

Verslag deskundigenbijeenkomst toetsing carcinogeniteit
7 feb 2025, 10.00-12.30, Utrecht

Aanwezigen:

Prof. Dr. M. van den Berg (Emeritus Universiteit Utrecht)
Prof. Dr. T. de Boer (Voorzitter CBG; Emeritus Universiteit Utrecht)
Dr. B. Bokkers (RIVM)
Prof. Dr. J. Duistermaat (Radboud Universiteit)
Dr. R. Godschalk (Universiteit Maastricht)
Dr. M. Luijten (RIVM)
Prof. dr. A. Ragas (Radboud Universiteit)
Dr. G. Schuur (RIVM)
Dr. K. Siezen (CBG)
Prof. dr. D. Sijm (NVWA)
Prof. dr. G. de Snoo (Universiteit Leiden)
Dr. M. van Raaij (directeur Ctgb)
Drs. R.J.T. van Lint (collegevoorzitter Ctgb)
Drs. R. van de Weerd, (collegelid Ctgb)
Prof. dr. H. Eijsackers (collegelid Ctgb)

Na een woord van welkom door Rob van Lint (voorzitter College Ctgb), legt Marcel van Raaij (gespreksleider) de aanleiding en het doel van de bijeenkomst uit. Glyfosaat en de herbeoordeling ervan hebben veel stof doen opwaaien vanwege de mogelijke relatie tussen het gebruik van glyfosaat en het ontstaan van kanker, ondanks de conclusie van de Europese agentschappen ECHA en EFSA dat glyfosaat niet geclassificeerd is voor carcinogeniteit. Ook in de media heeft deze discussie gespeeld. In een recente uitzending stelden Zembla en enkele wetenschappers kritische vragen over de methodiek voor statistische analyse die wordt gebruikt in proefdierstudies naar mogelijke carcinogene werking van chemische stoffen, zoals glyfosaat. En meer specifiek over het gebruik van 1-zijdige versus 2-zijdige testen.

Voorafgaand aan de discussie wordt in twee presentaties het veld geschetst:

Het proces van risicobeoordeling van chemische stoffen (Prof. Dr. Dick Sijm, BuRO, NVWA).

Chemische stoffen worden gereguleerd door verschillende wet- en regelgeving, veelal middels Europese verordeningen. Een overkoepelende verordening is de Europese Classification, Labeling and Packaging (CLP) verordening, die oorspronkelijk voortkomt uit een mondiale afspraak op het niveau van de Verenigde Naties over vervoer van gevaarlijke stoffen. De CLP verordening geeft regels hoe stoffen moeten worden ingedeeld (classificatie) en bevat een lijst met stoffen met hun geharmoniseerde classificatie. Voor het uitvoeren van studies ten bate van classificatie wordt mondiaal gebruik gemaakt van OECD testrichtlijnen. Die richtlijnen zorgen ervoor dat studies op een consistente en betrouwbare manier worden uitgevoerd en voorkomt daarmee dat testen voor verschillende autoriteiten iedere keer opnieuw zouden moeten worden gedaan. Aangesloten landen moeten studies die gedaan zijn volgens deze OECD richtlijnen accepteren (de studie als zodanig, niet per se de uitkomst). Studies zijn acceptabel als ze betrouwbaar en relevant zijn.

Het proces van classificatie & labeling van chemische stoffen (dr. Gerlienke Schuur, RIVM, lid Risk Assessment Committee (RAC)) – focus op carcinogeniteit.

Hoe komt het RAC tot een conclusie over classificatie van de gevaareigenschappen van een stof? Het proces bij ECHA wordt toegelicht: onder de gewasbeschermings- en biocideverordening is het verplicht om een dossier voor classificatie in te dienen (het CLH rapport). Het ingediende CLH-rapport wordt gecontroleerd door ECHA alvorens een publieke consultatie wordt gestart. Na deze consultatie geeft de lidstaat die het dossier heeft ingediend een reactie op alle ingediende commentaren. Daarna gaat het dossier naar het Risk Assessment Committee (RAC) waarvoor elke Europese lidstaat twee experts mag voordragen. Een concept RAC opinie wordt binnen RAC voor een schriftelijk en mondeling commentaar rondgestuurd en indien nodig in een vergadering besproken. De vastgestelde RAC opinie wordt op de ECHA website gepubliceerd en naar de Europese Commissie gestuurd, die na een consultatieronde een besluit neemt. Na deze besluitvorming wordt de geharmoniseerde classificatie opgenomen in de CLP verordening.

