

Van Muis naar Mens: De relatie tussen glyfosaat en kanker

Onderzoekresultaten (bij dieren) laten zich
maar moeilijk vertalen (naar mensen)

Door Dr. Marc Jacobs

MSJ Advies B.V.

Inhoudsopgave

Samenvatting	4
Inleiding.....	6
Methode	13
Literatuur zoeken	15
Data extractie	19
Data verwerking	22
Data analyse	23
Resultaten	26
Welke informatie missen we?	31
Naar welke soorten kanker is er gekeken?	34
Kenmerken van de reviews	40
Kenmerken van de studies	42
Verschillen tussen reviews met dezelfde studie	45
Risicoschattingen.....	56
Correctie voor confounders.....	60
Welke factoren hebben invloed op een risicoschatting?	62
Relatie tussen risicoschatting en de onzekerheid van die schatting.....	72
Verschillen tussen reviews in studies voor glyfosaat en kanker	81
Nabootsen van één meta-analyse	90
Invloed van correctiefactoren	98
Meta-analyses voor glyfosaat en kanker	102
Meta-analyses voor glyfosaat – odds-ratio's.....	108
Extra analyses voor de heterogeniteit	111
Publication bias.....	118
Prediction interval.....	120
Meta-analyses per kankersoort	121
Meta-analyses voor NHL en per statistiek.....	125
Meta-analyses voor NHL en per study design	128
Meta-analyses voor NHL odds-ratio's en exposure.....	129
Meta-analyses voor NHL odds-ratio's en days of exposure.....	135
Glyfosaat en kanker bij muizen en ratten	139
De Panzacchi studie	140

De relatie tussen glyfosaat en lymfoom bij dierstudies.....	174
Discussie.....	176
Wat ik heb gevonden bij mensen	177
Wat ik heb gevonden bij dieren	178
De vertaalslag van dier naar mens	179
Beperkingen van dit onderzoek	179
Conclusie	180
Figuren	181
Tabellen	184
Referenties.....	184

Samenvatting

Een zoektocht naar de mogelijke relatie tussen pesticiden en kanker moet zich beperken tot glyfosaat wanneer men tracht om dierstudies en studies onder mensen te combineren. Een statistische combinatie van gegevens bij mens en dier bleek uiteindelijk niet haalbaar door een gebrek aan combineerbare gegevens.

Onder mensen was alleen voor non-hodgkin lymfoom voldoende informatie beschikbaar om twee meta-analyses uit te voeren, waarvan alleen de eerste (wel/geen blootstelling) statistisch significant was met een verhoogd risico ongeveer 35% [14%; 61%]. Het 95% voorspellingsinterval liet een marginaal risico zien (0.02%; 83%). Een analyse voor de mate van blootstelling (*dose-response*) op basis van drie studies en 18 schattingen toonde een statistisch niet-significante *odds-ratio* van 24.1% [-0.1%; 55.7%] met een 95% voorspellingsinterval tussen de -45.3% en 281%. Met deze getallen kun je geen zinvol beleid informeren.

Bij dieren lukt het helemaal niet om een zinvolle meta-analyse uit te voeren waardoor het ook niet mogelijk was om een relatie te leggen met de bevindingen bij mensen. Het doel van dit rapport om te laten zien in welke mate de bevindingen bij dieren zich laten vertalen naar mensen is daardoor niet behaald. Samengevat zou men kunnen stellen dat in dit dossier de bevindingen bij dieren geen voorbode is (of kan zijn) voor wat ons mensen te wachten staat. Verder hoeven we niet ervan uit te gaan dat een mogelijke relatie bij mensen zich ook laat vertalen naar dieren.

Op basis van deze gegevens kan men in feite niet spreken over een relatie tussen pesticiden en kanker. Zo moet pesticiden worden vervangen door glyfosaat en 'met kanker door non-hodgkin lymfoom. Voor de andere 29 onderzochte kankersoorten bij mensen was er dus te weinig informatie om een zinvolle uitspraak te doen. Bij de meer dan 100 onderzochte kankersoorten bij dieren bleek er geen enkele serieuze *dose-response* relatie te zijn.

Beide onderzoeksvelden (mens én dier) hebben verder te maken met serieuze methodologische en statistische tekortkomingen waardoor appels niet meer met appels vergeleken kunnen worden. Gevonden relaties lijken hierdoor eerder het resultaat van hoogst sensitieve statistische screeningsmethoden. Daarbovenop is er een serieus gebrek aan beschikbare data en codes om gepubliceerde gegevens na te rekenen.

Samengevat kan men stellen dat de beschikbare gegevens rondom pesticiden en kanker haaks staat op de media-aandacht die dit onderwerp krijgt. Toekomstig onderzoek dient er op gebrand te zijn om zoveel mogelijk naar dossiervorming toe te werken waarin de gevonden gegevens in de ene studie dienst kunnen doen als zinvolle prior voor de volgende studie. Alleen op deze manier kan de waarde van meer informatie worden berekend waarmee zinvol beleid kan worden bepaald. Dat beleid dient de onvermijdbare aanwezigheid van onzekerheid te koesteren.

Inleiding

Dit is mijn vierde rapport dat betrekking heeft op de gezondheidsrisico's van pesticiden. In mijn eerste [rapport](#) deed ik een her-analyse van de review van Christopher Portier(1) uit 2020 die heeft gekeken naar de relatie tussen glyfosaat en kanker bij ratten en muizen. Deze studie werd door ZEMBLA-BNNVARA gebruikt om te laten zien dat de statistische toetsing bij pesticiden-onderzoek inadequaat is. Niet de tweezijdige toets, maar de eenzijdige toets zou men moeten hanteren, want glyfosaat kan onmogelijk beschermend werken. Met de eenzijdige toets zou een groot deel van de niet-significante bevindingen dan wel significant worden.

Ik heb toentertijd geprobeerd de bevindingen van Portier(1) te reproduceren, door de data op dezelfde wijze te gebruiken als Portier(1). Hoewel ik bij benadering vergelijkbare resultaten kreeg, stuitte ik ook op problemen die door niemand geadresseerd leken te worden. Zo werden er meer dan 200 statistische toetsen uitgevoerd zonder voor die hoeveelheid te corrigeren, en met correctie verdwenen alle statistisch significante effecten. Verder was men selectief in welke resultaten wel of niet werden meegeteld waardoor proporties als hoger werden gerapporteerd dan ze daadwerkelijk zijn. Verder bleek het vinden van een *dose-response* relatie uitermate lastig, ongeacht de gebruikte statistische techniek. Ook was er een substantiële kans op kanker voor de nul-dosering (de controlegroep). Dit alles maakt dat het zoeken naar een model dat recht doet aan de geobserveerde data, én aan de modelassumpties, buitengewoon lastig was.

In mijn [tweede en derde rapport](#) heb ik de uitkomsten beschreven van mijn onderzoek naar de invloed van alle mogelijke risicofactoren op de ziekte van Parkinson. Ik bestudeerde alle studies over de ziekte van Parkinson, die waren gebruikt voor meta-analyses, en vond uiteindelijk 885 studies met in totaal ruim 100 risicofactoren. Ik vond geen enkele risicofactor die er met kop en schouder bovenuit stak. Met gecombineerde [odds ratio's](#) van 1,0 tot 1,3 was het risico door pesticiden als zeer klein te bestempelen. Verder bleken de onzekerheden rondom deze *odds ratio's* groot te zijn. Zouden pesticiden echt een belangrijke oorzaak zijn, dan zouden deze er met kop en schouders bovenuit moeten steken bij de risicofactoren. Net als bij roken en longkanker. Ironisch genoeg sprong roken er wederom relatief sterk uit, maar deze keer als beschermende factor.

Ik was verder verrast dat er zo weinig studies waren die keken naar het verband tussen glyfosaat en de ziekte van Parkinson, aangezien hiervoor veel aandacht is bij media én politiek. Zo bleken alle glyfosaatstudies in relatie tot parkinson te zijn gedaan in de Verenigde Staten, en was het aantal beperkt. Wanneer samengevoegd laten ze weinig tot geen verband zien tussen glyfosaat en de ziekte van Parkinson.

Het feit dat jezelf inspuiten met glyfosaat leidt tot parkinson symptomen betekent nog niet dat het gebruik in het veld een risico is op de ziekte van Parkinson. Deze verkeerde manier van verbanden leggen noemt men in het Engels ook wel *atomistic fallacy*: de gedachte dat een individuele relatie zorgt voor een relatie op groepsniveau(2). Een andere manier van verkeerd redenen is de meer bekende *ecological fallacy*: het onterecht vertalen van relatie op groepsniveau naar het individu. Zo liet ik in mijn tweede en derde rapport al zien dat de 95% voorspelingsintervallen voor elk mogelijk risico op parkinson dermate groot was dat je nooit kon zeggen wie de ziekte zou krijgen, als men deze al zou krijgen.

De wens is de vader van de gedachte. Wie graag wil dat pesticiden parkinson of kanker veroorzaakt zal ook denken dat wel zo moet zijn. Studies die het tegendeel laten zien, of die niet voldoende bewijs kunnen leveren, worden al snel terzijde gelegd. Studies die wel een mogelijke relatie tonen worden daarentegen wel opgenomen en ook weer verspreid. Zo heerst al snel het idee dat er een relatie moet zijn tussen pesticiden en gezondheidsrisico's, ook wanneer de literatuur hier geen aanwijzingen voor kan vinden. Dat de meeste mensen niet eens de moeite nemen om zelf het landschap aan studies tot zich nemen, maar uitgaan van de media hen voert, spreekt dan eigenlijk voor zich. Dit is een van de redenen waarom ik mijn onderzoek publiekelijk beschikbaar maak door data en codes openbaar te delen.

In mijn optiek laat de afwezigheid van een zuivere verslaglegging zich het best tonen in wat de media schrijven over dierproeven. Zo worden de resultaten uit dierproeven met gemak geëxtrapoleerd naar mogelijke risico's voor mensen. Niet gehinderd door het gebrek aan kennis en kunde ontstaat zo het idee dat mens en muis hetzelfde zijn, en dat wat we bij muizen (en ratten) vinden ook wel voor mensen moet gelden. Ik zal in dit rapport expliciet maken waarom dit niet zomaar kan, en ik zal expliciet laten zien hoe stevig het bewijs is voor pesticiden en kanker bij muizen en mensen. Een stevige eerste conclusie heb ik in het eerste rapport al beschreven.

Ik zal in dit vierde rapport dus expliciet een duik nemen in de relatie tussen pesticiden en kanker, en mijn richtten op onderzoek bij mensen, muizen en ratten. Ik kies

bewust voor muizen en ratten omdat dit de studies zijn die veelvuldig in het nieuws terecht komen. Dat is ook de reden dat ik mij richt op kanker.

Nu is kanker al lang niet meer de dodelijke ziekte van weleer, maar wordt eerder gezien als een ernstig chronische, maar ook behandelbare aandoening¹. Met deze zin probeer ik uiteraard niet de impact van kanker te bagatelliseren, maar wel de verschuiving naar preventieve maatregelen uit te leggen. Zo is heel veel onderzoek, en heel veel geld, gemoeid met het identificeren van risicofactoren die van invloed zijn op het ontstaan (en de behandeling) van kanker.

Met het verschuiven van die is het vinden naar risicofactoren (voor kanker) *big business* geworden en de facto het uitgangspunt van de huidige epidemiologie. Daar waar de epidemiologie ooit begon met het beschrijven van (infectie)ziekten, groeide het langzaam uit naar een discipline waarin grote datasets worden gecombineerd met geavanceerde algoritmes om verbanden te vinden. Ik heb in het eerste rapport hopelijk al duidelijk laten zien wat overmatig statistiek gebruik kan doen voor een dossier. Ook in mijn tweede en derde rapport heb ik getracht om pas laat statistiek in te zetten en vooral beschrijven (en afbeeldend) te werk te gaan. Als een relatie tussen factoren namelijk echt overduidelijk aanwezig is, dan heb je geen statistiek nodig om dit te duiden. Een zuivere methodologie volstaat, maar is vaak niet haalbaar waardoor er wordt uitgeweken naar statistische oppoets-methoden. Dit zagen we allemaal in het tweede en derde rapport.

Onderzoek naar risicofactoren voor onze gezondheid leeft niet alleen op data en statistiek, maar ook op emoties. En pesticiden roepen veel angst op. Hoewel ik geen scheikundige ben, kan ik mij voorstellen dat het gebrek aan kennis over de chemische samenstelling en werking van de stof een unheimisch gevoel oproept. Pesticiden zijn geen natuurlijke middelen, maar het gevolg van scheikundige verbindingen samengesteld door mensen. Het zijn dus creaties.

