alleen stadsbestuur of inclusief OCMW en scholen) kan door de lokale besturen worden bepaald. In sommige gevallen zal het moeilijk zijn om bepaalde functieniveaus of profielen te testen omwille van de selectieprocedures. Zo is het bijvoorbeeld mogelijk dat nog voor het solliciteren met cv er eerst een selectieronde is die gebruik maakt van testen. Ook kan het zijn dat er reeds vanaf de eerste stap een gesprek is en pas nadien een cv-screening. Het is dan ook cruciaal om de bijzonderheden van de selectieprocessen op lokaal niveau te bestuderen om te zien in welke mate er oplossingen kunnen worden gevonden om het afnemen van correspondentietesten mogelijk te maken (bijvoorbeeld door middel van proceswijzigingen). Deze **praktische overwegingen** kunnen ook de selectie van de vacatures mee bepalen.

Er zijn ook **nadelen** verbonden aan correspondentietesten in de publieke sector. Het grotere aantal vacatures in de private sector zorgt er namelijk voor dat concretere uitspraken mogelijk zijn. Zo kan de steekproef bijvoorbeeld beperkt worden tot of nadien gecontroleerd worden op hoogopgeleide jobs of jobs in een bepaalde sector. Ook kan het type contract (bepaalde of onbepaalde duur) worden gefilterd. Omwille van het lagere aantal vacatures in steden of gemeenten is het minder evident om tot dezelfde fijnmazige conclusies te komen met statistische zekerheid.

Cv-sjablonen

Bij het selecteren van de cv-sjablonen staan twee zaken centraal: **variabiliteit** en **werkbaarheid**. Omwille van detectie is het belangrijk dat niet alle cv's op elkaar lijken. Microsoft Word heeft tegenwoordig al een hele selectie aan cv-templates ingebouwd dewelke gebruikt kunnen worden. Deze kunnen als basis genomen worden met wat variatie in kleur, lettertype, en opbouw. Er zijn ook tal van voorbeelden te vinden online, maar hier valt de werkbaarheid soms tegen. Met werkbaarheid wordt bedoeld dat de bestanden makkelijk aan te passen moeten zijn. Zo kan het goed krijgen van de vormgeving in Word

soms veel tijd kosten. Daarom raden we aan om dus bij elk nieuw sjabloon te testen hoe makkelijk of moeilijk het is om dit aan te passen naar een nieuw profiel. Zijn er 'lijntjes' of andere elementen die continu verspringen? Zijn er verborgen elementen in het document die de hele opmaak in de war sturen? Kunnen zaken zoals extra ervaring makkelijk worden toegevoegd of worden verwijderd? Dikwijls is het pas bij het uitproberen dat deze zaken duidelijk worden.

Eens een goede selectie is gemaakt, raden we aan om deze te nummeren om het randomiseren te vergemakkelijken. Het is moeilijk om een getal te plakken op het aantal benodigde sjablonen. In de mate van het mogelijke wordt een specifiek sjabloon slechts vier keer voor dezelfde selectieverantwoordelijke gespreid over de looptijd van het experiment gebruikt.³⁴

Lukrake toekenning

Het lukraak toekennen of randomiseren van elementen aan de profielen is cruciaal om geen **systematische fouten** (*bias*) te creëren.³⁵ Het idee hier is dat het onmogelijk is exact dezelfde informatie en indruk over te brengen via twee cv's. Vooral als deze genoeg moeten verschillen om niet gedetecteerd te worden. **Onbekende aspecten** kunnen een werkgever echter motiveren om een bepaalde kandidaat te verkiezen. Bijvoorbeeld iemand ziet zijn alma mater op een cv, of een bedrijf waarmee ze vertrouwd zijn en verkiest dit profiel (al dan niet onbewust). In principe is dit geen probleem als er genoeg willekeur is ingebouwd in het onderzoeksdesign. Er wordt dus bijvoorbeeld nooit steeds een profiel met opleiding aan de KU Leuven gekoppeld aan een vrouw en een opleiding aan de Universiteit Gent aan dat van een man. *Wel* wordt er afgewisseld zodat er gemiddeld genomen ongeveer evenveel mannen als vrouwen aan zowel de KU Leuven als de Universiteit Gent afstuderen. Dit balanceert deze voorkeuren van de werkgevers - voorkeuren die zich anders zouden kunnen vertalen in schijnbare discriminatie gekoppeld aan geslacht.

³⁴ Via Microsoft Word en wat online zoeken is het relatief makkelijk om 30 à 40 sjablonen te creëren. Het minimum van 200 vacatures in acht nemend bij duo-testen is het dus mogelijk om zelfs bij besturen met een erg kleine groep selectieverantwoordelijken (bv. vier personen) het aantal sjablonen te beperken tot 3 à 4 per persoon (400 profielen/4 selectieverantwoordelijken/30 sjablonen).

³⁵ Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991–1013.

Afwisseling in de profielen is dus belangrijk. Ook is het belangrijk dat de details van de profielen **gelijkwaardig** zijn bij het solliciteren voor eenzelfde vacature. Zo moeten opleidingen en werkervaringen gelijkwaardig blijven (tenzij we verschillen in deze elementen net willen testen). Binnen eenzelfde vacature wordt het niveau van de opleiding constant gehouden tussen de profielen. Ook de opleidingen zelf moeten gelijkwaardig zijn in hun relevantie voor de vacature. Voor een functie van jurist vergelijken we bijvoorbeeld geen profiel van master in IT met een profiel van master in de rechten. Een master in de rechten is namelijk veel relevanter voor deze vacature, en de verschillende reacties op deze sollicitaties kunnen dus toe te schrijven zijn aan de divergente opleiding eerder dan aan discriminatie. Hetzelfde geldt voor werkervaring en andere details van het cv, zoals hobby's, vrijwilligerswerk, of vaardigheden. Indien deze zaken binnen dezelfde vacature te sterk verschillen tussen profielen, dan kan ongelijke behandeling van de kandidaten niet uitsluitend worden toegewezen aan de discriminatiegronden (bv. geslacht of leeftijd). Het gelijkwaardig houden van de verschillende elementen van het cv en de profielen is dus de boodschap.

Hoewel elk detail van een profiel op basis van toeval kan worden gegenereerd (bv. woonplaats, opleiding, hobby's, etc.), is het voornamelijk van belang om dit te controleren met het oog op de discriminatiegronden die aan de basis liggen van het experiment. De meest voorkomende worden hier kort opgelijst. Beginnende bij het **cv-sjabloon** is het belangrijk dat deze lukraak worden toegewezen om te voorkomen dat de vormgeving, kleurschema's, of zelfs lettertypes een invloed kunnen uitoefenen op het oordeel van de werkgever. Is het doel van de test om bijvoorbeeld discriminatie tussen vrouwen en mannen te meten? Dan kan het nuttig zijn om de sjablonen te nummeren om tijdens de testen de spreiding over de twee groepen in het oog te houden.

Ook de **tijd en volgorde** zijn belangrijke aspecten. Dit heeft betrekking op een aantal punten. Zo is er de volgorde van solliciteren waarbij het een goede gewoonte is om af te wisselen in functie van de te testen discriminatiegronden. Er wordt dus niet steeds eerst het cv van het mannelijke en nadien dat van het vrouwelijke profiel verstuurt, maar er wordt afgewisseld. Ook is het aan te raden om te variëren doorheen de dagen. Solliciteer niet steeds op een maandag en dinsdag, maar verdeel dit over alle dagen van de week.

Dit heeft ook betrekking tot de **demografische gegevens**. In de eerste plaats gaat het hier om het lukraak toekennen van gegevens zoals woonplaats, leeftijd, opleiding, hobby's, en meer – telkens binnen de

grenzen van het steekproefkader. Indien bijvoorbeeld voor de leeftijdsgroep 32-46 jaar werd gekozen, dan moet binnen deze groep worden gerandomiseerd.