Classificatie en labelling is alleen gebaseerd op de intrinsieke 'eigenschappen' van een stof en zegt niets over de blootstelling of het daadwerkelijke risico. Indeling met betrekking tot carcinogeniteit gaat met name aan de hand van betrouwbare en relevante studies die conform OECD testrichtlijnen zijn uitgevoerd. Maar ook andere informatie kan deel uitmaken van het dossier. Van belang is de 'bewijskracht'. Bij de bewijskracht is statistische significantie in een studie belangrijk, maar bewijskracht omvat veel meer dan dat. Ook het zien van bijvoorbeeld een dosis response relatie is van groot belang. Een statistisch significant resultaat betekent niet per definitie dat het biologisch relevant is. Andersom kan ook: een effect kan biologisch relevant zijn zonder dat het statistisch significant is. De uitkomst van de statistische toets leidt niet automatisch naar een conclusie. Verder worden aanvullende bevindingen meegenomen die meewegen voor de uiteindelijke conclusie van de carcinogeniteit classificatie. Denk aan een plausibel werkingsmechanisme, indicaties uit andere studies zoals signalen van voorstadia van tumoren in de pathologie. Het RAC voert zelf geen nieuwe statistische analyses uit, maar kijkt naar de ingediende data en totale 'weight of evidence'.

Het rondetafelgesprek spitst zich vervolgens toe op een aantal aspecten:

Rapportage in (CLH) dossiers over 1- of 2-zijdig toetsen is niet transparant.

De internationale richtlijnen zijn een compromis en geven ruimte om zowel 1- als 2-zijdig te toetsen (zie kadertekst van de geldende OECD richtlijn in bijlage). Sommige gespreksdeelnemers vinden dat hier een duidelijke default (wetenschappelijke term voor "standaard") keuze zou moeten worden gemaakt. In ieder geval dient in een beoordeling duidelijk te worden aangegeven of 1- en/of 2-zijdig is getoetst en door wie. Komen de statistische resultaten vanuit het studierapport of publicatie of heeft de beoordelende instantie zelf statistische analyses uitgevoerd?¹ Tevens moet duidelijk worden aangegeven als er wordt afgeweken van een guidance. Dat mist soms, waardoor bijvoorbeeld onduidelijkheid is ontstaan in het glyfosaat dossier en waardoor veel vragen over de manier van toetsen zijn gekomen. Het verdient aanbeveling om in een guidance of testrichtlijn duidelijker op te nemen wat de default is en wanneer er 1-zijdig of 2-zijdig getoetst moet worden. Het is immers een hard technisch aspect dat vastgelegd kan worden. Nadeel van het hard vastleggen kan zijn dat er juist weer discussies ontstaan als er gemotiveerd wordt afgeweken van de default.

¹ Noot: bij gewasbeschermingsmiddelen gebeurt dat nu niet standaard met uitzondering soms van enkele landen.

Overweeg om in een richtlijn of guidance scherper vast te leggen hoe er moet worden getoetst en wat het referentiekader met betrekking tot statistiek (inclusief outliers) is, dit helpt duiding en voorkomt discussie.

Heeft 1- of 2-zijdig toetsen de voorkeur?

Statistische toetsing is slechts een van de vele aspecten in de evaluatie of een stof wel of niet carcinogeen is. Met behulp van statistische methoden kunnen de resultaten van experimenten systematisch worden geanalyseerd. Afhankelijk van de nul en de alternatieve hypothesen wordt er voor een één- of tweézijdige test gekozen. Is de nulhypothese: 'is er een verschil tussen de blootgestelde groep en de referentie c.q. controle' (2-zijdige toetsing), of is de nulhypothese: 'de blootgestelde groep heeft geen hoger aantal tumoren dan de controlegroep' (1-zijdige toetsing). Het is onlogisch om bij een stof die bedoeld is voor gewasbescherming te beoordelen of deze ook minder tumoren kan veroorzaken, je bent immers op zoek naar mogelijke carcinogeniteit en dus een toename van tumoren. Dat pleit voor 1-zijdig toetsen. Anderzijds is informatie over de fysiologische veranderingen die een stof veroorzaakt en daardoor ook kan leiden tot minder tumoren relevant om te weten. In de discussie over statistiek wordt nadrukkelijk gewezen op het belang van zowel type I als type II fouten. Met een type I fout wordt bedoeld het concluderen dat een stof een effect heeft terwijl dat eigenlijk niet zo is (vals positief). Een type II fout is de situatie dat iets niet statistisch significant is maar er eigenlijk wel een effect is (vals negatief). De foutmarge die acceptabel wordt geacht wordt meestal gesteld op 5%. De keuze voor een 1- dan wel 2-zijdige toets is relevant gezien de kans op type I en type II fouten: ze zijn met elkaar verbonden. De mogelijke impact van een type I fout en type II fout is geheel verschillend². Als de nulhypothese wordt verworpen kunnen we concluderen dat de nulhypothese niet waar is met de gekozen foutmarge (significantie niveau 5%). Echter wanneer de nulhypothese niet verworpen kan worden dan betekent dit niet dat we kunnen concluderen dat de nulhypothese waar is.

De gesprekspartners benadrukken dat een conclusie over carcinogeniteit niet alleen op basis van de statistiek kan worden getrokken maar een bredere afweging kent (zie onder), maar geven over het algemeen wel de voorkeur aan éénzijdige toetsing.