Toch zou het verkeerd zijn om ze allemaal op één hoop te gooien. Er zijn namelijk [verschillende soorten pesticiden](#) (fungiciden, insecticiden, herbiciden) en allemaal hebben ze een andere focus (schimmels, insecten, onkruid). Hun doel is wel hetzelfde, namelijk het beschermen van het organisme waarom het gaat. Dit kunnen groenten, fruit of planten zijn. Omdat pesticiden chemisch van aard zijn roepen ze per definitie al weerstand op. Nu ben ik

¹ Dit geldt zeker niet voor alle kankersoorten, maar er is wel een verschuiving gaande zoals bijvoorbeeld bij borstkanker.

geen scheikundige en kan dus niet beoordelen in hoeverre die weerstand (en angst) scheikundig verdedigbaar is. Wat ik wel kan doen is beoordelen of die weerstand een voldoende wetenschappelijke basis heeft wanneer we kijken naar alle uitgevoerde epidemiologische studies. Ik ben dus geen medisch specialist, noch scheikundige, noch (micro-)bioloog, maar wel onderzoeker met een specifieke focus (en track-record) in de statistiek en wiskundige modellen. Wat ik kan bijdragen aan dit dossier is een beoordeling van de resultaten gevonden in dierproeven, en bij studies met mensen, om te zien of deze resultaten zich enigszins laten koppelen.

Hoewel een resultaat bij dieren zich niet 1-op-1 laat vertalen naar mensen, valt de mentale koppeling die mensen toch maken goed te verklaren. Zo vormen dierstudies een vast onderdeel in de ontwikkeling van behandelingen bij mensen. Het gros van de dierproeven beslaat muizen en ratten, maar soms worden ook andere dieren zoals apen ingezet. Met het oog op genetische overlap proberen onderzoekers vervolgens om de resultaten bij dieren te vertalen naar mensen. Dat dit geen directe vertaling kan zijn blijkt alleen al uit het feit dat geneesmiddelen ook op mensen getest moet worden. Geneesmiddelen die na toelating toch schadelijk blijken te zijn verdwijnen uiteindelijk weer van de markt. De schade is dan groter dan oorspronkelijk gedacht en/of weegt niet op tegen de mogelijke positieve werking van het medicijn.

Nu zijn pesticiden geen geneesmiddelen, maar ook hier gelden strikte eisen en toelatingsprocedures. Ook hier worden producten na toelating verboden als aanvullend onderzoek laat zien dat het risico op gezondheidsschade te groot is. Uiteraard wil je dat dit proces vroeger dan later in gang wordt gezet. Ook wil je dat het bewijs vanaf het begin overduidelijk is. Helaas zijn beide wensen moeilijk te realiseren juist omdat gedegen onderzoek tijd kost én omdat statistiek werkt met onzekerheden. De wereld is inherent onzeker en alleen bij een extreem causale relatie is het makkelijk om snel en eenduidig te handelen. Verder is het onmogelijk om te bewijzen dat een product (waaronder) geneesmiddel te allen tijde 100% veilig is. Dit onvermogen mag daarentegen niet worden aangedragen om dan maar alles bij voorbaat te verbieden.

Toch blijkt deze gedachte steeds vaker het uitgangspunt te zijn van mensen die graag zien dat bepaalde producten als schadelijk worden bestempeld. Niet omdat ze bewezen schadelijk zijn, maar omdat ze onbewezen onschadelijk zijn. Ik zie dit gedachtegoed het meest duidelijk bij stichtingen die graag zien dat pesticiden, waaronder glyfosaat,

verdwijnen. De twee meest uitgesproken tegenstanders zijn Pesticide Action Network (PAN) Netherlands (<https://www.pan-netherlands.org/>) en Meten=Weten (<https://metenweten.nl/>). Een klein bezoek aan beide websites onthult al snel de toon en richting van de berichtgeving: pesticiden zijn slecht, schadelijk en dienen te worden verbannen, en elk onderzoek dat ruimte biedt aan die stelling wordt op een standaard gezet. Maar dit is niet hoe wetenschap werkt. Wetenschappers dienen te beschouwen en niet te betogen. Dat wil niet zeggen dat er uit een beschouwing geen conclusie mag komen die met een bepaalde kracht mag worden gecommuniceerd. Maar dat is nog geen betoog. Een betoog is namelijk eenrichtingsverkeer waarin de boodschap, en niet het bewijs, de boventoon voert. In de wetenschap mag je hier geen ruimte voor reserveren.

In de media eigenlijk ook niet, maar in de afgelopen jaren zien we een forse verschuiving door berichtgeving vanuit zogenaamde onderzoeksjournalisten. Een bekend voorbeeld is glyfosaat en dat is ook weer niet zonder reden. Zo heeft men² in maart 2015 glyfosaat ingedeeld als waarschijnlijk kankerverwekkend voor de mens (groep 2A) op basis van beperkt bewijs voor kankerverwekkendheid bij de mens en dan met name bij non-hodgkin lymfoom (NHL). We zullen later uitgebreid bij deze relatie stil gaan staan, omdat er inderdaad heel veel studies zijn die naar NHL hebben gekeken. Ook zal ik uitgebreid stil gaan staan bij de hoeveelheid bewijs die er nu eigenlijk is om dergelijke classificaties te onderbouwen. Dat bepaalde media-bronnen deze berichtgeving zomaar overnemen is niet zonder gevaar en benadrukt de verantwoordelijkheid van de lezer om te bepalen wat betrouwbare informatie is of niet.

Zo heb ik in drie rapporten beschreven hoe uitspraken gedaan in de media geen houvast vinden in de onderzoeksliteratuur. Althans, niet als je alle beschikbare gegevens meeneemt. Maar dat is niet wat de media doen, en ook niet wat stichtingen zoals PAN en METEN=WETEN doen. Selectie voert de boventoon en een prachtig voorbeeld zien we in de berichtgeving rondom het 'nieuwste' onderzoek van het Italiaanse [Ramazzini instituut](#) uit 2025 (3). Dit onderzoeksinstituut doet al langere tijd onderzoek naar de invloed van glyfosaat (4) en is initiatiefnemer van de internationale Global Glyphosate Study groep (<https://glyphosatestudy.org/>).

² Het International Agency for Research on Cancer (IARC) en het agentschap voor kanker van de Wereldgezondheidsorganisatie (WHO).

Zo [berichtte](#) BNNVARA op 11 juni 2025 dat het Ramazzini instituut een verband had gevonden tussen glyfosaat bij kanker en proefdieren. Dit artikel werd door meerdere Nederlandse media opgepikt waarna het College voor de Toelating van Gewasbeschermingsmiddelen en Biociden ([Ctgb](#)) zich genoodzaakt voelde om te [reageren](#) en deed dat met een voorlopige [appreciatie](#). Samenvattend kwam het Ctgb tot de volgende conclusie:

Wanneer we de informatie uit deze studie meewegen in samenhang met de al beschikbare relevante data uit het glyfosaatstofdossier (zes ra7enstudies) zien we dat er a) géén sprake is van een dosis-respons relatie, b) de in de Panzacchi-studie gevonden tumoren slechts in een enkel geval voorkomen en dat c) sommige tumoren in het geheel niet gevonden worden in andere studies waarin bovendien met veel hogere concentraties glyfosaat getest is. Op basis van de nu beschikbare informatie zou de conclusie vooralsnog zijn dat de bevindingen in de studie van Panzacchi et al. (2025) niet leiden tot een andere conclusie over de kankerverwekkendheid van glyfosaat dan eerder al op Europees niveau genomen.

Voor sommigen is de reactie van het Ctgb een geruststelling. Voor anderen is het een stuk bewijs dat laat zien dat het Ctgb zich voor het karretje van de zogenaamde 'pesticidenlobby' laat spannen. Nu is die houding richting pesticiden en toelatingsprocedures niet geheel ongegrond, want door de jaren heen zijn er aardig wat pesticiden weer van de markt gehaald. Maar met die redenatie vergeet men vaak [Hanlon's Razor](#): de stelling dat iets niet aan kwade opzet moet worden toegeschreven als domheid een afdoende verklaring is. Nu is 'domheid' een uiterst negatieve manier van het verklaren van fouten, maar de stelling spreekt duidelijk over de noodzaak tot het heroverwegen van (andere negatieve) verklaringen.

Het gebrek aan het toepassen van Hanlon's razor zien we expliciet in het boek 'Het pesticidenparadijs' van Dirk de Bekker waar we kunnen lezen wat de schade van sommige pesticiden is, en hoe er sprake is van een heuse lobby. Hierdoor raakt wetenschap verstrengelt met belangen en spreekt men zelfs van moedwillig [misbruik](#). Het boek spaart ook het Ctgb niet en voedt daarmee de gedachte dat in Nederland pesticiden willens en wetens worden ingezet ten koste van onze gezondheid.

Het boek zelf blijkt uiteindelijk [naïef](#) als het om wetenschap gaat. Zo wordt ook in dit boek het dossier glyfosaat aangehaald, maar ontbreekt het Dirk de Bekker aan de nodige kennis (of moeite) om alle informatie die er te vinden is ook daadwerkelijk mee te nemen.,

Er moet namelijk een verhaal verteld worden. Daarmee leest het boek eerder als een betoog dan als een beschouwing, waardoor het (in mijn ogen) per definitie niet wetenschappelijk kan zijn. Althans, het voldoet niet aan de norm zoals ik die ken tijdens mijn promoveren.

Het hoofddoel van dit onderzoek is om wel iets wetenschappelijks neer te zetten en dan specifiek in het verbinden van de onderzoeksresultaten zoals gevonden bij dierstudies met de onderzoeksresultaten zoals gevonden in menselijke studies. De focus ligt op de relatie tussen pesticiden en kanker waarbij er geen *a priori* focus is op welk soort pesticide of kanker dan ook. Van belang is wel dat, om aan dit hoofddoel enigszins te kunnen voldoen, er studies zijn bij zowel dieren als bij mensen. Zodra er voor een bepaald pesticide wel dierstudies zijn maar geen studies bij mensen, of vice versa, dan kan de vergelijking niet worden gemaakt en wordt het pesticide ook niet opgenomen in dit onderzoek. Het soort kanker neem ik wel op, omdat ik van mening ben dat ik volledig moet zijn in wat er aan risico's wordt gecommuniceerd in de literatuur. Deze eis maakt dat de zoektocht naar relevante studies minstens net zo belangrijk is dan welk resultaat dan ook dat uit die studies komt. We zien namelijk vaak genoeg dat de media berichten over wederom een nieuw 'bewijs' dat een bepaald soort pesticide kankerverwekkend is, maar dit zijn vaak dierstudies waarvoor geen menselijke tegenpool bestaat óf een studie bij mensen welke niet wordt belicht in de context van andere studies. Wat er overblijft zijn losse flodders waarbij het oproepen van negatieve emoties belangrijker lijkt dat het in context plaatsen van een complex landschap aan beschikbare kennis.

Wanneer het over wetenschappelijke resultaten gaat (resultaten verkregen door onderzoek) is het beter om met gepaste nederigheid te berichten, zeker als de materie waarover je bericht verder gaat dan de eigen kennis. Ik zal mij daarom niet richten op de scheikunde van pesticiden, noch de microbiologische onderdelen waar het mogelijk een effect op kan hebben. Mijn kennis is compleet ontoereikend om daar ook maar enigszins uitspraken over te kunnen doen. Wat ik wel wil (en kan doen) is rapporteren over welke studies hebben gekeken naar de relatie tussen pesticiden en kanker en welke informatie deze studies ons kunnen bieden. Dit onderzoek en het rapport wat hieruit voortvloeit richt zich dus uitsluitend op epidemiologie, data en statistiek en doet dat in de vorm van een literatuurstudie met bijbehorende analyses. Het is uiteindelijk aan de lezer om te beoordelen wat de kwaliteit van dit werk is.

Methode

Deze [literatuurstudie](#) heeft als doel het beschrijven en onderzoeken van het landschap aan studies dat heeft onderzocht wat de relatie is tussen pesticiden en kanker³. Dit heb ik gedaan voor dierstudies én voor menselijke studies, en ik heb mij voornamelijk gericht op informatie die tot stand is gekomen door middel van overzichtsstudies: zogenaamde [systematic reviews](#) (5). Een specifiek soort *systematic review* is de [meta-analyse](#) die zich kenmerkt door studies statistisch met elkaar te combineren (6, 7). Omdat de meeste epidemiologische studies in het pesticiden landschap case-control studies zijn, zijn ze ook klein van opzet wat betekent dat het lastig is om een signaal te bepalen zelfs als dat signaal er is. Het idee van een meta-analyse is dat door deze kleine studies statistisch te combineren het signaal toeneemt omdat de studiegrootte toeneemt. Dit werkt alleen als de toename sterker is dan de toename in variatie én als de studies zich verantwoord laten combineren(8, 9)⁴. Die toename is haast onvermijdelijk: studies zijn nou eenmaal verschillend ook als de intentie hetzelfde is. Wat overblijft is een statistische combinatie die wordt uitgedrukt als *summary estimate*: een risicoschatting bepaalt door een combinatie van studies(10). Ik ga (onder andere) met deze *summary estimates* aan de slag.

Ook zal ik mij bezig gaan houden met elke *estimate* van elke studie die is gerapporteerd in een meta-analyse. Ik heb dus niet geprobeerd om de losse data van alle mogelijke studies te achterhalen. De hoofdreden is dat ik te weinig tijd om dit te doen én dat het mij niet zou lukken om alle studies bij elkaar te verzamelen. Het zou waarschijnlijk meer dan een jaar duren voordat ik alles zou hebben en dan is het maar de vraag hoeveel ik verzameld zou krijgen. Een deel van de studies bevindt zich namelijk achter een betaalmuur. Verder is het maar de vraag of onderzoekers hun data met mij zouden delen.