Technische aspecten

Het toekennen van data op basis van toeval kan dus op verschillende manieren. Indien de beschikbare elementen beperkt zijn, zoals bijvoorbeeld geslacht of etnische afkomst, dan is de meest eenvoudige oplossing om telkens **af te wisselen** (bv. vacature 1: solliciteren met Marokkaanse naam, nadien met Belgische naam; vacature 2: solliciteren met Belgische naam, nadien met Marokkaanse naam). Echter, zaken zoals adressen zijn moeilijker te alterneren omwille van de grote hoeveelheid gegevens die wordt gehanteerd. Vooral bij correspondentietesten in de openbare sector waar detectie een groter risico vormt omdat een beperkter aantal mensen meer fictieve cv's te zien krijgt, wordt het belangrijk om een goed systeem op te zetten om echt te **randomiseren**. De gemakkelijkste en meest toegankelijke manier is om te werken met een **Excel-file** puur gericht op het genereren van informatie om profielen mee op te stellen. Voor het **randomiseren van cijfers** kan de RANDBETWEEN-functie worden gebruikt:

`=RANDBETWEEN(X,Y)`

X staat hier voor de laagste waarde, bijvoorbeeld 0, en Y staat voor de hoogste. Dit kan handig zijn bij het genereren van huisnummers of cv-sjablonen. Ook voor het genereren van willekeurige verjaardagen is deze functie interessant. Zo zijn data (bv. 04/09/1980) in Excel eigenlijk gewoon unieke nummers. De jongste en oudste verjaardagen worden in twee cellen gezet (bv. A3 en A4). Vervolgens wordt de RANDBETWEEN-functie gebruikt met deze twee cellen als een boven- en ondergrens (Y en X in de bovenstaande functie). Dan dient het formaat van de cel met RANDBETWEEN-functie nog te worden aangepast om een datum weer te geven (rechterklik op de cel en selecteer *cell format* en vervolgens het gewenste datum formaat).

Voor het **randomiseren van tekst-data**, zoals een lijst met namen, scholen, of straatnamen kan deze functie worden gecombineerd met de INDEX-functie en de COUNTA-functie:

`=INDEX(data,RANDBETWEEN(1,COUNTA(data)),1)`

Het woordje “data” staat in deze formule voor aan data-range. Dit kan bijvoorbeeld een kolom zijn waar onze te randomiseren namen staan opgelijst (bv. A3:A86). Aan de rechterkant staat de COUNTA-functie. Deze telt het aantal cellen waar effectief ook data in staan, ongeacht of deze tekst, cijfers, of iets anders zijn. Deze functie berekent op de achtergrond (dus zonder dat dit zichtbaar is) een cijfer waar vervolgens de RANDBETWEEN-functie op dezelfde manier als hierboven van gebruik maakt: deze selecteert willekeurig een cijfer tussen 1 en het cijfer dat als resultaat van de COUNTA-functie werd gegenereerd. Op de achtergrond komt deze RANDBETWEEN-functie dus terug met een willekeurig cijfer dat overeenkomt met een bepaalde cel (een locatie) in het geheel van data waar onze gegevens staan (bv. cel A18 in de range A3:A86). Als laatste stap is er de INDEX-functie. Deze wordt slechts beperkt gebruikt: “data” staat voor diezelfde data-range (bv. de namenlijst), vervolgens is er een cijfer dat werd gegenereerd dat de rij-positie binnen de data-range aanduidt. Het derde argument in de formule (1) is de kolom-positie. Deze formules samen vormen een makkelijke manier om stukken tekst, zoals straatnamen of gemeentes uit een lijst te randomiseren via Excel.³⁶

Bij sollicitaties in de openbare sector is het soms ook verplicht om een **rijksregisternummer** in te geven. Hier wordt best zo realistisch mogelijk te werk gegaan zonder dat echte personen betrokken worden bij de testen. Een rijksregisternummer bevat 11 cijfers die drie groepen omvatten:³⁷

- 1) De eerste groep van zes cijfers omvat de **geboortedatum** in de volgorde van jaar, maand, en dag (**xx.xx.xx-yyy.zz**). Dit moet overeenkomen met de geboortedag van de fictieve kandidaat (bv. het rijksregisternummer van iemand geboren op 18 mei 1972 heeft de vorm **72.05.18-yyy.zz**).
- 2) Vervolgens is er een tweede groep van drie cijfers, de **dagteller**. (**xx.xx.xx-yyy.zz**). Dit nummer is even voor een vrouw en oneven voor een man en gaat van 001 tot en met 998. Hier wordt ook realistisch te werk gegaan door een willekeurig cijfer te nemen in de correcte vorm afhankelijk van het geslacht. Natuurlijk is er geen echte dagteller voor onze fictieve kandidaat aangezien deze nooit geboren is. Desondanks kan dit deel wel door toeval overeenkomen met een echt

³⁶ Via het Centraal Referentieadressenbestand (CRAB) dat wordt onderhouden door de overheid, kan een volledige lijst van gemeenten en straten worden bekomen.

³⁷ https://www.ibz.rn.fgov.be/fileadmin/user_upload/nl/rr/instructies/IT-lijst/IT000_Rijksregisternummer.pdf (geraadpleegd 12 januari, 2022)

persoon. Dit deel op basis van toeval genereren is dus niet voldoende om fictieve rijksregisternummers te bekomen.

- 3) De derde groep is een **controlegetal** dat uit twee cijfers bestaat ($xx.xx.xx-yyy.zz$). Dit getal wordt berekend door de aaneenschakeling van de eerste twee groepen, in totaal negen cijfers, en modulo-rekenen ($xx.xx.xx-yyy(mod\ 97)$). De rest van deze deling wordt vervolgens afgetrokken van 97. Voor personen die geboren zijn vanaf 2000 moet hier nog een “2” voor het getal van negen cijfers worden geplaatst alvorens het te delen door 97.

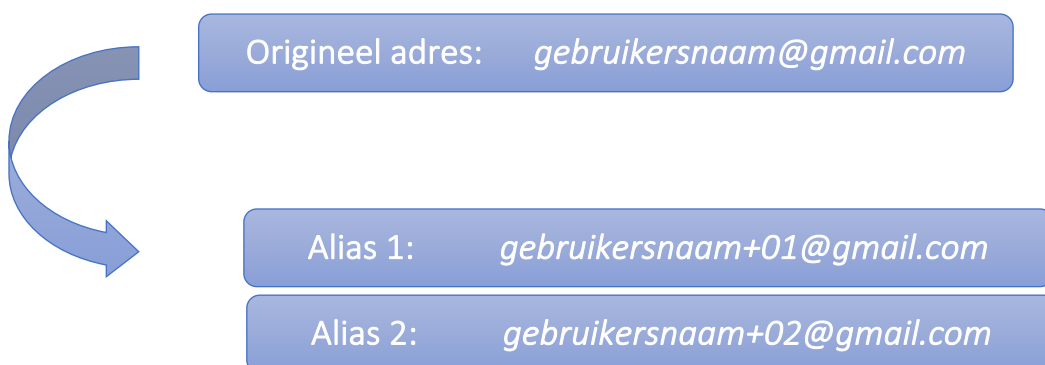
Dit controlegetal is het deel van het rijksregisternummer dat wordt aangepast om volledig fictieve cijfers te creëren. Indien het controlegetal afwijkt van het resultaat dat de officiële berekening zou bekomen op basis van de modulus 97, dan kan dit niet overeenkomen met een bestaand rijksregisternummer. Om dit controlegetal ongedaan te maken of te vervormen, wordt 97 simpelweg vervangen door een ander getal zoals bv. 34. Ook al zou door toeval de geboortedatum ($xx.xx.xx-yyy.zz$) en de dagteller ($xx.xx.xx-yyy.zz$) wel toebehoren aan een echt persoon, de laatste twee cijfers zullen verschillen. Zo wordt dus vermeden dat bestaande rijksregisternummers, en bijgevolg echte personen, per toeval betrokken worden bij fictieve sollicitaties.