Indien de oorspronkelijke (ruwe) data beschikbaar zijn kan een beoordelende instantie alsnog ook andere type statistische toetsen uitvoeren (inclusief het combineren van data uit verschillende studies). Daar is principieel geen belemmering voor. Hoewel sommige gesprekspartners liefst zouden zien dat er dan maar één soort toets wordt weergegeven in een beoordelingsrapport, bestaat in Europa een beoordelingsrapport uit verschillende onderdelen waarin weergegeven wordt wat de oorspronkelijke studieresultaten zijn (incl. statistiek) en wat de beoordelende instantie daarvan vindt. Juist vanuit oogpunt van transparantie. Het is wel mogelijk om de resultaten van zowel eenzijdige- als tweezijdige toetsing weer te geven. Discussie kan dan wel ontstaan over welke resultaten dan uiteindelijk leidend zijn.

² Indien een type I fout zou leiden tot een onterecht afwijzing van een stof kan dat een onjuist beeld geven van een stof en leiden tot economische schade vanwege het niet kunnen gebruiken van zo'n stof. Indien een type II fout zou leiden tot onterechte toelating van een stof vergroot het kans op onacceptabele risico's voor de mens.

Overweging: Bij carcinogeniteitstudies heeft het meestal de voorkeur om éézijdig te toetsen tenzij er inhoudelijke redenen zijn om anders te besluiten. Neem in overweging om als beoordelende instantie om een eenzijdige toets additioneel uit te voeren indien dat niet in de oorspronkelijke studie is gedaan. Leg dit helder uit en leg dit vast in de beoordeling van een dossier.

Het proces om te concluderen of een stof carcinogeen is.

Het is belangrijk om de meest toepasselijke statistische toets uit te voeren en daarover helder te rapporteren. De meeste gesprekspartners geven aan dat de uitkomst van de statistische toets slechts één van de elementen is die meegewogen moet worden om uiteindelijk tot de conclusie te komen of een stof kankerverwekkend is en of die geclassificeerd zou moeten worden. Een andere toetsing leidt dus niet automatisch tot een andere conclusie. Het wel of niet aanwezig zijn van een dosis-respons-relatie is de essentie van de toxicologie dus van groot belang in de afweging. Ook een plausibel werkingsmechanisme (hoe veroorzaakt de stof het ontstaan van tumoren en is dat logisch gegeven de eigenschappen van de stof?) moet worden meegewogen. De biologische relevantie van de waargenomen tumoren of het werkingsmechanisme (vooral voor de mens) is een afwegingspunt net als observaties in andere proefdierstudies en pathologische veranderingen in weefsels die in korter durende studies zijn gezien die een aanwijzing kunnen zijn voor carcinogeniteit. Regelmatig is in proefdierstudies alleen in de hoogste dosisgroep een toename van tumoren te zien maar dat is ook een dosis waarbij er al vaak ernstige overall toxiciteit is opgetreden. De afwegingen rondom mechanisme en relevantie zijn dan extra belangrijk.

Bij het vaststellen van carcinogeniteit is niet alleen de statistiek van belang maar dienen alle aspecten (weight of evidence) te worden meegewogen om tot een conclusie te komen. Bovenal moet transparant en volledig worden gerapporteerd welke afwegingen gemaakt worden.

Overall beeld en mogelijke vervolgstappen:

Het gesprek leidt niet tot één ultieme route die altijd gevolgd moet worden maar geeft wel aanleiding om nog transparanter te zijn over de standaard werkwijze en de keuzes die worden gemaakt. Deelnemers geven aan dat eenzijdige statistische toetsing bij carcinogeniteit de voorkeur heeft en geven aan dat ook indien een studie met tweezijdige toetsing is gerapporteerd, de beoordelende autoriteit alsnog een éézijdige toetsing kan uitvoeren. Desalniettemin wordt aangegeven dat uiteindelijk de totale 'weight of evidence' bepalend moet zijn voor het trekken van conclusies. Aangezien het Ctgb opereert in een internationale context en niet eigenstandig internationale guidances kan aanpassen, zullen de bevindingen van dit gesprek worden ingebracht in de diverse (wetenschappelijke) gremia waar Ctgb en anderen in participeren.

Rob van Lint dankt alle deelnemers voor hun open en constructieve bijdrage aan het gesprek. Het gespreksverslag wordt ter informatie naar het ministerie van LNV gestuurd.

Bijlage

De huidige geldende internationale guidance voor statistische toetsing afkomstig uit OECD guideline 116:

The choice of whether to use a one- or two-side test should be made at the design rather than the analysis stage. A two-sided statistical hypothesis test tests for a difference from the negative control (in a pair-wise comparison) in either direction. A one-sided comparison tests for a difference in only one pre-specified direction, but as a consequence has more power. In a carcinogenicity study, the expectation is often that the change will be an increase in tumours in the treated group so a onesided test may be considered more appropriate, although this can be controversial. If the treatment could also be protective (i.e., reduce tumour incidence or delay it) then a two-sided comparison may be more appropriate. Regulatory authorities may have specific opinions. For instance, the US EPA (2005) notes that either “a two-tailed test or a one-tailed test may be used”.