³ De opzet van dit onderzoek komt grotendeels overeen met de opzet die ik heb gebruikt bij het [onderzoek](#) naar risicofactoren voor de ziekte van Parkinson, en een [meta-analyse](#) voor de relatie tussen glyfosaat en de ziekte van Parkinson. Dit komt omdat ik in feite aan het herhalen ben wat ik al eerder heb gedaan, maar nu met een andere focus. Daarom zijn grote delen van de methodesectie haast 1-op-1 overgenomen van teksten uit eerdere rapporten: het zou in mijn optiek onnodig meer werk zijn om hetzelfde anders op te schrijven met als enige reden om het er anders uit te laten zien. Daar waar ik ben afgeweken van eerder onderzoek zal ik dit ook uitgebreid beschrijven. Onderdelen die methode technisch van belang zijn, maar die al ergens anders uitgebreid zijn beschreven, zal ik niet herhalen maar naar verwijzen. Op deze manier hoop ik dat dit rapport door de lezer wordt opgenomen als onderdeel van een groter dossier.

⁴ Iets 'verantwoord' combineren is een handeling die nooit op basis van statistiek gemaakt kan worden. Het is een menselijke handeling die, ondanks verwoede pogingen om het gehele proces zo objectief als mogelijk te maken, altijd subjectief zal zijn. Vandaar dat het uiterst belangrijk blijft om onderzoek te bespreken en resultaten te zien als een begin van een gesprek en niet het einde.

Daarom heb ik besloten om alleen aan de slag te gaan met de gepubliceerde meta-analyses. Een review zonder meta-analyse is dus niet opgenomen in deze review wat kan betekenen dat ik studies mis.

Dit literatuuronderzoek kijkt dus alleen naar meta-analyses wat maakt dat dit rapport ook wel een *umbrella review* mag worden genoemd (11): een review van reviews. Er bestaat geen officiële manier om een *umbrella review* te doen. Wel zijn er richtlijnen (12), maar die laat ik voor wat ze zijn. Uiteindelijk wil ik mij grotendeels richten op de studies die zijn gerapporteerd in de reviews en niet de reviews zelf. Deze keuze maakt mij minder afhankelijk van de kwaliteit van de gepubliceerde meta-analyses, omdat ik het grootste deel van het gepubliceerde onderzoek naar eigen inzicht kan herhalen. Uiteraard blijf ik wel afhankelijk van wat er gepubliceerd is en of er data ter beschikking wordt gesteld.

Systematische reviews kennen namelijk hun eigen problemen (13). Een analyse van gepubliceerde data vanuit meta-analyses wordt ook wel meta-epidemiologie genoemd (14). Een *umbrella review* rondom pesticiden en kanker is mijns inziens nog niet eerder gepubliceerd, hoewel er wel studies zijn die hebben getracht om de kwaliteit van bestaande reviews te beoordelen, maar dit is op een meer kwalitatieve en narratieve manier tot stand gekomen. Een *umbrella review* is meer dan dat: het is een kwantitatieve beoordeling van alle systematische reviews waarbij wordt uitgerekend wat de heterogeniteit tussen de studies is op basis van wat de reviews ter beschikking stellen. De focus ligt dus op de variatie in én tussen studies (en ook tussen reviews) en niet noodzakelijkerwijs op de schatting van het effect.

De gehanteerde methodiek maakt dat ik nu drie belangrijke speerpunten in dit landschap kan beoordelen⁵:

1. Welke studies zijn er uitgevoerd om te ontdekken wat de relatie is tussen pesticiden en kanker bij dieren én mensen?
2. In hoeverre zijn de studies het met elkaar eens wanneer ze dezelfde risicofactoren onderzoeken?

⁵ Waarbij het dus heel goed kan zijn dat ik tot de conclusie kom dat ik het niet goed kan beoordelen. De intentie van dit onderzoek is, nogmaals, niet om met een bepaalde risicoschatting te komen, maar te beoordelen hoe betrouwbaar een risicoschatting in dit veld is. Statistiek leent zich altijd wel tot het maken van uitspraken in de vorm van een cijfer, maar vertelt ons niks over de waarde van dat cijfer, zelfs als we de onzekerheid kunnen meten. Die waarde is afhankelijk van het menselijk oog. Met die reden zal ik, wederom, alle data, codes en analyses ter beschikking stellen.

3. In hoeverre zijn de bevindingen uit dierstudies te koppelen aan de bevindingen bij mensen?⁶

Literatuur zoeken

De grootste medische database van de wereld is het Amerikaanse [PubMed](#). In hun database heb ik gezocht naar alle mogelijk relevante studies door de volgende kernwoorden met elkaar te combineren: "PESTICIDES" AND "CANCER". Dit is een combinatie van kernwoorden die leidt tot een brede zoektocht en daarmee tot een groot aantal mogelijke relevante studies. Ik heb vervolgens nog een tweede zoektocht opgestart door de volgende kernwoorden met elkaar te combineren: "SYSTEMATIC REVIEW" AND "META-ANALYSIS" AND "FARMING" AND "CANCER"⁷. Door de woorden *systematic review* en *meta-analysis* toe te voegen wordt de zoektocht alweer een stuk kleiner, maar ik kan dan wel zien wat de mate van overlap is. In beide zoektochten heb ik geen onderscheid gemaakt tussen dierstudies en studies bij mensen⁸. De zoektocht in Pubmed heb ik later herhaalt in [Google Scholar](#)⁹. Verder heb ik de referentielijst van elke review gekeken of ik reviews heb gemist. Mijn laatste zoektocht is op 4 december 2025 geweest.

Nu is het zo dat veel studies niet publiekelijk beschikbaar zijn. Daarom heb ik via [ResearchGate](#) de auteurs gecontacteerd om die publicaties te sturen waar ik zelf niet bij kon. Verder heb ik gebruik gemaakt van [SciHub](#)¹⁰. Van elke publicatie heb ik verder getracht om zogenaamde *supplementary information* te downloaden wat vaak essentieel was om de basale informatie uit een meta-analyse te halen. Ook hielp het om extra informatie te verkrijgen waarmee ik mijn eigen database kon aanvullen. De gevonden publicaties heb ik met de desbetreffende PDF bestanden in een [EndNote](#) database gezet die ik op een publieke [GitHub pagina](#) heb geplaatst. De EndNote database heb ik verder gebruikt om dubbeltellingen eruit te filteren én om een Excel file te maken die ik kon gebruiken voor verdere dataverwerking, waaronder verwerking met kunstmatige intelligentie. Deze Excel

⁶ Met andere woorden: kunnen we wel spreken over één landschap op dit gebied?

⁷ De reden voor die tweede zoektocht was de gedachte dat ik met de eerste wellicht studies zal missen die wel kijken naar pesticiden, maar met een andere focus. Zo heb ik in de mijn rapport over de ziekte van Parkinson ook gezien dat sommige studi es zich sterk richten op pesticiden, terwijl anderen dat doen door te kijken naar beroep en landbouw in het algemeen.

⁸ Het zou mij echter sterk verbazen als een zoektocht naar de relatie tussen boer zijn en pesticiden dierstudies oplevert. Toch wilde ik in dit stadia niet té selectief zijn.

⁹ Voor zover dat mogelijk is, want zoeken in Google Scholar is een stuk beperkter dan wat ik aan kernwoorden kan aangeven in Pubmed.

¹⁰ De hier gedeelde link kan mogelijk al niet meer werken omdat de websites (en servers) keer op keer verplaatst moeten worden. Men is niet blij met deze website omdat wat voorheen alleen toegankelijk werd door betaling nu vrij toegankelijk is.

file laat zien welke studies ik gevonden heb, welke studies passend waren bij mijn criteria, welke studies ik wilde gebruiken voor het maken van deze database én welke studies ik uiteindelijk heb gebruikt¹¹.

Zoals al eerder beschreven leverde de combinatie van brede kernwoorden een grote lijst van mogelijk relevante studies op. Om dit werk enigszins behapbaar te maken heb ik gebruik gemaakt van een [Plus abonnement van ChatGPT](#) om zo meerdere rondes van beoordelingen te doorlopen. Zo heb ik aan ChatGPT de opdracht gegeven om de studies te rangschikken op relevantie en te bepalen of deze voldoen aan de inclusie- en exclusiecriteria.

Omdat ik meerdere rondes aan beoordelingen heb, heb ik ook meerdere inclusiecriteria gehanteerd¹², maar in het algemeen heb ik het redelijk simpel gehouden. Zo ben ik eigenlijk alleen geïnteresseerd in meta-analyses die kijken naar de relatie tussen pesticiden en kanker óf landbouw (*farming*) en kanker. Ik heb geen specifieke voorkeur voor een soort kanker, soort pesticide of de leeftijd van de onderzoeksgroep. Het enige wat ik wilde was zien of een review (of studie) meerdere studies statistisch met elkaar gecombineerd heeft om zo tot een nieuwe schatting van een mogelijke relatie te komen. De poging van een review om tot een nieuwe schatting te komen biedt mij verder ook weer informatie over de heterogeniteit tussen studies. De manier waarop dit door een studie werd gedaan was voor mij uiteindelijk niet interessant, omdat ik zelf met de studies aan de slag ga. Dit rapport overstijgt dus per definitie elke tot nu toe gepubliceerde review.

Omdat mijn inclusiecriteria breed zijn, en ik dus een hoop resultaten voor de kiezen kreeg, leek het me relevant én kosteneffectief om de eerste iteratie van dit proces aan ChatGPT over te laten. Bij elke prompt heb ik de opdracht gegeven om elke stap te beschrijven en deze in een kolom te plaatsen in de Excel file. Dit heeft uiteindelijk geleid tot een aantal iteraties waarbij ik niet altijd een blij was met het resultaat, omdat het (1) niet voldeed aan wat ik vroeg en (2) niet overeenkwam met de tekst in de publicatie zelf. Ik heb daarom voortdurend beoordeeld of de werking van ChatGPT consistent was door op meerdere manieren om hetzelfde te vragen. Zo heb ik ook gebruik gemaakt van meerdere sessies om blanco te beginnen.

¹¹ Sommige studies waren erg interessant om mee te nemen, maar kon ik niet opnemen door een gebrek aan toegang. Het opvragen van de studies was soms succesvol, maar is vaker wel dan niet onbeantwoord gebleven. In de Excel database kun je exact zien welke PDF bestanden ik uiteindelijk niet heb kunnen meenemen in de beoordelingsrondes.

¹² Meer informatie hierover staat in de resultaten sectie van dit rapport.

De werking van ChatGPT was zeker niet consistent. Hoewel je inconsistenties (deels) kunt ondervangen door de prompt aan te passen, leek het mij dus beter om op verschillende manieren hetzelfde te vragen om op die manier te achterhalen of de resultaten consistent genoeg zijn om deze ook als betrouwbaar te bestempelen. Ook hielp het door de referentielijst van elke meta-analyse te doorlopen om te zien of ik geen studie zou missen¹³.

In de Excel file die uiteindelijk alle gemaakte stappen toont (door steeds weer kolommen toe te voegen) kan de lezer zien wat ik gedaan heb om tot een betrouwbare lijst te komen. Het is aan de lezer om zich hier tegenaan te bemoeien als die denkt dat wat gedaan is, incorrect gedaan is.

Wie de Excel file tot zich neemt ziet al snel dat de kolommen verschillende kleuren hebben. Dat komt dus omdat ik meerder zoektochten en handelingen met elkaar heb gecombineerd. De kolommen met deze **kleur** tonen de resultaten van de eerste en tweede zoektocht. De kolommen met deze **kleur** tonen een (derde) aparte zoektocht waarbij ik expliciet heb gezocht naar dierstudies die keken naar de relatie tussen glyfosaat en kanker na 2019¹⁴. Ik heb in PubMed gebruik gemaakt van de kernwoorden "CANCER" en "GLYPHOSATE" en mij beperkt tot dierstudies en studies na 2019¹⁵. Dit heb ik daarna herhaald in Google Scholar.

Het uitvoeren van een derde zoektocht had meerdere redenen. Ten eerste kon ik in de eerste twee zoektochten al zien dat er wat betreft dierstudies geen enkele meta-analyses te vinden zijn, behalve de meta-analyse van Portier(1) die ik in een [eerder rapport](#) al grondig onder de loep heb genomen. Het blijkt dus nu ook de enige meta-analyse te zijn die kijkt naar de resultaten van dierstudies bij pesticiden. Daarmee is het onderzoeksveld van dit rapport in één klap verkleint tot glyfosaat, want ik wil namelijk beoordelen in hoeverre de resultaten bij dierstudies zich laten vertalen naar studies bij mensen.

Maar dan ben ik er nog niet, want nu ik weet dat ik mijn focus moet verkleinen kon ik het niet laten om toch nog een vierde zoektocht te doen waarbij ik exclusief naar studies heb gezocht waarvan de focus lag op glyfosaat, kanker en mensen. De resultaten staan in de kolommen met deze **kleur**. Ik heb wederom gebruikt gemaakt van PubMed en de kernwoorden "GLYPHOSATE" AND "CANCER" AND "REVIEW" gebruikt met als extra

¹³ Vanwege de media-aandacht heb ik ook in Google zelf gezocht naar artikelen die schrijven over pesticiden en kanker.