We kiezen om op het controlegetal in te werken omdat dit deel van het rijksregisternummer minder zichtbaar is voor mensen die de gegevens van onze kandidaten verwerken. Als de eerste groep niet overeenkomt met de geboortedag, is dit snel te zien met het blote oog. Iemand die rijksregisternummers een beetje kent zal misschien ook op de eerste oogopslag een even of oneven getal zien in de dagteller en dus het geslacht herkennen. Het controlegetal is echter complexer om te berekenen. Het is onwaarschijnlijk dat dit zomaar visueel, zonder diepgaande controle kan worden gedetecteerd. Het genereren van rijksregisternummers kan ook weer makkelijk via Excel worden geautomatiseerd door elke van deze groepen van cijfers te genereren en vervolgens te combineren in het gewenste formaat.³⁸

³⁸ Excel heeft een modulo (=MOD()) functie die deze berekening makkelijk kan uitvoeren. De echte berekening van de laatste twee cijfers van het rijksregisternummer zou dan op de volgende manier worden weergegeven in excel: =MOD(cel met de aaneenschakeling van de eerste negen cijfers, 97). Vervolgens dient dit cijfer afgetrokken te worden van 97.

Bij de meeste sollicitaties is het nodig een e-mailadres en gsm-nummer op te geven. **E-mailadressen** kunnen gratis worden aangemaakt bij grote providers zoals Google of Microsoft. Alternatieven zijn natuurlijk ook mogelijk, maar deze zijn dikwijls gebundeld aan een klantendossier (bv. Telenet of iCloud van Apple) of zijn slechts mogelijk tegen betaling. Ook is het cruciaal detectie te vermijden. Zo zou het opzetten van een eigen e-mailserver met e-mailadressen snel argwaan wekken indien er na het @-tekentje consistent dezelfde specifieke woorden staan. Hoewel grote providers zoals Google en Microsoft hier dus oplossingen bieden, zijn er de laatste jaren wel verschillende hordes bijgekomen omwille van o.a. gegevensbescherming. Deze hordes hebben voornamelijk betrekking op situaties waarbij er tientallen of honderden adressen moeten worden aangemaakt. Zo is bij Google (*gmail*-adressen) nagenoeg altijd een **gsm-nummer** nodig om de aanmaak van een account te verifiëren. Elk nummer kan telkens ook maar een handvol keer worden gebruikt. Online-diensten die tijdelijk nummers aanbieden om sms-codes te ontvangen werken niet. Wie goed zoekt online kan soms wel manieren vinden om de gsm-verificatie te omzeilen, maar Google sluit deze *loopholes* erg snel.

Gmail biedt wel de mogelijkheid om 30 **aliassen** te maken wat van pas kan komen als de e-mailadressen niet echt zichtbaar zijn. Aliassen zijn alternatieve vormen van hetzelfde adres die gekoppeld worden aan het originele adres en mailbox. Het is dus niet nodig om telkens een nieuwe account te maken om een uniek e-mailadres te bekomen. Wel is de vorm van dit alias-adres niet regulier dus is enige voorzichtigheid omwille van detectie noodzakelijk. Concreet werkt Gmail met het plusteken (+). Alles dat na het plusteken maar voor het @ teken komt, telt in principe niet mee en wordt dus verbonden aan het originele adres. De gebruiker of een systeem waarin wordt gesolliciteerd ziet dus een nieuw e-mailadres, maar Google beschouwt het als het originele:



Het is tevens relatief eenvoudig om een **autoforward** in te stellen in Gmail, iets dat bij Outlook of Hotmail bij beginnende accounts niet werkt. Dit kan worden omzeild door via *rules* te werken en simpelweg een regel aan te maken die alle inkomende berichten doorstuurt naar een algemeen verzameladres. De diensten van Microsoft zijn ook meer toegankelijk als er vele adressen tegelijk moeten worden aangemaakt. Wel is hier een **VPN** nodig omdat nieuwe accounts op het moment van schrijven beperkt zijn tot twee per IP-adres. Na het aanmaken van twee nieuwe Outlook of Hotmail-accounts moet dus via de VPN telkens een nieuw IP-adres worden gecreëerd. Tot twee aliassen aanmaken per keer is ook mogelijk bij Outlook, maar niet bij Hotmailadressen. Ook kan het gebeuren dat er toch gsm-verificatie nodig is wanneer men voor een tweede of derde keer inlogt. Wanneer dit gebeurt is moeilijk te voorspellen. Zowel Google als Microsoft doen dit wanneer zij “iets verdachts vaststellen”. Deze verificatie is echter minder strikt dan die van Google. Eenzelfde nummer kan voor meerdere bevestigingen worden gebruikt.

Gmail	Outlook	Hotmail
<ul style="list-style-type: none">• aanmaak met gsm-nummer• soms <i>loopholes</i> bruikbaar, maar worden snel gesloten.• veel aliassen mogelijk• eenvoudige autoforward	<ul style="list-style-type: none">• aanmaak zonder gsm-nummer• twee accounts per IP-adres• soms na aanmaak verificatie met gsm nodig• aliassen beperkt mogelijk• autoforward via ‘rules’	<ul style="list-style-type: none">• aanmaak zonder gsm-nummer• twee accounts per IP-adres• soms na aanmaak verificatie met gsm nodig• geen aliassen mogelijk• autoforward via ‘rules’

Zoals reeds aangehaald, wordt er ook een **autoforward-adres** aangemaakt waar al de andere mailboxen naar verwijzen. Het continue inloggen en controleren van tientallen mailboxen per week zou ons te veel tijd kosten. Door al het verkeer door te verwijzen naar één mailbox, is het makkelijker het overzicht te bewaren.³⁹

³⁹ Voor meer technische details over het gebruik van e-mails, zie Crabtree C (2018) An introduction to conducting email audit studies. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer, pp.103–117.

De aankoop van **GSM-nummers** is beperkt tot tien per persoon per provider. Gemiddeld genomen rekenen providers 10 à 15 euro aan per simkaart. Indien de onderzoeker overtuigd is dat de nummers niet worden gebruikt, kan worden geopteerd voor fictieve nummers. De echte nummers worden aan een naam gekoppeld, en er wordt een voicemail opgezet. De *callbacks* van selectieverantwoordelijken of werkgevers worden nooit rechtstreeks beantwoord. Indien er een boodschap wordt ingesproken, wordt er per email gereageerd om het proces af te breken.

Correspondentietesten in de openbare sector eisen een grote variatie in cv's en profielen. Soms kan het dan ook moeilijk zijn om de nodige informatie te verschaffen om geloofwaardig te solliciteren. Dat is een probleem dat zich voornamelijk voordoet bij specialistische vacatures. Een goede bron van inspiratie is **LinkedIn** waarin via zoektermen gezocht kan worden naar mensen met gelijkaardige werkervaring. Hier zijn echter wel beperkingen aan verbonden. Zo is a) voor veelvuldig zoeken een *premium*-account nodig wat de kosten gepaard gaande met het experiment verhoogt. Ook b) laat LinkedIn voornamelijk *network*-based zoekacties toe. Dit wil zeggen dat deze zoekacties via een volledig nieuwe account zonder connecties veel minder zoekresultaten zullen opleveren. Idealiter wordt dus een account gebruikt van één van de onderzoekers of medewerkers die al enige jaren actief is. Een betere optie is om een algemene account van de stad of gemeente bij te houden die doorheen de jaren een netwerk uitbouwt. Dit account kan dan voor correspondentietesten worden gebruikt. Hier moet wel worden gelet dat de mensen die getest worden (bv. selectieverantwoordelijken) geen rechtstreekse toegang hebben. LinkedIn houdt namelijk de zoektermen bij en een zoekgeschiedenis overeenkomstig met de open vacatures zou snel tot detectie leiden.

Een andere mogelijkheid is om online andere vacatures te raadplegen voor **gelijkaardige openingen** (bijvoorbeeld in de private sector; al dan niet op LinkedIn). Deze vacatures gunnen de onderzoeker een beter inzicht in concrete taken en andere aspecten van de functie die kunnen worden gebruikt bij het invullen van het fictieve cv of het beantwoorden van inhoudelijke vragen.