¹⁴ Dit komt omdat uit de eerste en tweede zoektocht bleek dat voor dieren er alleen maar meta-analyses zijn voor glyfosaat. Zie de resultaten sectie voor een meer uitgebreide beschrijving.

¹⁵ Deze beperkingen kun je in PubMed duiden via de *advanced search*: <https://pubmed.ncbi.nlm.nih.gov/advanced/>

toevoeging dat ik alleen studies van mensen wilde opnemen. Ik heb nu expliciet gezocht met het kernwoord "REVIEW" omdat dit breder is dan *systematic review* of *meta-analysis* is en ik daarmee meer resultaten krijg. De toename in dubbeltellingen nam ik voor lief vanuit de wens om bovenal uitputtend te zijn. Ik wilde geen studies missen.

Vervolgens heb ik aan ChatGPT gevraagd om de resultaten van de vierde zoektocht te vergelijken met die van de eerste en tweede zoektocht. De bevindingen van ChatGPT heb ik daarna systematisch onderzocht door te zoeken naar inconsistenties en dat bleken er (wederom) best wat te zijn. Daarom heb ik van honderden artikelen zelf handmatig het abstract en de volledige tekst doorlopen om te zoeken of in deze studies wel degelijk is gekeken naar glyfosaat en kanker. Ik ben pas gestopt toen ik zeker wist dat ik rondom glyfosaat en kanker alles heb gevonden wat er maar te vinden is¹⁶.

Uiteindelijk hebben deze exercities mij gebracht tot twee lijsten met resultaten. In de ene lijst staan alle reviews die keken naar pesticiden en kanker bij dieren. We weten nu al dat alleen de meta-analyse van Portier(1) in aanmerking komt en die studie heb ik al onderzocht in een eerder [rapport](#). Dit betekent dat ik eigenlijk alleen nog maar hoeft te zoeken naar nieuwe dierstudies bij glyfosaat en kanker. Ik laat hier dus expliciet de eis los dat elke studie een meta-analyse moet zijn. Met het oog op de media-aandacht voor de Ramazzini studie(3), waarvan we weten dat deze de relatie tussen glyfosaat en kanker heeft onderzocht, zou het niet gepast zijn om deze studie niet op te nemen in mijn eigen meta-analyse rondom glyfosaat en kanker bij dieren. En om volledig te zijn heb ik dus maar gezocht naar alle mogelijk nieuwe studies die vanaf 2019 zijn gepubliceerd.

In de tweede lijst staan alle meta-analyses die keken naar glyfosaat en kanker bij mensen. Door uit elke meta-analyse de bevindingen te extraheren kon ik een database bouwen waarmee ik eigenhandig en naar eigen inzicht grafische én statistische analyses kan uitvoeren. Deze analyses overstijgen de analyses van de losse meta-analyses die elk berusten op een klein deel van deze database¹⁷.

¹⁶ Waarbij 'zeker' wederom een subjectief gevoel is en geen objectieve maat. Ook hier is het weer aan de lezer om te beoordelen of datgene wat ik gedaan heb correct is gedaan.

¹⁷ Mijn intentie is dus niet om te herhalen wat er al gedaan is om te achterhalen of wat er gedaan is goed gedaan is. Eerder wil ik ontdekken wat er allemaal aan informatie is verzameld en hoe bruikbaar die informatie is. Ook heb ik niet geprobeerd om de losse data van de studies te achterhalen verder dan wat direct beschikbaar zou kunnen zijn in de publicatie of de supplementen. De reden hiervoor is tijd (het zou waarschijnlijk meer dan een jaar duren voordat ik alles zou hebben), maar de grote aanwezigheid dat ik die data niet zou krijgen. Daarom heb ik besloten om alleen aan de slag te gaan met de meta-analyses zelf.

Data extractie

Het mag nu al duidelijk zijn dat de resultaten uit de verschillende zoektochten bepalen welke onderdelen ik wel of niet heb meegenomen in de data-extractie. Zo heb ik al een data-analyse verricht bij dierstudies wat betekent dat de data extractie bij een nieuwe studie exact hetzelfde protocol moet volgen. De gehanteerde methode staat in dat [rapport](#) beschreven en zal ik hier daarom niet nogmaals noemen.

Alleen voor de studies onder mensen moet ik een nieuwe, meer uitgebreide data extractie doen waarbij ik zowel op het niveau van de review als op het niveau van de studie kijk wat er aan informatie beschikbaar is. Ook nu leun ik sterk op eerder gemaakte keuzes, zoals beschreven in [twee eerdere rapporten](#) van mijn hand. Op basis van die resultaten heb ik besloten om een aantal variabelen niet meer uit de reviews of de studies te extraheren.

Zo ben ik gestopt met het extraheren van informatie met betrekking tot hoe de meta-analyses zijn uitgevoerd. Ook ben ik niet meer gaan zoeken naar data die iets vertelt over de kwaliteit van de studies is beoordeeld, noch wat die kwaliteit exact is. Voor die keuzes zijn meerdere redenen. Zo bleek uit mijn eerdere [rapport](#) dat beide onderdelen geen noemenswaardig effect hebben op de resultaten, ook al zou men dit wel verwachten. Zo denken we vaak dat de uitkomst van een methodologisch meer hoogwaardige studie ook meer betrouwbaar is, maar het is best lastig om te bepalen (en dus te kwantificeren) hoe groot die stijging zou kunnen zijn. Verder verschillen studies vaak noemenswaardig in hun opzet waarbij de opzet ook gebonden is aan wat toentertijd bekend én gangbaar was. We kunnen dus niet zomaar zeggen dat een studie met een bepaalde kwaliteitsscore direct een bepaalde hogere mate van kwaliteit laat zien. Een dergelijke kwantitatieve uitspraak is eigenlijk onmogelijk om te maken, omdat de kwaliteit zelf bepaald is door wat er aan informatie wordt beschreven in de publicatie.

Eenzelfde redenatie is ook van toepassing op de kwaliteit van de meta-analyse zelf(6, 15, 16) en dat geldt ook voor omgevingsepidemiologie: het veld waar deze en eerdere rapporten zich mee bezig houdt(17-19). Omdat ik van elke meta-analyses de data van elke studie extraheer en zelf analyses doe is de methode van de meta-analyse eigenlijk niet meer relevant¹⁸. De kwaliteit van de studie is zelf wel relevant, maar mijn [tweede rapport](#) liet al

¹⁸ Behalve de inclusie- en exclusiecriteria. Ik hoop dit juist te ondervangen door alleen te zoeken naar meta-analyses en vervolgens te zien welke studies wel of niet in een meta-analyse zijn opgenomen. Ik ga niet expliciet op onderzoek uit naar de beschreven rede voor in- of exclusie, maar laat liever de grafieken hun werk doen wanneer ik alle rapportages bij elkaar zet.

zien dat het onduidelijk is hoe relevant een kwaliteitsscore is ten opzichte van bijvoorbeeld de grootte van de studie.

Zo weten we dat hoe kleiner een studie is, hoe onzekerder de schatting van het effect¹⁹. Omdat we bij mensen niet kunnen experimenteren missen we verder randomisatie en blinding. Voeg daarbij de problemen toe die te maken hebben met vragenlijstonderzoek (wat valt er nog te herinneren?), onderzoek dat kijkt naar de mate van blootstelling (hoe definieer je blootstelling?), en onderzoek waarbij eerste naar de ziekte wordt gezocht en vervolgens pas naar risicofactoren (*case-control* studies) en je beseft al snel dat een kwaliteitsscore weinig zegt: elke studie is per definitie al serieus beperkt.

Dit wil niet zeggen dat een *case-control* studie of een *cohort* studie alleen goede studies kunnen zijn naar hun maatstaven, maar niet naar de maatstaven van wat een goede studie in het algemeen is. Toch kunnen we vaak niet anders. Ook hebben beide soorten studies ons absoluut zinvolle informatie gebracht in andere onderzoeksvelden, zoals roken en kanker, maar de kracht van de studie lag in het daadwerkelijke effect en niet de studie zelf. Met andere woorden: de relatie is zo overduidelijk dat je hem altijd wel vindt. Bij omgevingsepidemiologie ligt dat vaak anders.

Maar het zou verkeerd zijn om hier te stoppen door elke studie (en het veld) bij voorbaat al af te schieten. Het doel van dit rapport is namelijk niet om een onderzoeksveld af te vallen, maar om te beoordelen hoeveel informatie er eigenlijk beschikbaar is voor dit gebied én hoe betrouwbaar die informatie is. Ik wil namelijk graag weten of de media-aandacht gerechtvaardigd is. Daarom heb ik uit elke studie wel degelijk informatie geëxtraheerd, maar beperkter dan in vorig onderzoek op basis van geleerde lessen.

Uiteindelijk heb ik uit elke meta-analyse elementen geëxtraheerd die mij vertellen wat er precies is onderzocht en hoe het is onderzocht. Verder heb ik uit elke studie informatie gehaald met betrekking tot glyfosaat en kanker. Essentieel waren voor mij die studies waarvan de zogenaamde *risk estimates* werden genoemd: dit zijn de schattingen van het mogelijk risico op kanker door de desbetreffende risicofactor. Voorbeelden zijn de eergenoemde *odds-ratio's* en *risk-ratio's*(20-25).

Uit elke review heb ik de volgende elementen gehaald die betrekking hebben op de review: (1) auteur, (2) jaar, (3) titel en (4) of het een meta-analyse is. Vervolgens heb ik

¹⁹ Dit liet ik duidelijk zien in mijn [eerste rapport](#).

elementen uit elke review gehaald die betrekking hebben op de studie: (5) auteur, (6) titel, (7) jaar, (8) land, (9) periode, (10) *study design*, (11) kankersoort, (12) soort pesticide, (13) soort blootstelling waarmee is gerekend, (14) hoeveelheid blootstelling, (15) geslacht, (16) aantal mensen risicogroep, (17) aantal mensen controlegroep, (18) aantal mensen blootgesteld in de risicogroep, (19) aantal mensen blootgesteld in de controlegroep, (20) risico statistiek, (21) risicoschatting, (22) risicoschatting laag, (23) risicoschatting hoog, (24) betrouwbaarheidsinterval, (25) aanpassingen voor *confounders* (ja/nee) en (26) voor welke *confounders* de analyse werd aangepast. Al deze elementen heb ik een Excel database gezet.

Het mag duidelijk zijn dat omdat dit werk door één persoon wordt uitgevoerd het rapport de zienswijze van één persoon volgt. Ook is het goed mogelijk dat er fouten zijn gemaakt in de extractie²⁰. De toepassing van AI in mijn zoektocht naar mogelijk interessante studies, en het extraheren van gegevens, heeft mij geholpen door op vele verschillende manieren verschillende rondes van beoordelingen tegen elkaar af te zetten. Verder worden de uitkomsten in meta-analyses vaak in figuren gezet waarvan ik de informatie met ChatGPT heb geëxtraheerd. Dat ging niet altijd vlekkeloos²¹, en de kans op fouten is door ChatGPT zeker niet tot nul gebracht. Eerder worden er nieuwe soorten fouten geïntroduceerd die vroegen om extra controles van mijn kant²².

Om er toch grotendeels zeker van te zijn dat de data accuraat zijn overgenomen heb ik door middel van de database de data van studies vergeleken die in meerdere reviews zijn opgenomen. Dit leverde vaak verschillen op die ik niet kon verklaren, zoals bijvoorbeeld een verschil in jaar of een verschil in land. Verschillen tussen schattingen heb ik gelaten zoals ze zijn, omdat sommige reviews ervoor kiezen om schattingen op te nemen die zijn berekend vanuit complexe modellen terwijl andere studies schattingen opnemen die op een meer simpele manier tot stand zijn gekomen²³. Deze verschillen zijn juist interessant om te ontdekken. Ook heb ik niet bepaald of elke review wel de juiste data heeft opgenomen door in de originele publicatie te gaan graven: ik heb deze waarden voor waar aangenomen juist omdat lezers er ook vanuit (mogen) gaan dat dit correct is. Ook al is dat zeker niet altijd het

²⁰ Daarover later meer in de discussie sectie.

²¹ Problemen die ik heb ondervonden wisselen tussen het niet kunnen extraheren van de data tot aan het zelf maar invullen van gegevens die achteraf niet blijken te kloppen.

²² De cijfers worden vaak correct opgenomen door ChatGPT, maar is altijd het probleem van consistentie wat inherent is aan de kansrekening aan de achterkant van het algoritme.

²³ Raw (crude) vs adjusted estimates.

geval. Ik heb alleen dan data aangepast wanneer door de auteurs een correctie werd gepubliceerd.

Data verwerking

In verschillende reviews werd belangrijke informatie over studies niet verstrekt²⁴. Wanneer dit het geval was heb ik gekeken of deze studie ook in andere reviews vermeld werd.

Wanneer dit het geval was én in meerdere reviews de data wel voorhanden was heb ik gekeken of die reviews het met elkaar eens waren. Als dit het geval was heb ik die data overgenomen zodat ik bij het maken van de grafieken geen informatie verlies²⁵. Ik had er uiteraard voor kunnen kiezen om te laten zien welke review welke informatie wel of niet vermeld, maar dat was niet mijn doel. Mijn doel was om te achterhalen welke studies er beschikbaar zijn, naar welke kankersoort(en) onderzoek is gedaan en hoe groot het risico op kanker was. Vervolgens wilde ik zien of de gerapporteerde risico's vanuit dezelfde studies verschillend werden gerapporteerd in de reviews. Noemenswaardige verschillen zijn in mijn optiek dus belangrijker dan het ontbreken van informatie, tenzij die informatie in elke review ontbreekt.

In tegenstelling tot mijn [rapport](#) over risicofactoren voor de ziekte van Parkinson heb ik deze keer wel schattingen ingevuld met behulp van de data van andere reviews. Dit heb ik alleen gedaan als het gros van de reviews exact dezelfde data hebben opgenomen. Dat sterkte mij in de gedachte wat de missende waarde zou kunnen zijn. Hoewel ik uiteindelijk per studie, per kankersoort en per blootstelling maar één waarde kan opnemen in de meta-analyse, geeft deze manier van invullen een idee hoe bruikbaar het landschap als. Als studies duidelijk verschilden in hun risicoschatting, maar het onduidelijk was of er ook echt iets anders geschat werd, heb ik de missende waarden niet ingevuld. Er was in mijn optiek sprake van te veel divergentie²⁶. Deze studies heb ik wel opgenomen in de database omdat

²⁴ De schatting voor het risico op kanker was voor mij het meest belangrijke. Als een studie deze schatting niet had werd hij wel in mijn database opgenomen, maar zal geen plek krijgen in de verdere analyses. Vaak was het zo dat die studie in een andere review wel volledig werd neergezet, met als gevolg opname in de database, maar dan wel zonder de mogelijkheid om te vergelijken. De mogelijkheid om schattingen en informatie tussen meta-analyses te vergelijken was voor mij een van de speerpunten om een idee te krijgen van de kwaliteit van de reviews én de mate waarin methodiek kan verschillen.

²⁵ Dit heeft geen invloed op de meta-analyses.

²⁶ Een ander noemenswaardig verschil was het rapporten van de statistiek. Zo heb ik vaak exact dezelfde waarden gevonden over reviews heen, maar werden ze beschreven als *odds-ratio*'s of *risk-ratio*'s. Je kunt deze statistieken niet zomaar inwisselen. Ik heb deze verschillen niet overschreven. Wat ik wel heb gedaan is dat als de gebruikte statistiek niet werd beschreven voor een studie in een review, maar de rest van de reviews het eens waren over de gebruikte statistiek, dan vulde ik die missende waarde in op basis van de rest.

ik aan de lezer wil laten zien welke studies dat zijn. Ze zullen hoe dan ook niet worden meegenomen in de meta-analyse.

Data analyse

Het grootste doel van dit rapport is om het landschap van studies te beschrijven die keken naar de relatie tussen glyfosaat en kanker bij mensen en dieren. Op die manier ben ik in staat om te zien of schattingen bij dieren stand houden bij mensen. Ook wil ik kijken in hoeverre reviews het met elkaar eens zijn én in hoeverre studies het met elkaar eens zijn. En als ze het niet met elkaar eens zijn dan wil ik ontdekken hoe dit wellicht komt. Ik ben dus, samengevat, op zoek geweest naar het signaal tussen de ruis waarbij de ruis voor mij uiteindelijk interessanter is dan het signaal(26, 27).

Het is belangrijk om te blijven benoemen dat we eigenlijk drie soorten analyses hebben. In de eerste analyse kijk ik naar de relatie tussen glyfosaat en kanker bij mensen. In de tweede analyse kijk ik wat het toevoegen van nieuwe studies doet voor mijn eerdere analyse naar de relatie tussen glyfosaat en kanker bij dieren. De bevindingen uit die twee analyses gebruik ik dan om mijn laatste analyse te doen en dat is een direct vergelijk tussen dieren en mensen. Ik zal de analyse splitsen per type kanker om het vergelijk meer robuust te maken. De ene kankersoort is namelijk de andere niet. Dit weerhoudt mij er niet van om ook over alle kankersoorten heen te kijken om te zien hoe de grootte van de ratio's veranderen en of we eenzelfde mate van relatie zien in beide soorten onderzoek. Ik ben dus, wederom, op zoek naar signalen waarbij ik de ruis minstens net zo interessant vindt.

De vraag wanneer een signaal interessant is om te vermelden is trouwens geen gemakkelijke. Dit noemen we ook wel een *effect size* oftewel de grootte van het effect(25, 28). Zoals verwacht mag worden is er geen eenduidig antwoord voor wat men een kleine, gemiddeld of grootte *odds-ratio* kan noemen. Zo stelt de [University of Cambridge](#) dat 1.5, 3.5 en 9 respectievelijk synoniem staan voor klein, middel en groot, maar dit is geen vaste indeling. Een andere studie komt met waarden van 1.68, 3.47 en 6.71 (29). Nog een andere studie geeft een reeks weer van wat gezien mag worden als kleine, gemiddelde en grootte *odds-ratios*. Een gemiddelde *odds-ratio* begint vanaf 2.48 (30).

Om de verhouding tussen signaal en de ruis te tonen, werkt het in mijn optiek het beste om grafisch aan het werk te gaan. Wanneer ik alleen maar getallen zou uitrekenen

introduceer ik (onbedoeld) ook weer ruis²⁷. Daarom zal ik in de resultaten sectie in het begin vooral beschrijvend én grafisch te werk gaan. Zo kan ik laten zien wat ik gevonden heb, wat deze bevindingen (eventueel) betekenen, en vooral hoe de bevindingen zich tot elkaar verhouden én het grotere doel verhouden. Het is aan de lezer om hier conclusies uit te trekken, hoewel ik mij uiteraard niet zal onthouden van eigen commentaar. Uiteindelijk is het, nogmaals, niet de bedoeling om een risicofactor definitief te ankeren als 'beschermend' of 'risico', 'significant' of 'niet significant'. Het is eerder de bedoeling om een decenniagroot landschap inzichtelijk te maken.

Met het opzetten van een database neemt ook de verleiding toe om met allerlei statistische methoden aan de slag te gaan. Zo zou ik oneindig veel meta-analyses en meta-regressies kunnen uitvoeren en elk van deze analyses kunnen uitbreiden met verdere subgroep analyses. Deze haast onophoudelijk manier van data analyseren noemen we ook wel *p-hacking* (31)²⁸. Door data te blijven splitsen lukt het eigenlijk altijd wel om iets te vinden wat statistisch significant is en ik heb dit uitgebreid besproken in mijn eerste [rapport](#).

Ondanks mijn wens om vooral beschrijvend en grafisch te werk te gaan, ben ik in mijn analyse niet helemaal vrij gebleven van de statistiek. De reden is dat ik graag overschrijdend te werk wil gaan: ik wil zien wat er gebeurt als ik alle studies voor één bepaalde kankersoort combineer. Welke schatting krijg ik dan? Hoeveel variatie zit er tussen studies? Hoe onzeker is die schatting? Het is uiteindelijk aan de lezer om te bepalen of ik (te) ver ben gegaan.

Om die beoordeling te kunnen maken zal ik zowel de data als de analyses ter beschikking stellen op de [GitHub pagina](#). Ook zal ik zo goed en kwaad als ik kan beschrijven wat ik gedaan heb en waarom ik dat gedaan heb. Deze uitleg volgt in de resultaten sectie.

Ik zal verder later laten zien hoe moeilijk het is om een duidelijk signaal op te pikken als er zo veel verschillen zijn tussen niet alleen de reviews, maar ook tussen de studies zelf²⁹. Als een meta-analyse ergens voor bedoeld is, dan is het wel om te laten zien hoe verschillend studies zijn in hun schattingen.

Om de data te analyseren heb ik gebruik gemaakt van het statistiek programma R en Rstudio. De gebruikte statistische pakketten zijn te achterhalen in de meegeleverde codes én

²⁷ Het interpreteren van statistiek is een vak apart en wordt te vaak en voornamelijk in de media afgedaan als een makkelijke beoordeling van een cijfer.

²⁸ Zoekt en gij zult vinden.

²⁹ En ook dezelfde studies per review.

zal ik in dit rapport vermelden wanneer ik vind dat dit essentieel is. De afbeeldingen, data en codes zijn allemaal te vinden op de [GitHub pagina](#).

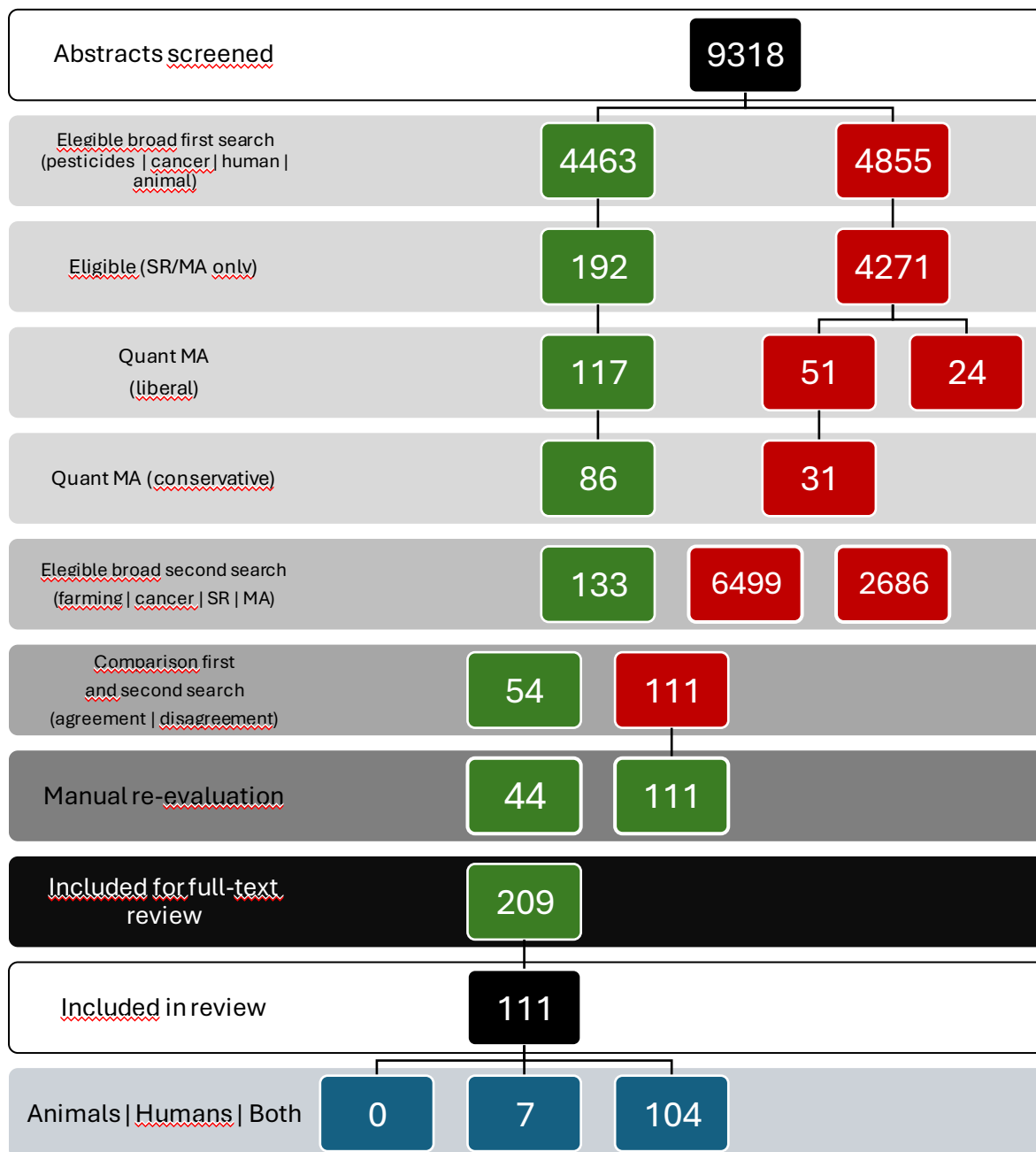
Hoewel dit rapport in het Nederlands is geschreven zal ik zo nu en dan gebruik maken van Engels terminologie en dat heeft verschillende redenen. Ten eerste zijn alle studies in het Engels en in die publicaties zul je dezelfde terminologie vinden. Het is dus goed als de lezer hier bekend mee raakt. Het overnemen van deze termen maakt het mijns inziens dus alleen maar makkelijker om de gebruikte informatie te verifiëren. Verder is de database ook in het Engels en dat is omdat ik graag zou zien willen dat deze gegevens verder gebruikt worden door anderen³⁰. Met die reden zijn de grafieken ook in het Engels opgemaakt, hoewel ik ze in dit rapport zo goed en duidelijk als ik kan zal verhelderen met in het Nederlands geschreven tekst.

³⁰ Al is het maar om fouten eruit te halen.

Resultaten

Zoals al beschreven is het zoekproces door ChatGPT ondersteund (maar uitdrukkelijk door mij gecontroleerd en aangepast) en zijn de resultaten zoals hier beschreven weer het resultaat van het samenvoegen van in totaal vier zoektochten. De resultaten van de eerste twee zoektochten zien we in **Figuur 1**.