Een expert in Excel zal misschien snel tot de verleiding komen om zowel het genereren van profielen en demografische informatie, als het registreren samen te voegen in één groot bestand, of verschillende bestanden automatisch te linken. Dit raden we af. Het zou namelijk niet de eerste keer zijn dat omwille van technische moeilijkheden (zoals *macro errors*) bestanden plots onbruikbaar worden. Als dit tijdens

een experiment gebeurt, kan er zodoende veel tijd verloren gaan in het zoeken naar de meest recente werkende back-up bestanden of het proberen te repareren van het bestand. Het **genereren en registreren van data wordt best gesplitst** en de data wordt best manueel van het eerste naar het tweede bestand gekopieerd. Dit tweede registratie-bestand (en vooral de registratie-*sheet*) blijft best zo licht en simpel mogelijk zodat dit altijd bruikbaar blijft.

Fase 2: Uitvoering

Opsturen van fictieve sollicitaties

Op de private arbeidsmarkt gebeurt het opsturen van vacatures bij correspondentietesten vandaag de dag meestal via e-mail. Hoewel we zien dat in sommige gemeentes en steden er kan worden gesolliciteerd via e-mail, merken we toch dat dit een uitzondering is. Veelal zijn er **online formulieren**, vragenlijsten, en profielen die moeten worden opgesteld. Dit vergt natuurlijk meer voorbereidend werk. Zoals reeds uitgebreid besproken in het onderdeel 'lukrake toekenning' is het cruciaal dat het verzenden van sollicitaties **geen merkbaar patroon** volgt. Indien de onderzoeker er de gewoonte op nahoudt om elke dag om 9u te solliciteren, dan kan dit beginnen opvallen. Vooral als er nog andere elementen zijn in het cv die minder standaard zijn in vergelijking met de gemiddelde echte kandidaat, dan kan dit sneller leiden tot detectie. Ook met betrekking tot de te testen discriminatiegrond is het belangrijk de **volgorde af te wisselen** zodat niet altijd de kandidaat met een Belgische naam eerst solliciteert. Indien de selectieverantwoordelijke een voorkeur heeft voor vroege vogels, kan dit namelijk schijnbare discriminatie tot gevolg hebben omdat de Belgische (vroege) kandidaat sneller tot de tweede ronde wordt uitgenodigd.

Solliciteren gebeurt normaal gezien in eenzelfde periode, met **24 tot 48 uur** tussen de fictieve sollicitanten. Veel hangt af van het soort vacature. Is het een erg specialistische vacature, dan is het beter wat langer te wachten tussen het solliciteren van de eerste en tweede kandidaat. Het is namelijk mogelijk dat de selectieverantwoordelijke slechts een aantal cv's binnenkrijgt. Door de opzet van correspondentietesten zijn de cv's van de fictieve kandidaten gelijkwaardig. Dus het ontvangen van twee gelijkwaardige (zelfde mate van kwaliteit en detail) cv's in opeenvolgende dagen terwijl er anders amper respons is, verhoogt weer de kans op detectie. In de publieke sector wordt er in de vacature dikwijls een uiterste sollicitatiedatum gespecificeerd. Staat er iets zoals "solliciteren kan tot en met [datum]" dan is een lagere frequentie binnen deze tijdsperiode mogelijk (bv. 48u tussen de twee

sollicitaties). Is er géén duidelijke looptijd, dan kan het zijn dat de werkgever al tot een beslissing komt meteen na het lezen van het cv van onze eerste sollicitant. Wordt er dan te lang gewacht, dan is het mogelijk dat de vacature informeel reeds is gesloten. Zolang wordt afgewisseld met de profielen heeft dit echter geen meetbaar effect.

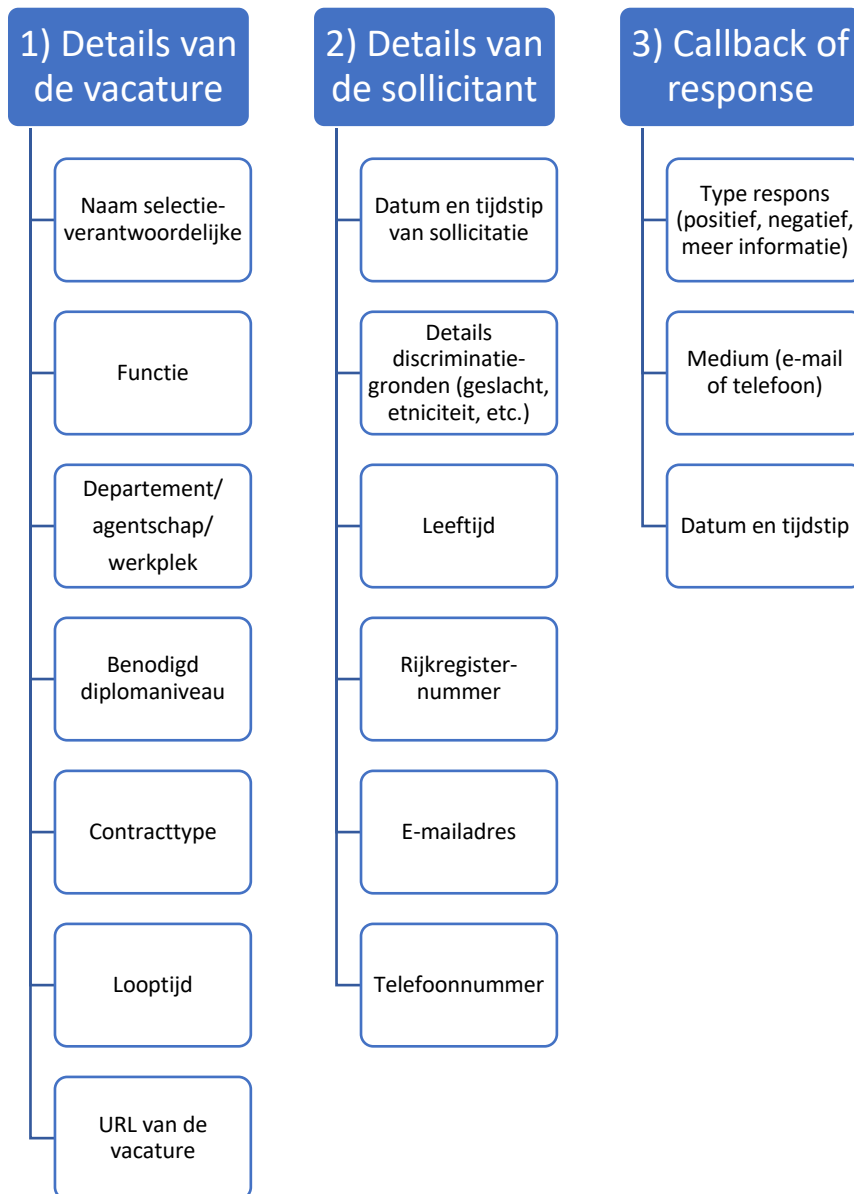
Registratie van sollicitaties en reacties

Er wordt aangeraden om alle informatie van de vacature en sollicitatie te registreren omwille van verschillende redenen. Vooreerst is het cruciaal om tijdens de **analyse** toegang te hebben tot al de relevante variabelen. Misschien is er een patroon van discriminatie dat alleen zichtbaar is als de laaggeschoolde profielen of contracten van bepaalde duur worden bekeken. Ook registreert de onderzoeker telkens de data en tijdstippen. Het is aangewezen om in deze fase zo nauwkeurig mogelijk te werken, en nadien te abstraheren indien nodig. Bijvoorbeeld, werk niet met “voormiddag” of “namiddag” maar met exactere maten zoals 11u42 of 18u22. Bij de meeste elementen wordt er niet actief op zoek gegaan naar discriminatie. Met statistische software is het echter eenvoudig dit wetenschappelijk te controleren zolang de data voorhanden is. Wordt bepaalde informatie niet geregistreerd, dan kan er nooit worden uitgesloten dat deze elementen geen effect hebben gehad op de resultaten van het experiment.