Figuur 1. Flow-chart eerste twee studies



Uiteindelijk hebben we 9318 abstracts gevonden waarvan er 4463 keken naar de relatie tussen pesticiden en kanker. Van die 4463 werden er 192 aangemerkt als systematische review / of meta-analyse. Van die 192 werden er 117 beschreven als kwantitatief wanneer je er met een wat meer liberale bril naar keek en 86 wanneer je er meer conservatief naar keek³¹.

In dit stadium betekent het dus dat we van 4855 abstracts weten dat deze niet voldoen aan de inclusiecriteria. Meer specifiek kunnen we lezen dat de studies om de volgende redenen niet voldoen:

1. Kijkt niet naar pesticiden én kanker (N=289).
2. In-vitro studie (N=3).
3. Noemt kanker maar niet pesticiden (N=2860).
4. Noemt pesticiden maar niet kanker (N=1883).

Omdat ik gebruik heb gemaakt van ChatGPT wilde ik de resultaten niet zomaar overnemen en een deel systematisch controleren. De resultaten zie je in de vakken die rood gekleurd zijn en die, volgens ChatGPT, niet voldoen aan de inclusiecriteria van dit onderzoek. Ik kom later op deze getallen terug.

In **Figuur 1** staan dus ook de resultaten van de tweede zoektocht (vanaf de zesde regel) van waaruit 133 systematische reviews of meta-analyses werden geïdentificeerd. Als we kijken naar de redenen voor exclusie in de tweede zoektocht dan zien we de volgende redenen verschijnen:

1. Geen focus op landbouw en kanker (N=6499).
2. Geen systematische review of meta-analyse (N=2686).

Als ik die resultaten toevoeg aan de eerste behaalde resultaten dan lijkt het alsof ik 325 mogelijke interessante reviews heb³². Maar dit is niet het hele verhaal, want van die 192 uit de eerste zoektocht zijn er maar 86 die waarschijnlijk echt een kwantitatieve schatting hebben gedaan. Verder is mogelijk sprake van overlap.

³¹ Het verschil tussen liberaal en conservatief zit hem in het wel of niet samenvoegen van meerdere studies

³² 192+133 = 325

Wanneer ik inderdaad kijk naar de mate van overlap dan zie ik dat 9207 abstracts in beide rondes hetzelfde worden beoordeeld. Daarvan worden er 54 opgenomen in zowel de eerste als de tweede zoektocht met bijbehorende inclusiecriteria. Van 111 studies zie ik dat ze óf vanuit de eerste zoektocht worden geïncludeerd óf vanuit de tweede zoektocht. Dit zijn studies die ik handmatig wil evalueren om te zien waarom er geen overlap is. Om er zeker van te zijn dat de 54 studies daadwerkelijk interessant zijn wil ik ook hier zelf bepalen of ze moeten worden opgenomen. Dit betekent dat ik in totaal 165 studies handmatig heb beoordeeld, op basis van de volledige tekst, of ze voldoen aan de inclusiecriteria. Uiteindelijk wil ik van 44 andere studies ook weten of ze voldoen³³ wat maakt dat het totaal aan handmatig beoordeelde studies op 209 neerkomt.

Helaas konden we van 47 studies niet de PDF bestanden downloaden wat er op neerkomt dat we maar 162 volledige artikelen hebben weten te beoordelen. Van die 162 voldeden er 111 aan de inclusiecriteria. Een splitsing naar soort onderzoek toont, volgens ChatGPT, dat er nul studies zijn die alleen naar dieren keken, zeven studies die bij mensen onderzoek hebben gedaan en 104 studies naar beiden hebben gekeken³⁴.

De resultaten uit de vierde zoektocht laten zien dat er 63 studies zijn die voldoen aan de inclusiecriteria. Specifieke redenen om niet te worden opgenomen zijn als volgt:

1. Glyfosaat wordt niet genoemd (N=672).
2. Glyfosaat wordt niet genoemd en niet naar kanker gekeken (N=85).
3. Glyfosaat wordt niet genoemd en niet naar kanker gekeken en geen systematische review of meta-analyse (N=819).
4. Glyfosaat wordt niet genoemd en geen systematische review of meta-analyse (N=6139).
5. Niet naar kanker gekeken (N=65).
6. Niet naar kanker gekeken en geen en geen systematische review of meta-analyse (N=1224).
7. Geen systematische review of meta-analyse (N=249).
8. Alleen dieren meegenomen (N=1).

³³ Op basis van verschillende redenen.

³⁴ Maar omdat we al kennis hebben van de Portier studie, die ook een meta-analyse is, weten we dat deze aantallen niet kunnen kloppen. De reden dat ik ze hier nu toch presenteer is om duidelijk te maken hoe belangrijk het is om werk uitgevoerd door kunstmatige intelligentie te controleren.

9. Artikel is teruggetrokken (N=1).

Van de 111 studies die ik uit de eerste en twee ronde heb opgenomen voor volledige beoordeling zie ik maar 12 studies terug in de vierde zoektocht. Ook zie ik twee glyfosaatstudies die ik eerder nog niet had opgenomen. Dit vraagt dus om een extra ronde controle, wat maakt dat ik uiteindelijk 78 studies (nogmaals) beoordeel. Van die 78 studies kan ik er van 14 geen PDF downloaden en beoordeel ik er 44 als ongeschikt. De specifieke reden is als volgt:

1. Gaat niet over kanker (N=1).
2. Geen risicoschattingen opgenomen (N=28).
3. Geen meta-analyse (N=4).
4. Gaat niet over glyfosaat (N=7).
5. Geen schattingen voor glyfosaat (N=1).
6. Gaat niet over mensen (N=2).
7. Gaat niet over kanker (N=1).

Uiteindelijk heb ik 20 studies opgenomen voor data-extractie. Ik kan nu terug naar de derde zoektocht die zicht richt op dieren. Omdat ik vanuit de eerste twee zoektochten geen enkele meta-analyse heb gevonden behalve de bekende studie van Portier(1), ben ik maar gaan zoeken naar losse dierstudies om mijn voorheen gerapporteerde meta-analyse te herhalen. Uiteindelijk heb ik maar één studie gevonden die aan de inclusiecriteria voldoet en dat is de eerdergenoemde Ramazzini studie(3). Ik zal de verdere verwerking van die data later apart bespreken.

De focus van wat nu volgt heeft dus alleen betrekking op de studies uitgevoerd bij mensen en de 20 reviews die ik heb geïncludeerd(32-51) :

- 1 P. J. Mink, 2012
- 2 L. Zhang, 2019
- 3 G. C. Kabat, 2021
- 4 J. Acquavella, 2023
- 5 E. T. Chang, 2016
- 6 J. Acquavella, 2016
- 7 D. D. Alexander, 2007
- 8 D. D. Weisenburger, 2021
- 9 F. Donato, 2020
- 10 P. Boffetta, 2021
- 11 A. Khan, 2022
- 12 L. Schinasi, 2014

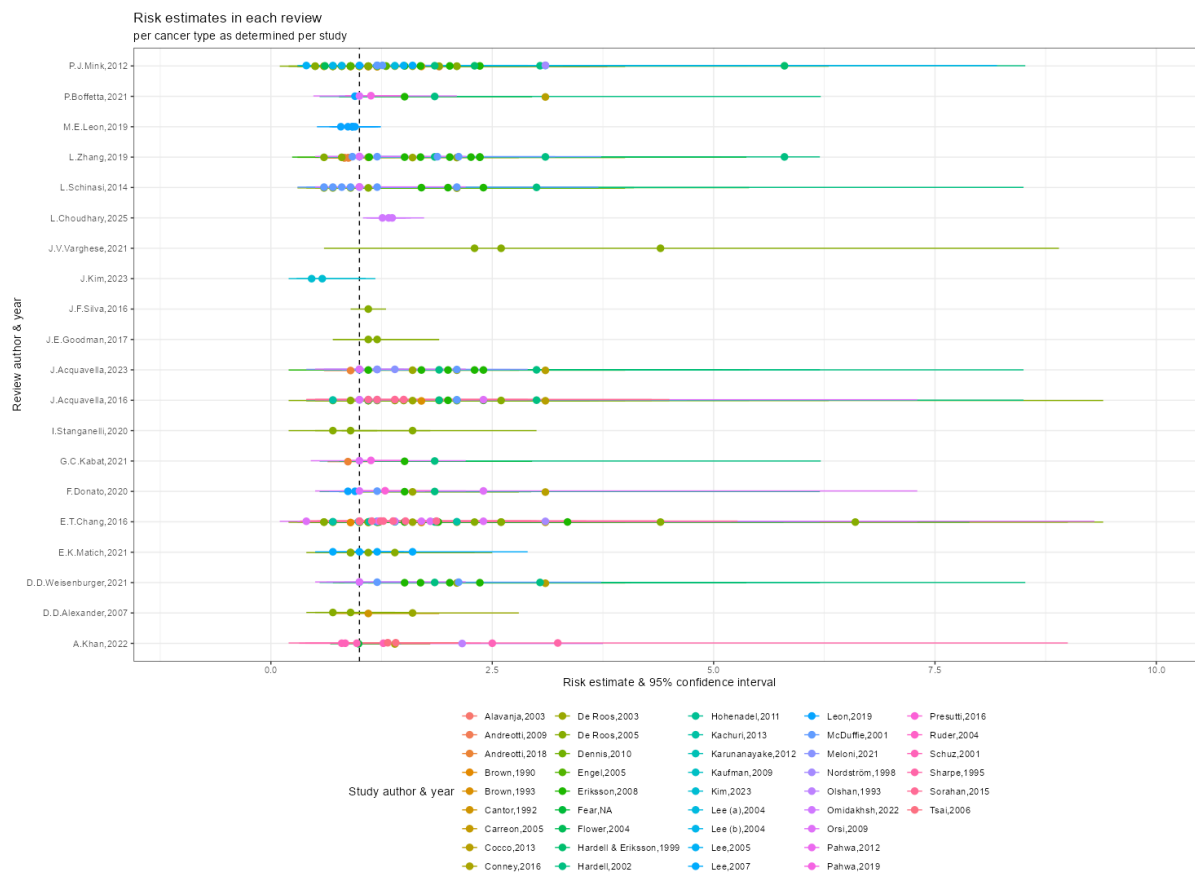
13 E.K.Matich,2021
14 J.V.Varghese,2021
15 I.Stanganelli,2020
16 J.E.Goodman,2017
17 J.F.Silva,2016
18 J.Kim,2023
19 M.E.Leon,2019
20 L.Choudhary,2025

Op basis van die die 20 reviews heb ik een Excel database gebouwd waarin de data van 42 publicaties staan beschreven³⁵ (50, 52-91):

1 Alavanja,2003
2 Andreotti,2009
3 Andreotti,2018
4 Brown,1990
5 Brown,1993
6 Cantor,1992
7 Carreon,2005
8 Cocco,2013
9 Conney,2016
10 De Roos,2003
11 De Roos,2005
12 Dennis,2010
13 Engel,2005
14 Eriksson,2008
15 Fear,NA
16 Flower,2004
17 Hardell,2002
18 Hardell & Eriksson,1999
19 Hohenadel,2011
20 Kachuri,2013
21 Karunanayake,2012
22 Kaufman,2009
23 Kim,2023
24 Lee,2005
25 Lee,2007
26 Lee (a),2004
27 Lee (b),2004
28 Leon,2019
29 McDuffie,2001
30 Meloni,2021
31 Nordström,1998
32 Olshan,1993
33 Omidakhsh,2022
34 Orsi,2009
35 Pahwa,2012
36 Pahwa,2019
37 Presutti,2016
38 Ruder,2004
39 Schuz,2001
40 Sharpe,1995
41 Sorahan,2015
42 Tsai,2006

³⁵ Ik zeg nu expres 42 publicaties (en niet studies), omdat ik later zal laten zien dat verschillende studies leunen op hetzelfde cohort (maar dan op andere data) of een her-analyse zijn van dezelfde data. Dit maakt de interpretatie wat lastiger omdat het niet per se nieuwe data betreft (maar wellicht wel nieuwe informatie). Ook zal de lezer in de grafieken later zien dat er soms maar over 37 studies wordt gesproken, maar dat is alleen wanneer ik op basis van auteur kijk. In dit dossier hebben verschillende auteurs vaker gepubliceerd.

In totaal heb ik 291 rijen aan gegevens geëxtraheerd waarbij elke rij één schatting is van de relatie tussen glyfosaat en kanker³⁶. Een eerste grafiek die alle risicoschattingen per studie per review laat zien we in **Figuur 2**.



Figuur 2. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat. Verschillende reviews hebben verschillende studies opgenomen.

Welke informatie missen we?

Als we kijken naar de gegevens die zijn opgenomen in de database om te bepalen hoeveel data we hebben voor elk van de 26 kolommen, dan kan ik met de volgende grafiek (**Figuur 3**)³⁷ goed aanduiden wat we wel en niet hebben.

Laten we eerst beginnen met de reviews. Hoewel het mijn doel was om voornamelijk meta-analyses te includeren, blijken negen(32, 35, 37-39, 44, 45, 48, 51) van de 20 studies

³⁶ Nogmaals, dit zijn niet allemaal unieke gegevens in de zin dat het schattingen zijn van andere studies. Het zijn 291 regels data met voor een deel gedeelde informatie.