Ook komt deze manier van registreren de **efficiëntie** ten goede. Als het geweten is welke selectieverantwoordelijke (onlangs) welke cv's heeft gezien, dan kunnen deze cv's mits kleine aanpassingen (bv. demografische gegevens) gerecycleerd worden bij vacatures van andere selectieverantwoordelijken. Dit recyclen kan de benodigde **tijd halveren** in vergelijking met het telkens opnieuw creëren van profielen. Ook kunnen zo eventuele bronnen van detectie beter in het oog worden gehouden. Randomiseren en alterneren heeft altijd de voorkeur, maar soms zorgt dit ongewild toch voor patronen. Zo kan bijvoorbeeld selectieverantwoordelijke X telkens met een profiel met handicap Y worden geconfronteerd. Indien dit profiel onder echte kandidaten zo vaak niet voorkomt, kan dit argwaan wekken en de kans op detectie verhogen. Dit soort controles kunnen in principe worden ingebouwd met functies en formules in Excel, maar soms is het makkelijker en sneller om dit visueel en manueel af te toetsen. Complexiteit wordt voorbehouden voor de zaken waar dit nuttig is.

Bij correspondentietesten in de private sector kunnen een aantal zaken zoals demografische gegevens worden verbonden aan een beperkt aantal cv-profielen. Zo kan een profiel telkens dezelfde bijkomstige

informatie bevatten (opleiding, werkervaring, hobby's, adres). Dit profiel kan vervolgens willekeurig aan een cv-sjabloon en de te testen discriminatiegronden worden toegewezen. Dit zorgt ervoor dat de gegevens die bijkomstig zijn niet telkens moeten gegenereerd worden. Voor correspondentietesten in de openbare sector worden echter honderden verschillende sets van gegevens gebruikt wat het moeilijk maakt om met cv-profielen te werken. Om het onderzoek dus navolgbaar te maken, worden de volgende zaken onderverdeeld in drie groepen geregistreerd:



1) **Details van de vacature:** Hier geldt dat de details geregistreerd worden, ook al vormen ze misschien niet het centrale onderdeel van het onderzoek. De naam van de **selectieverantwoordelijke** is, indien bekend, in de eerste plaats belangrijk met het oog op

detectie. Eenzelfde persoon mag namelijk niet overstelpt worden met fictieve sollicitaties in een te korte periode. Ook wordt er gevarieerd in de details die ze te zien krijgen via de sollicitatie (werkervaring, cv-sjabloon, etc.). Vervolgens wordt de naam van de **functie**, de **werkgever**, het benodigde **diplomaniveau**, **contracttype**, de **looptijd** van de vacature, en uiteindelijk nog de URL van de **vacature** geregistreerd. Deze informatie wordt enerzijds geregistreerd om nadien mee op te nemen in de analyses indien de steekproef voldoende groot is (bv. controleren voor contracten onbepaalde en bepaalde duur). Anderzijds is het belangrijk om de efficiëntie zo hoog mogelijk te houden tijdens het testen zelf, en te recyclen wat mogelijk is. Indien vacatures voor gelijkaardige functies op een later moment of onder een ander departement en selectieverantwoordelijk terugkomen, dan kunnen deze hergebruikt worden mits kleine aanpassingen (bv. demografische gegevens, werkervaring een beetje aanpassen, etc.).

- 2) **Details van de sollicitant:** In het volgende deel van het Excel-bestand wordt alle relevante data van onze fictieve sollicitanten geregistreerd, beginnende met de **datum en tijd** waarop ze solliciteren. Ook worden de geteste **discriminatiegronden**, de **leeftijd** (dit kan natuurlijk ook een discriminatiegrond zijn) en het fictieve **rijksregister** nummer geregistreerd. Dit laatste is belangrijk om na het experiment de officiële statistieken van de stad of gemeente te corrigeren. Bij een intern officieel overzicht van sollicitanten moeten de fictieve sollicitanten namelijk worden verwijderd. Als deze intern geregistreerd blijven, dan zorgen de fictieve sollicitaties voor een vertekend beeld (nl. proportioneel meer kandidaten met de te testen discriminatiegronden). Vervolgens wordt ook het **e-mailadres** en het gebruikte **gsm-nummer** geregistreerd. Dit is cruciaal om te weten welke adressen en nummers reeds zijn gebruikt. Optioneel kunnen ook andere zaken worden vastgelegd zoals opleiding, werkervaring of woonadres. Deze kunnen nuttig zijn om detectie te vermijden of verdere analyses te doen na het experiment.
- 3) **Callback of response:** In dit deel wordt vastgelegd wat de reactie is van de werkgever/selectieverantwoordelijke op de sollicitatie. Het belangrijkste aspect is hier het type respons. Analoog met de wetenschappelijke literatuur kan er een onderverdeling worden gemaakt binnen de **positieve respons**. Een positieve respons in ruime zin wordt gedefinieerd als om het even welke positieve reactie (uitnodiging job gesprek, voorstel voor andere positie, vraag om meer informatie). Een respons in enge zin is een concrete uitnodiging voor de volgende ronde (testen, interviews). Een **negatieve respons** kan zowel bestaan uit een negatieve email of

telefoon (bv. niet voldoen aan de vereisten) als het uitblijven van een reactie. Idealiter speciëren we in het Excel bestand door middel van een code om wat voor soort positieve of negatieve reactie het gaat (bv. via cijfers: 1: positieve eng, 2: positief ruim, 3: geen reactie, 4: negatieve reactie). Tot slot wordt het **medium, datum, en tijdstip** genoteerd. Deze laatste zijn interessant omdat voorgaand onderzoek structurele verschillen heeft vastgesteld in hoe snel groepen antwoord krijgen.⁴⁰ Indien deze informatie voorhanden is, kan daarop gecontroleerd worden.

Afhankelijk van het onderzoeksontwerp en de beschikbare informatie in de vacatures kan deze lijst er anders uitzien. Idealiter wordt dit alles horizontaal vastgelegd in een Excel-bestand (één rij = één sollicitant) beginnende met een uniek ID voor ieder sollicitant (bv. 00001, 00002, etc.). Een alternatieve methode die omslachtiger is indien met triotesten wordt gewerkt, is de verschillende sollicitanten in één rij zetten (dus één rij is één vacature). Ook worden er enkele kolommen toegevoegd ter referentie. Zo kan worden bijgehouden of de uitnodiging voor de volgende ronde al is beantwoord. Ook kunnen er een paar controles worden ingebouwd door bijvoorbeeld de namenlijst of e-maillijst te linken aan dit bestand. Via een functie in *conditional formatting* kan de cel met de naam of het e-mailadres rood worden gemaakt indien dit al gebruikt was bij een vorige vacature.

`=COUNTIF(data range,cel met e-mail of naam)>0`

Als laatste (of eerste) stap, wordt er ook nog een kopie van de vacature gemaakt ter archivering. Dit kan via *screenshots* maar ook door gewoon de website te printen in pdf via de webbrowser, en deze pdf samen met de cv's voor die vacature op te slaan in een folder.

⁴⁰ Kaas, L., & Manger, C. (2011). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1–20; Duguet, Parquet, L'horty, & Petit. (2015). New Evidence of Ethnic and Gender Discriminations in the French Labor Market Using Experimental Data: A Ranking Extension of Responses from Correspondence Tests. *Annals of Economics and Statistics*, 117/118, 21.

Fase 3: Analyse

Na het afronden van de dataverzameling van het experiment, rest ons nog de analyse. Er zijn verschillende manieren om de resultaten van het experiment uit te drukken. We overlopen hier de meest voorkomende.

Nettodiscriminatiegraden

Nettodiscriminatiegraden worden in de internationale wetenschappelijke literatuur frequent gebruikt om de mate van ongelijke behandeling uit te drukken en te analyseren. Deze wordt als volgt berekend:

$$\text{Nettodiscriminatiegraad} = \frac{Obs_{i1} - Obs_{j1}}{Obs_{ij1}}$$

Obs_{i1} het aantal observaties waarbij enkel de meerderheidskandidaat een gunstige uitkomst kent

Obs_{j1} het aantal observaties waarbij enkel de minderheidskandidaat een gunstige uitkomst kent

Obs_{ij1} het aantal observaties waarbij minstens één van beiden een gunstige uitkomst kent

Deze formule kan worden toegepast op elke vorm van positieve reactie (enge of ruime zin). Concreet specificeert deze maat het risico voor bepaalde groepen om gediscrimineerd te worden rekening houdend met een aantal zaken. Zo houdt deze formule alleen rekening met de vacatures waar minstens één van de kandidaten een reactie kreeg. Vacatures waar beide kandidaten een negatieve reactie krijgen worden dus buiten beschouwing gelaten omdat de achterliggende reden hiervoor niet gekend is. We weten dus ook niet of hier discriminatie aan- of afwezig is. Ook is dit een formule om een *netto*-graad te berekenen. Indien er vacatures zijn waar enkel de testpersonen (de sollicitanten die toebehoren aan een minderheidsgroep) een positieve reactie krijgen, dan is dit te wijten aan een toevallige ongelijke behandeling. Dit toeval zou in dezelfde mate moeten voorkomen bij zowel de test- als de controlegroep. Als er sprake is van positieve discriminatie ten voordele van de minderheidsgroep dan onderschat deze maat echter de totale discriminatie. Dit wil zeggen dat situaties van toevallige ongelijke behandeling (ten voordele/nadele van bepaalde groepen) worden uitgebalanceerd.⁴¹

⁴¹ In het geval van positieve discriminatie wordt de formule $(Obs_{j1} - Obs_{i1}) / Obs_{ij1}$. De interpretatie blijft gelijk maar is van toepassing voor de meerderheidsgroep (bv. bij 33% van de vacatures werden kandidaten van de meerderheidsgroep gediscrimineerd).

Het uitdrukken van deze maat gebeurt in percentages. Een nettodiscriminatiegraad van 33%, wil dan zeggen dat **de kandidaten van de minderheidsgroep in 33% van de vacatures waarbij minstens één kandidaat een positieve reactie kreeg gediscrimineerd** werden. Dit wil zeggen dat kandidaten van de minderheidsgroep in ongeveer **1 op 3 van de vacatures waarbij minstens één van beide kandidaten werd uitgenodigd, discriminatie** onderging.

Uitnodigingskansen

Uitnodigingskansen drukken de **kans uit voor leden van een bepaalde groep** (bv. de test- of controlegroep) **om uitgenodigd te worden** voor een volgende ronde. Deze volgende ronde neemt meestal de vorm aan van interviews of testen (bv. online of bij een rekruteringsagentschap zoals Hudson). In tegenstelling tot de nettodiscriminatiegraad, is de uitnodigingskans dus specifiek voor een groep. Indien groepen met elkaar vergeleken worden, worden meerdere kansen vergeleken om inzicht te verwerven in welke groep sneller uitgenodigd werd.

De uitnodigingskans van een bepaalde groep wordt als volgt berekend: Het aantal uitnodigingen door de werkgever gedeeld door het totaal aantal sollicitanten van de specifieke groep. Als bijvoorbeeld 100 fictieve kandidaten met een Marokkaans klinkende naam solliciteerden en dit 22 positieve reacties⁴² opleverde, dan is de uitnodigingskans voor sollicitanten met een Marokkaans klinkende naam 22% (22/100).

$$p_{uj} = \frac{Obs_{j1}}{Obs_{totaal}} = \frac{Obs_{j1}}{Obs_{j1} + Obs_{j0}}$$

P_{uj}	de kans op uitnodiging voor groep j
Obs_{j1}	het aantal observaties waarbij de kandidaat van groep j wordt uitgenodigd
Obs_{totaal}	= $Obs_{j1} + Obs_{j0}$ of het totaal aantal observaties of de som van het aantal observaties waarbij de kandidaat van groep j wordt uitgenodigd, en het aantal observaties waarbij deze niet wordt uitgenodigd

⁴² De definitie van deze positieve reactie kan in lijn met het voorgaande variëren naargelang het soort reactie. Zo zou bijvoorbeeld de uitnodigingskans in ruime en enge zin berekend kunnen worden waarbij de kans in enge zin logischerwijs kleiner zal zijn dan die in ruime zin.

Uitnodigingskans wordt uitgedrukt in percentages. Zo kunnen bijvoorbeeld sollicitanten met een Marokkaanse naam uitgenodigd worden in 22% van de gevallen (onze testgroep), terwijl dit voor sollicitanten met een traditioneel 'Belgisch' klinkende naam bijvoorbeeld 25% is. Het **verschil in uitnodigingskansen** tussen beide groepen in **procentpunten** is 3% (25%-22%).

$$\Delta_{p_{uij}} = p_{ui} - p_{uj}$$

$\Delta_{p_{uij}}$	het verschil in kans op uitnodiging tussen groep i en j
p_{uj}	de kans op uitnodiging voor groep j
p_{ui}	de kans op uitnodiging voor groep i

Het verschil in **percentage** in ons voorbeeld is 12% en wordt als volgt berekend: De uitnodigingskans van de controlegroep (25%) verminderd met de uitnodigingskans van de testgroep (22%), gedeeld door de uitnodigingskans van de controlegroep (25%). Met andere woorden, in ons voorbeeld hebben de sollicitanten met een Marokkaanse naam **12% minder kans om uitgenodigd te worden** in vergelijking met kandidaten met een typisch 'Belgisch' klinkende naam. Anders geformuleerd worden sollicitanten met een Marokkaans klinkende naam **ongeveer een achtste minder vaak uitgenodigd**.

$$\Delta_{p_{uij}} = \frac{(p_{ui} - p_{uj})}{p_{ui}}$$

Logistische regressieanalyse

Regressieanalyses zijn ingewikkelder dan de voorgaande analysemethodes om discriminatie in kaart te brengen. De mathematische details en achtergrond omtrent regressieanalyses uit de doeken doen is niet het doel van deze methodiek. Wel wordt er hier kort stilgestaan bij de toepassing van deze techniek op correspondentietesten.

Vooreerst is het onderzoeksdesign bepalend voor het **type regressie**. Er kan een **simpele logistische regressie** op de kans op uitnodiging (de afhankelijke variabele) worden toegepast. Ook, mits voldoende data voorhanden zijn, kan er voor een **logistische multilevel regressieanalyse** worden gekozen. De vacatures (niveau 1) zijn zodoende ingebed in een hoger niveau zoals dat van selectieverantwoordelijke, een bepaalde dienst of gemeente (niveau 2).

Regressieanalyses zijn ook uitermate geschikt om diepgaande analyses uit te voeren. Zo kan snel en overzichtelijk gecontroleerd worden of de verschillen in discriminatie tussen twee groepen ook afhangt van andere factoren van de kandidaat, vacature (bv. functieniveau of sector) of het selectieproces (bv. dienst, afkomst of geslacht van de selectieverantwoordelijke). De discriminatiegrond waarvoor wordt getest – bijvoorbeeld gender – kan zo in **interactie** worden gebracht met deze andere kenmerken. Met andere woorden kan er achterhaald worden in welke mate er heterogeniteit zit in de gemeten discriminatie naar bepaalde kenmerken.

6. Sensibilisering

Correspondentietesten in het kader van arbeidsmarktdiscriminatie in de publieke sector verschaffen inzicht in discriminerend gedrag tijdens de eerste ronde van selectie. Als er na een nulmeting wordt vastgesteld dat er inderdaad een significant patroon is van discriminatie, dan kunnen lokale besturen ervoor opteren om hier dan ook iets aan te doen.

De wetenschappelijke literatuur is hier minder uitgebreid dan die met betrekking tot monitoring. Er onderscheiden zich twee paradigma's: langs de ene kant een model waar **afschrikking** (*deterrence*) centraal staat, en langs de andere kant **overtuiging** (*compliance*).⁴³ Hoewel de opdrachtgever hier de focus legt op sensibiliserende acties, levert een combinatie met handhaving de beste resultaten op. Bij sommigen kan een sensibiliserende aanpak goed werken terwijl anderen pas tot inzicht komen bij bestraffing. Omdat het onmogelijk is te weten te komen tot welke groep discriminerende agenten behoren, is een tweevoudige aanpak te verkiezen. Ook is een zekere logische volgorde van deze paradigma's belangrijk. De cognitieve neurowetenschap leert ons echter dat in het algemeen positieve, sensibiliserende acties beter werken dan een afschrikkende aanpak.⁴⁴ Het handhavingsbeleid dient dus een logische opbouw te volgen die begint met overtuigen en sensibiliseren en pas nadien overschakelt op bestraffen wanneer duidelijk wordt dat sensibiliseren niet langer werkt en wanneer er sprake is van discriminatie. Deze opdracht richt zich uitsluitend op het model van compliance of het geheel van **sensibiliserende acties**.