³⁷ De afbeeldingen zijn expres in het Engels opgeleverd om ervoor te zorgen dat het rapport ook internationaal leesbaar is. Hoewel tooling zoals [DeepL](#) prima in staat is om tekst te vertalen wil het met figuren nog wel eens mis gaan en ik ga er gemakshalve van uit dat elke geïnteresseerde Nederlander voldoende Engels kan om met deze figuren uit de voeten te kunnen.

die stap nooit genomen te hebben. Verder weet ik van twee studies niet de titel (Olsan, Fear) en beiden zijn opgenomen in één enkele review (42) waarvan de referentielijst niet lijkt te kloppen.

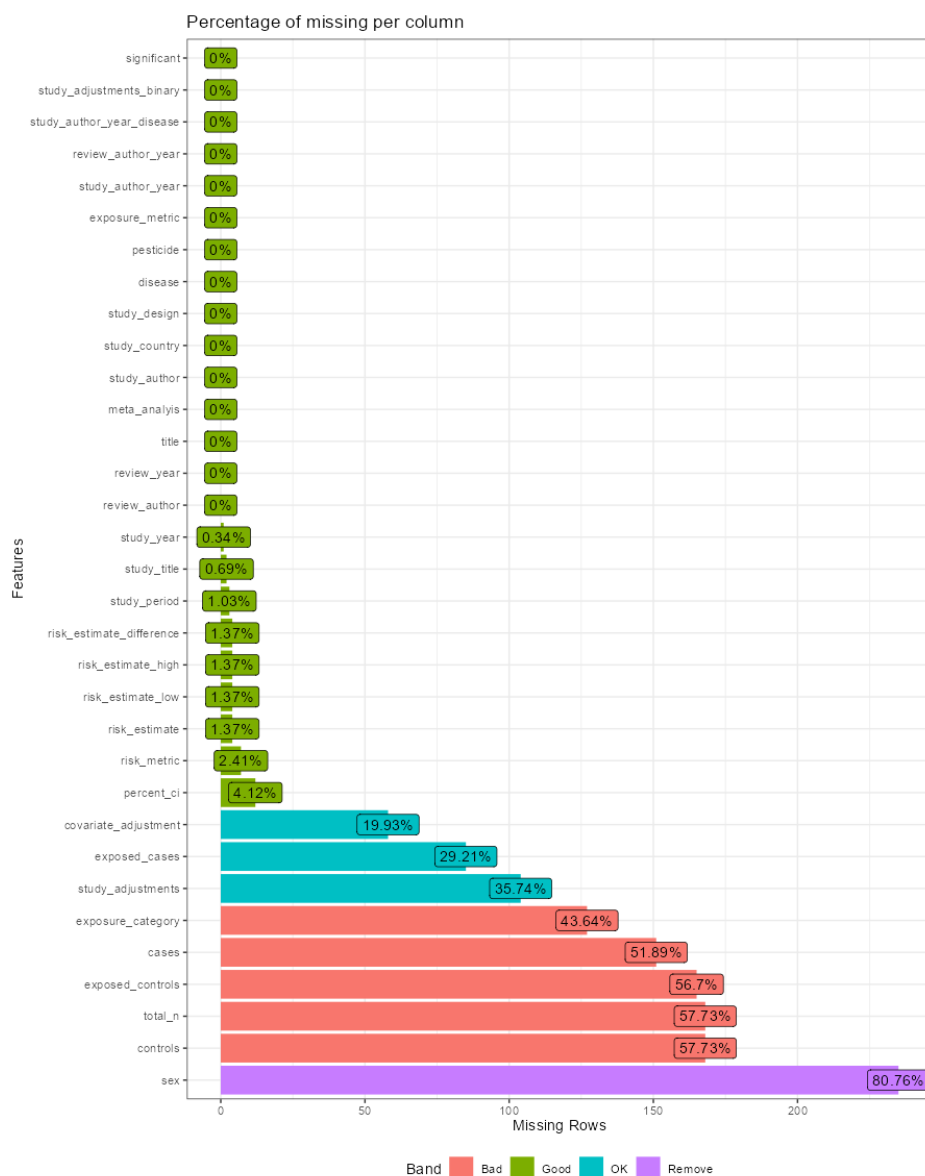
Wat ook opvalt is dat de meeste reviews de studiegrootte van hun opgenomen studies niet vermelden. Althans, ze doen dat niet expliciet genoeg door te vermelden hoe groot de risicogroep was, hoe groot de controlegroep was en hoeveel events er in elke groep heeft plaatsgevonden³⁸. Deze informatie is wel nodig om een *odds-ratio* uit te rekenen wat betekent dat ik de gerapporteerde *odds-ratio* niet steekproefsgewijs kon controleren. Ik moet dus maar aannemen dat wat gerapporteerd is ook klopt. Wat helpt is dat veel studies in meerdere reviews terugkomen, waardoor er eigenlijk altijd wel een betrouwbaar beeld ontstaat wat de risicoschatting is (of zou moeten zijn). Maar dit werkt dus alleen als je de moeite neemt om door het gehele landschap van studies te waden. Pak je lukraak één enkele review, dan vertekent dat waarschijnlijk je beeld, zelfs als het de meest recente review is. Daarvoor zijn er nog te veel vrijheidsgraden in het uitvoeren van een meta-analyse.

Een ander probleem wat zich voordoet door het ontbreken van informatie over de studiegrootte is dat ik nu essentiële informatie mis bij het uitvoeren van een meta-analyse. Dit komt omdat de studiegrootte een directe invloed heeft op het gewicht die een studie krijgt in de meta-analyse. Wederom kunnen we dan alleen maar aannemen dat het gewicht zoals deze is toegekend in de gepubliceerde meta-analyse correct is. Zelf narekenen kan moeizaam zijn (met een aantal assumpties) of lukt soms helemaal niet. Omdat ik uiteindelijk mijn eigen meta-analyses zal uitvoeren, en omdat elke studie 95% betrouwbaarheidsschattingen toont, kan ik hier (deels) omheen werken, maar dit is niet zonder uitdaging. Mijn pogingen en uitdagingen in het nabootsen van die schattingen komen overheen met de uitdagingen die ik ook heb beschreven in mijn [tweede](#) en [derde](#) rapport en zal ik daarom niet nogmaals uitgebreid beschrijven. Voor nu volstaat dat het uitvoeren van een meta-analyse op basis van gepubliceerde informatie geen eenvoudige exercitie is en die omslachtigheid zie je pas wanneer je er aan begint. Het identificeren van deze uitdagingen

³⁸ Ik beseef dat deze informatie het best past (en zelfs noodzakelijk is) voor *case-control* studies en niet zozeer cohort studies. Dus, als deze gegevens ontbreken in een cohort studie dan is dat geen ramp, maar wat je wel wil zien is hoe groot de cohort studie is én hoe vaak een bepaalde kankersoort is gevonden.

vertelt daarmee ook iets over de bewijskracht die we vanuit dit veld kunnen opzetten en/of het gebrek eraan.

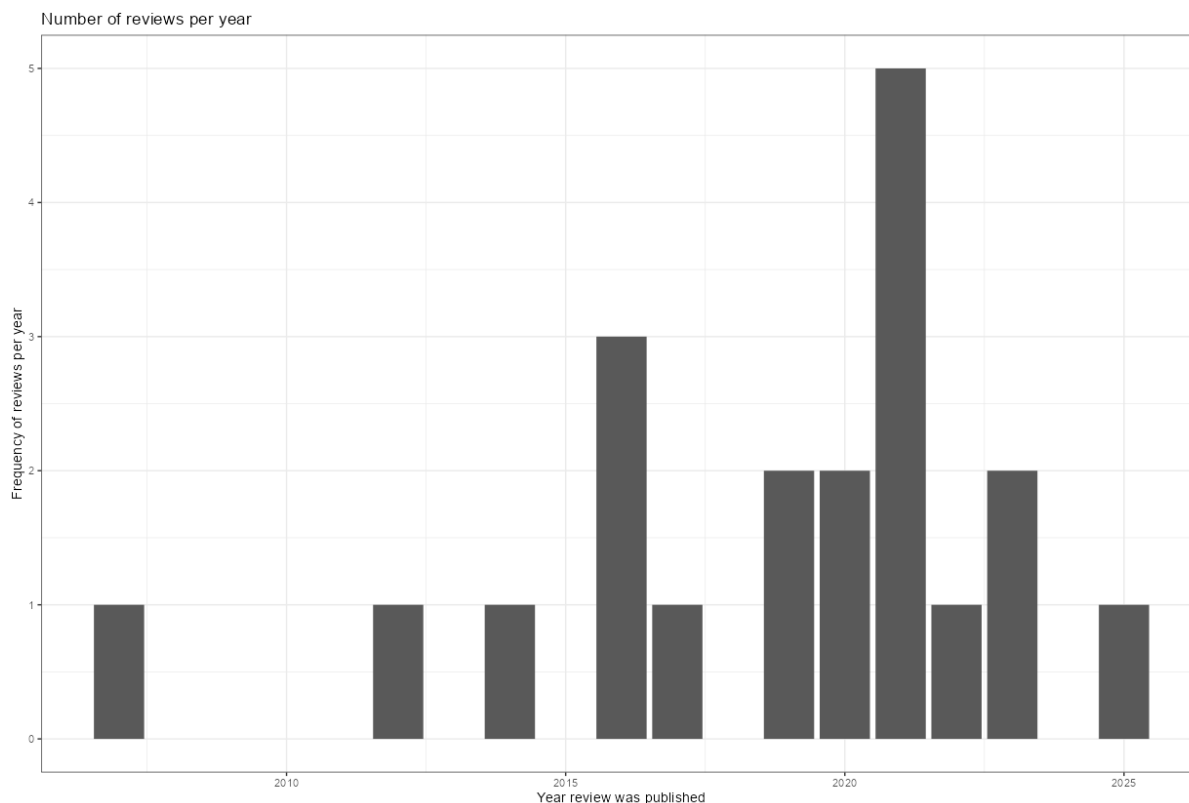
Andere informatie die veelvuldig mist is informatie met betrekking tot aanpassingen die eventueel wel of niet zijn gedaan. De informatie die ontbreekt voor geslacht (*sex*) is waarschijnlijk niet echt ontbrekende informatie: het is informatie die gewoon niet genoemd is omdat de studies zowel op mannen als op vrouwen betrekking heeft.



Figuur 3. Het aantal gegevens per kolom in de database op basis van 20 reviews en 42 publicaties.

Naar welke soorten kanker is er gekeken?

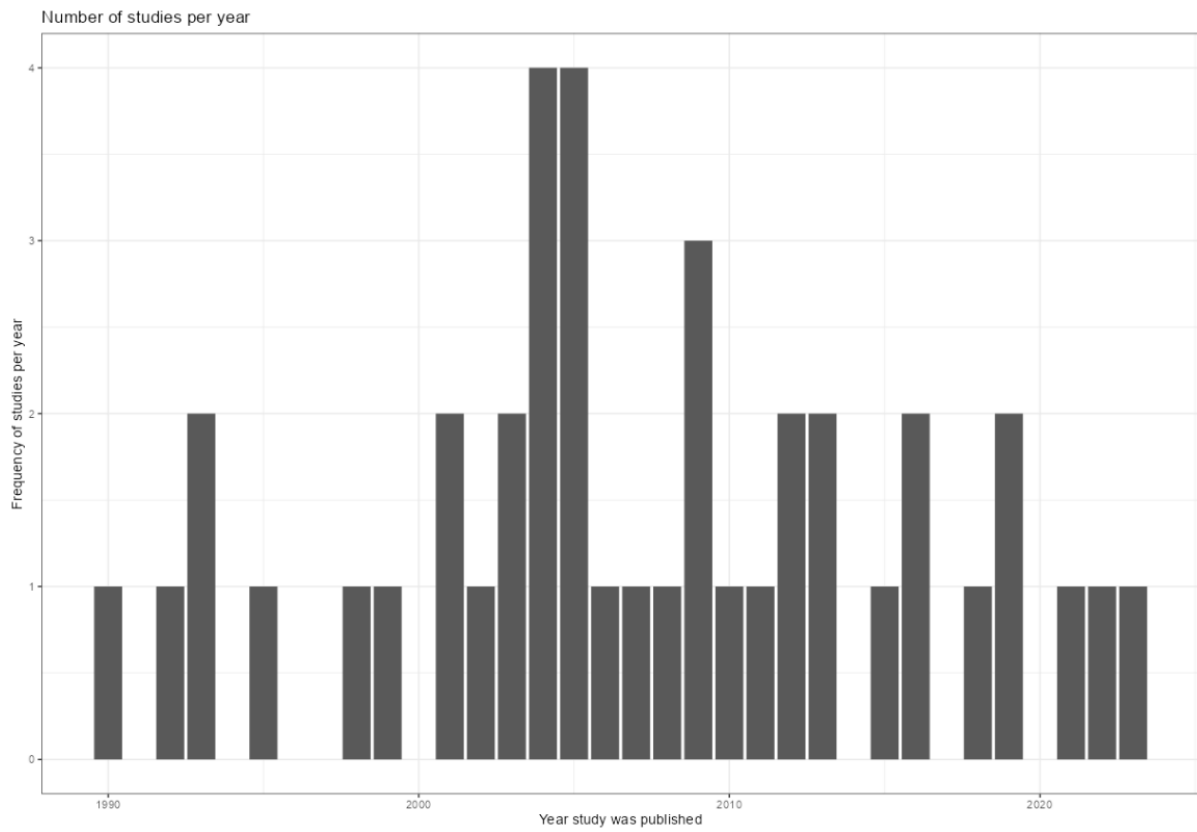
De reviews die zijn opgenomen in deze database zijn tussen 2007 en 2025 gepubliceerd (**Figuur 4**). De studies die zijn opgenomen zijn tussen 1990 en 2023 gepubliceerd: dat is dus ongeveer 33 jaar aan gegevens (**Figuur 5**). Het gros daarvan is na het jaar 2000 gepubliceerd.