De focus in de **compliance-benadering** ligt in de eerste plaats op het verhogen van de naleving van regels door in te werken op het rationele of via het bewust maken van afwijkend gedrag. Het gaat met andere woorden om het **overtuigen** van mensen om quasi-automatisch de regelgeving na te leven. Hutter (1997) maakt een onderscheid binnen deze benadering: de a) **overtuigende** of b) **aandringende aanpak**.⁴⁵

⁴³ Zie ook Baldwin, R., Cave, M., & Lodge, M. (2012). *Understanding regulation: Theory, strategy, and practice* (2nd ed). (pp.238-243). Oxford University Press en Kramer, R. M., & Carnevale, P. J. (2003). Trust and Intergroup Negotiation. In *Blackwell Handbook of Social Psychology: Intergroup Processes* (pp. 431–450). Blackwell Publishing en Verstraete, J., Vermeir, D., De Decker, P., & Hubeau, B. (2017). *Een Vlaams antidiscriminatiebeleid op de private huurmarkt. De mogelijke rol van zelfregulering*. Leuven: Steunpunt Wonen.

⁴⁴ Alleen als bepaald gedrag echt moet worden afgeleerd heeft bestraffen ook zijn nut (Sharot, T. (2017). *The influential mind: What the brain reveals about our power to change others*. Henry Holt and Company).

⁴⁵ Hutter, B.M. (1997). *Compliance: Regulation and the Environment*. Oxford: Oxford University Press.

a) De overtuigende aanpak

In deze aanpak staat het **ondersteunen** van actoren met betrekking tot het naleven van de regelgeving centraal. Dit kan bijvoorbeeld in de vorm van **opleidingen** of het aanbieden van **advies**. De wetenschappelijke literatuur toont dat deze eerste aanpak zeker zijn vruchten kan afwerpen - vooral als er aan twee zaken voldaan wordt. Eerst is het cruciaal dat er een **wil tot verandering** is.⁴⁶ Vervolgens is het **gradueel beïnvloeden of omscholen** van de actor cruciaal. Stapsgewijs iemand overtuigen werkt dus beter dan geconfronteerd worden met een boodschap die lijnrecht tegenover de eigen houding staat.⁴⁷

We ontwaren een aantal soorten van interventies binnen deze overtuigende aanpak:

- a. **Communicatie van het onderzoek:** Communicatie kan zowel voor als na een nulmeting plaatsvinden – telkens met een ander functie. Zo heeft het aankondigen van (correspondentietest)en bij werving van lokale besturen de functie van **controle**. Zoals reeds aangehaald in fase 1 van het stappenplan (zie *hoofdstuk 5*) heeft communicatie gevolgen voor het experiment. Het aankondigen van testen kan ertoe leiden dat actoren zich gecontroleerd voelen en bewuster worden van hun gedrag, met gevolg minder discriminatie.⁴⁸ Na een meting met de resultaten in handen kan ook in gesprek worden gegaan met alle actoren die betrokken zijn bij de onderzochte processen op geaggregeerd niveau om zo eventuele vooroordelen en discriminerend gedrag te belichten (**bewustmaking**).

Omwille van de hoge mate van validiteit en betrouwbaarheid, kunnen correspondentietesten dus al een methode van sensibilisering zijn indien de metingen wetenschappelijk werden uitgevoerd en de conclusies voldoende sterk zijn. Vandaag de dag gebeurt het immers vaak dat actoren zich niet bewust zijn van de mate waarin ze discrimineren. Objectieve metingen kunnen zo verborgen structuren en patronen op lokaal geaggregeerd niveau blootleggen. Met andere woorden, monitoring, zoals in de vorm van

⁴⁶ Wood, W., & Quinn, J. M. (2003). Forewarned and forearmed? Two meta-analysis syntheses of forewarnings of influence appeals. *Psychological Bulletin*, 129(1), 119–138.

⁴⁷ Hogg, M. A., & Cooper, J. (Eds.). (2007). *The SAGE handbook of social psychology*. SAGE.

⁴⁸ Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, 64(11), 4988–4995.

correct uitgevoerde correspondentietesten, kan zelf een sterk sensibiliserende waarde hebben. Het eindrapport van het Centraal Expertencomité raadt ook aan om de nodige opvolgmetingen te ondernemen. Correspondentietesten in de publieke sector zijn echter onderworpen aan een aantal specifieke voorwaarden die de uitvoering in zekere mate omslachtiger en tijdsintensiever maken. Hierdoor zijn veelvuldige opvolgmetingen niet de meest praktische manier van sensibiliseren voor lokale besturen.

- b. **Training en opleiding:** Dit type overtuigende aanpak is relatief eenduidig. Het gaat hier om het opleiden en trainen van selectieverantwoordelijken, juryleden, en werkgevers met het oog op het vermijden van discriminerend handelen tijdens de aanwerving en op de werkvloer. De uitvoering van correspondentietesten gebeurt best in een **breder kader waarin er ook aandacht is voor structurele ondersteuning en begeleiding**. Het belang van het structurele aspect mag niet onderschat worden. Het is cruciaal om ten eerste inzicht te krijgen in de realiteit achter de cijfers van de correspondentietesten. Dit inzicht komt van pas bij de specifieke uitwerking van de begeleiding. Vervolgens is herhaaldelijke training nodig om niet te hervallen in het ongewenste gedrag. In veel gevallen is dit namelijk onbewust en is correctie dus iets waar continu aan moet gewerkt worden. Ten derde is er het aspect van personeelsverloop onder de werknemers die in het proces van aanwerving zitten. Van zodra nieuwe werknemers worden aangenomen, moet er een optie zijn tot begeleiding. Het loont niet te wachten tot een volgende monitoring vaststelt dat er problemen zijn. Preventief ingrijpen kan hier dus deel van uitmaken. Daarom pleiten we voor aanvullende positieve ondersteuning die kan worden voorzien zelfs zonder specifieke problemen met discriminerend gedrag.

Hier moet echter worden opgemerkt dat de huidige trend van *diversity training* in de private sector weinig wetenschappelijke onderbouwing heeft. Onderzoek leert ons dat de manier waarop deze trainingen meestal worden ingevuld weinig tot geen effect heeft.⁴⁹

⁴⁹ Dobbin, F., & Kalev, A. (2018). Why Doesn't Diversity Training Work? The Challenge for Industry and Academia. *Anthropology Now*, 10(2), 48–55.

b) De aandringende aanpak

De aandringende aanpak onderscheidt zich van de vorige aanpak door het afbakenen van de grenzen van tolerantie door toezichthouders. De aandringende aanpak is in vergelijking met de overtuigende dan ook minder flexibel en stelt concrete resultaten centraal.⁵⁰ Meer doelgerichte (individuele) **communicatie** en **ondersteuning** behoren hier tot de mogelijkheden.

Lokale besturen kunnen in het kader van deze opdracht enkel gebruik maken van de eerste groep overtuigende aanpakken. Om echter het complete plaatje te geven, focussen we ook kort op *deterrence* of afschrikking. Een combinatie van *deterrence* en *compliance* wordt immers aanbevolen in de wetenschappelijke literatuur. Deze aanpak is in feite alleen maar remediërend en werkt niet in op de psychisch en sociale basis van discriminatie. Bij afschrikking staat het **sanctioneren** centraal door middel van formele procedures. Sanctioneren gaat uit van een simpele optelsom: worden de kosten om te discrimineren te hoog, dan zal het individu dit niet meer doen. We wensen deze kosten niet te dragen dus we discrimineren minder. *Deterrence* doelt zowel op het afschrikken van discriminerend gedrag (dus nog vóór discriminatie plaatsvindt) door middel van mogelijke sancties, als op het sanctioneren na het vaststellen van discriminatie. Hoewel deze aanpak zijn nut heeft bij enkelingen waarbij de overtuigende aanpak minder werkt, is het eerder een laatste-kans aanpak (zie ook 3.2.3 *ethische reflecties*). Het is efficiënter zowel voor de werkgever als de maatschappij in het algemeen om overtuigend te werken met deze aanpak als stok achter de deur. Ook is het bij deze aanpak steeds noodzakelijk aanvullend juridisch advies in te winnen om mogelijke problemen en aandachtspunten in kaart te brengen.

⁵⁰ Heijden, J. van der. (2008). *Competitive enforcement: Comparative analysis of Australian building regulatory enforcement regimes*. IOS Press.

7. Samengevat

We concluderen dat correspondentietesten ook voor lokale besturen een waardevolle methode zijn om discriminatie objectief te meten tijdens de eerste fase van het aanwervingsproces. Het is echter belangrijk om voldoende voorbereidingstijd in te plannen. Zo zijn er een aantal technische en procesmatige hindernissen (bv. beperkt aantal selectieverantwoordelijken of de diplomachecks) die het afnemen van de testen sterk kunnen bemoeilijken indien deze niet worden aangekaart tijdens de voorbereiding. Ook is het cruciaal zich ten alle tijden bewust te zijn van de verschillen in vergelijking met correspondentietesten op de private arbeidsmarkt. Deze verschillen vertalen zich in (kleine) aanpassingen die nagenoeg elk aspect van het experiment beïnvloeden en zodoende een rode draad vormden doorheen deze methodiek. Deze aanpassingen en de daarbij gepaard gaande coördinatie en kosten indachtig, is het lokaal samenwerken een punt om mee te nemen en verder uit te werken naar de toekomst.

Ter gelegenheid van het afronden van deze methodiek wensen we als auteurs onze bijzondere dank uit te spreken aan iedereen die heeft bijgedragen: de medewerkers rekrutering, selectie en personeelszaken, beleidsmedewerkers gelijke kansen en diversiteit, en diensthoofden van de lokale besturen Antwerpen, Genk, Gent, Leuven, Kortrijk, en Mechelen. Ook bedanken we de medewerkers van VVSG, Diverscity, en niet in het minst de Dienst Diversiteitsbeleid van Vlaamse overheid als opdrachtgevers. Tenslotte bedanken we ook onze collega's verbonden aan de VUB en andere Vlaamse universiteiten voor hun waardevolle inbreng.

Referenties

- Acquisti, A., & Fong, C. (2020). An experiment in hiring discrimination via online social networks. *Management Science*, 66, 1005–1024.
- Baert, S. (2013). *Career Lesbians: Getting Hired for Not Having Kids?* (Discussion Paper No. 7767; IZA Discussion Paper).
- Baert, S. (2016). Wage subsidies and hiring chances for the disabled: Some causal evidence. *The European Journal of Health Economics*, 17, 71–86.
- Baert, S., Dalle, A., Louis, L., & Malfait, L. (2021). *Discriminatie op de Gentse arbeidsmarkt anno 2021: Rapport*.
- Baert, S., Norga, J., Thuy, Y., & Hecke, M. V. (2016). Getting Grey Hairs in the Labour Market: An Alternative Experiment on Age Discrimination. *Journal of Economic Psychology*, 57, 86–101.
- Baldwin, R., Cave, M., & Lodge, M. (2012). *Understanding regulation: Theory, strategy, and practice (2nd ed)*. Oxford University Press.
- Bellemare, C., Goussé, M., Lacroix, G., & Marchand, S. (2020). Video Resumes and Job Search Outcomes: Evidence from a Field Experiment. *SSRN Electronic Journal*.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94, 991–1013.
- Crabtree C (2018) *An introduction to conducting email audit studies*. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. New York: Springer, 103–117.
- Dobbin, F., & Kalev, A. (2018). Why Doesn't Diversity Training Work? The Challenge for Industry and Academia. *Anthropology Now*, 10, 48–55.
- Duguet, Parquet, L'horty, & Petit. (2015). New Evidence of Ethnic and Gender Discriminations in the French Labor Market Using Experimental Data: A Ranking Extension of Responses from Correspondence Tests. *Annals of Economics and Statistics*, 117/118, 21.
- Edo, A., Jacquemet, N., & Yannelis, C. (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household*, 17, 349–376.
- Fossati, F., Liechti, F., & Auer, D. (2020). Can signaling assimilation mitigate hiring discrimination? Evidence from a survey experiment. *Research in Social Stratification and Mobility*, 65, 100462.
- Gaddis, M. (2018). An Introduction to Audit Studies in the Social Sciences. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3–44). Springer.
- Heijden, J. van der. (2008). *Competitive enforcement: Comparative analysis of Australian building regulatory enforcement regimes*. IOS Press.
- Hogg, M. A., & Cooper, J. (Eds.). (2007). *The SAGE handbook of social psychology* (Concise student ed., reprinted). SAGE.

Kaas, L., & Manger, C. (2011). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13, 1–20.

Kramer, R. M., & Carnevale, P. J. (2003). *Trust and Intergroup Negotiation*. In *Blackwell Handbook of Social Psychology: Intergroup Processes* (pp. 431–450). Blackwell Publishing.

Lahey, J. en Beasley, (2018). Technical Aspects of Correspondence Studies. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 81–102). Springer.

Lippens, Louis, et al. (2020). Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence. *Journal of Ethnic and Migration Studies*, 1-34.

Lippens, L., Vermeiren, S., & Baert, S. (2021). *The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments* (Discussion Paper No. 972; GLO Discussion Paper). GLO.

Martiniello, B., & Verhaeghe, P.-P. (2021). Signaling Ethnicity and Race Through Names? The Perception of Names from an Intersectional Perspective. *SSRN Electronic Journal*.

Möbius (2020). *Eindrapport van het Centraal Expertencomit  tegen discriminatie*. Brussel: M bius.

Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34, 181–209.

Pope, D. G., Price, J., & Wolfers, J. (2018). Awareness reduces racial bias. *Management Science*, 64, 4988–4995.

Purssell, E., Drey, N., Chudleigh, J., Creedon, S., & Gould, D. J. (2020). The Hawthorne effect on adherence to hand hygiene in patient care. *Journal of Hospital Infection*, 106, 311–317).

Quillian, L., & Midtb en, A. H. (2021). Comparative Perspectives on Racial Discrimination in Hiring: The Rise of Field Experiments. *Annual Review of Sociology*, 14, 25.

Riach, P. A., & Rich, J. (2004). Deceptive Field Experiments of Discrimination: Are they Ethical? *Kyklos*, 57, 457–470.

Sharot, T. (2017). *The influential mind: What the brain reveals about our power to change others* (First edition). Henry Holt and Company.

Verhaeghe, P.P., Martiniello, B., & Ghekiere, A. (2020). *Discriminatie door makelaars op de huurwoningmarkt van Antwerpen*. Brussel: Vakgroep sociologie – Vrije Universiteit Brussel.

Verhaeghe, P.-P., & Van der Bracht, K. (2017). Praktijktesten: Van onderzoeksmethode naar beleidsinstrument tegen discriminatie? *Sociologos*, 38, 19-31.

Verhaeghe, P.-P., & Van der Bracht, K. (2020). *How many correspondence tests are enough to detect discrimination among single agents? A longitudinal study on the Belgian real estate market* (Discussion Paper No. 678; GLO Discussion Paper, p. 23). GLO.

Verstraete, J., Vermeir, D., De Decker, P., & Hubeau, B. (2017). *Een Vlaams antidiscriminatiebeleid op de private huurmarkt. De mogelijke rol van zelfregulering*. Leuven: Steunpunt Wonen.

Weichselbaumer, D. (2015). Testing for Discrimination against Lesbians of Different Marital Status: A Field Experiment. *Industrial Relations: A Journal of Economy and Society*, 54, 131–161.

Wood, W., & Quinn, J. M. (2003). Forewarned and forearmed? Two meta-analysis syntheses of forewarnings of influence appeals. *Psychological Bulletin*, 129, 119–138.

Zschirnt, E. (2019). Research Ethics in Correspondence Testing: An Update. *Research Ethics*, 15, 1–21.