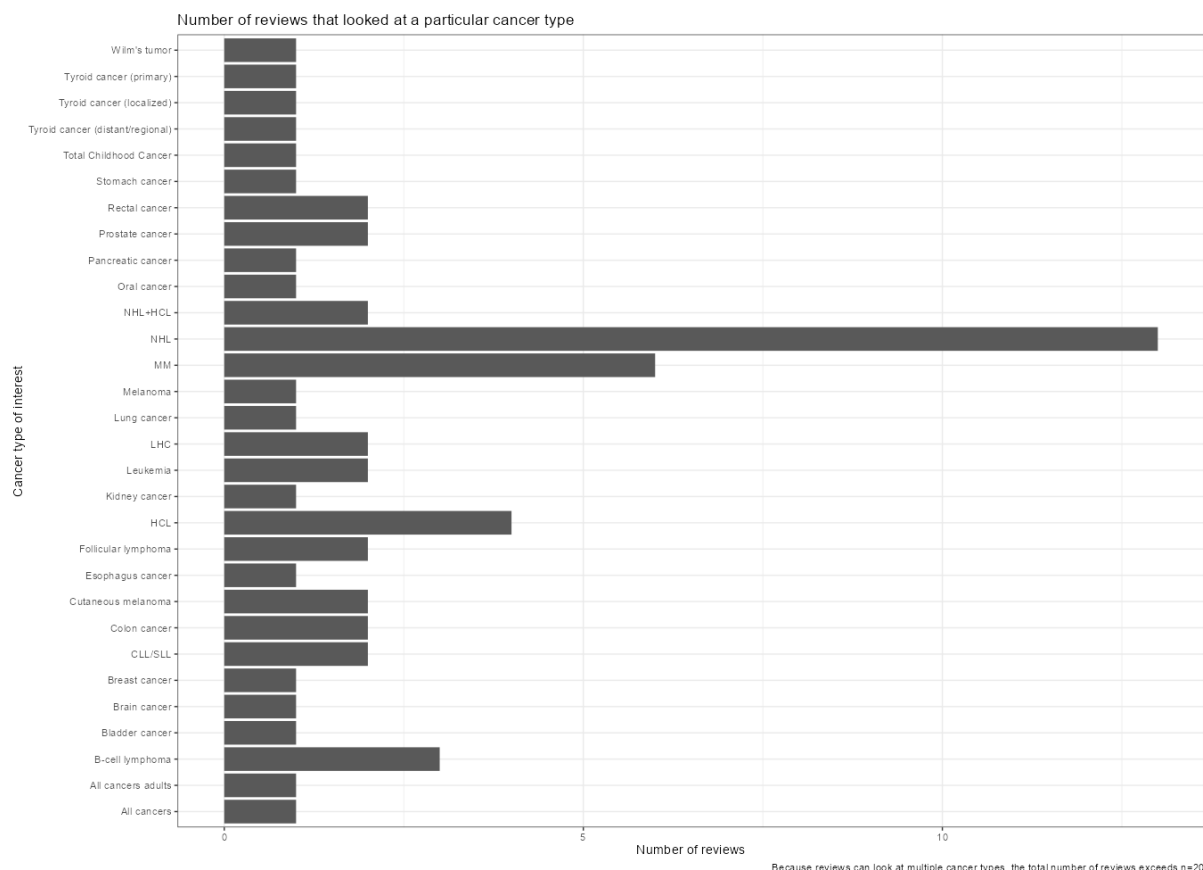
Figuur 4. Het aantal reviews per jaar.

Tezamen hebben deze reviews naar 30 verschillende soorten kanker gekeken (**Figuur 6**). De top 3 aan kankersoorten waarnaar onderzoek is gedaan is non-hodgkin lymfoom (NHL, N=13), beenmergkanker (*multiple myeloma* of MM, N=6) (32, 36, 37, 40, 45, 50) en hairy cell leukemie (HCL; N=4)(32, 33, 36, 90). Zo hebben de volgende reviews aandacht aan non-hodgkin lymfoom besteed ((32-41, 43, 47, 50):

- 1 L. Zhang, 2019
- 2 G.C.Kabat, 2021
- 3 J.Acquavella, 2023
- 4 P.J.Mink, 2012
- 5 E.T.Chang, 2016
- 6 D.D.Alexander, 2007
- 7 D.D.Weisenburger, 2021
- 8 F.Donato, 2020
- 9 P.Boffetta, 2021
- 10 J.Acquavella, 2016
- 11 L.Schinasi, 2014
- 12 J.E.Goodman, 2017



Figuur 5. Het aantal studies gepubliceerd per jaar.



Figuur 6. Het aantal reviews dat naar een bepaalde risicofactor heeft gekeken.

Als ik vervolgens kijk naar welke studies er allemaal voor NHL zijn, dan zie ik de volgende lijst(50, 54, 57, 59-61, 64, 66-68, 73, 76, 77, 80, 81):

- 1 Andreotti, 2018
- 2 Cantor, 1992
- 3 Cocco, 2013
- 4 De Roos, 2003
- 5 De Roos, 2005
- 6 Eriksson, 2008
- 7 Hardell, 2002
- 8 Hardell & Eriksson, 1999
- 9 Hohenadel, 2011
- 10 Lee (a), 2004
- 11 Leon, 2019
- 12 McDuffie, 2001
- 13 Meloni, 2021
- 14 Orsi, 2009
- 15 Pahwa, 2019

Wanneer we elke type kanker afzetten tegen de jaren waarover er is gepubliceerd kunnen we de volgende grafiek maken voor het aantal reviews (**Figuur 7**) en het aantal studies (**Figuur 8**). We kunnen dus, zoals verwacht, zien dat sommige kankersoorten een continue vorm van aandacht krijgt, terwijl er voor andere kankersoorten minder ruimte is³⁹. Wat we

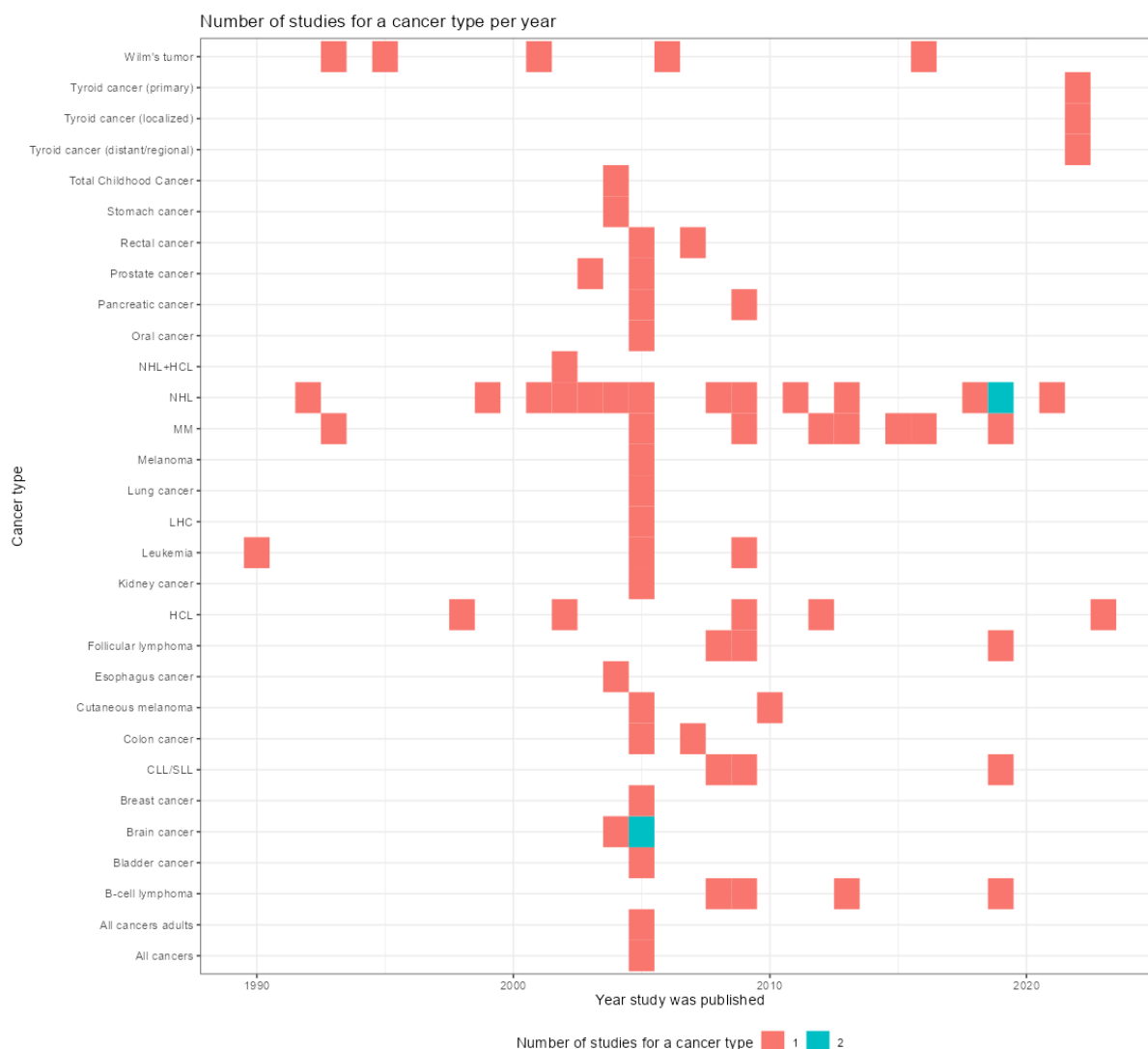
³⁹ Of het is minder makkelijk om er onderzoek naar te doen vanwege de ziekteprogressie.

ook kunnen zien is dat er reviews zijn waarin er maar naar één soort kanker is gekeken, terwijl andere reviews een hele lijst van kankersoorten tot zich hebben genomen. Dit is waarschijnlijk de reden waarom het gros van de kankersoorten in maar één of twee reviews is opgenomen. De totale lijst van het aantal reviews per kankersoort staat hieronder beschreven:

```
# A tibble: 30 x 2
  disease                n
  <chr>                  <int>
1 NHL                    13
2 MM                     6
3 HCL                    4
4 B-cell lymphoma       3
5 CLL/SLL                2
6 Colon cancer          2
7 Cutaneous melanoma    2
8 Follicular lymphoma   2
9 LHC                   2
10 Leukemia              2
11 NHL+HCL               2
12 Prostate cancer       2
13 Rectal cancer         2
14 All cancers           1
15 All cancers adults    1
16 Bladder cancer        1
17 Brain cancer          1
18 Breast cancer         1
19 Esophagus cancer      1
20 Kidney cancer         1
21 Lung cancer           1
22 Melanoma              1
23 Oral cancer           1
24 Pancreatic cancer     1
25 Stomach cancer        1
26 Total Childhood Cancer 1
27 Thyroid cancer (distant/regional) 1
28 Thyroid cancer (localized) 1
29 Thyroid cancer (primary) 1
30 Wilm's tumor          1
```



Figuur 7. Het jaar waarin een review is gepubliceerd per kankersoort die is onderzocht in de review.



Figuur 8. Het jaar waarin een studie is gepubliceerd per kankersoort die is onderzocht in de review.

Hoewel er dus naar een divers aantal kankersoorten is gekeken kunnen we eigenlijk nu al zeggen dat er maar een aantal zijn waar echt voldoende informatie voorhanden is. Een wellicht betere lijst is er eentje waar we niet het aantal reviews, maar het aantal studies voor elke kankersoort laten zien. Die lijst ziet er als volgt uit:

```
# A tibble: 30 x 2
  disease          n
  <chr>          <int>
1 NHL             15
2 MM              8
3 wilm's tumor   6
4 HCL            5
5 B-cell lymphoma 4
6 Brain cancer   3
7 CLL/SLL        3
8 Follicular lymphoma 3
9 Leukemia       3
10 Colon cancer  2
11 Cutaneous melanoma 2
12 Pancreatic cancer 2
```

13	Prostate cancer	2
14	Rectal cancer	2
15	All cancers	1
16	All cancers adults	1
17	Bladder cancer	1
18	Breast cancer	1
19	Esophagus cancer	1
20	Kidney cancer	1
21	LHC	1
22	Lung cancer	1
23	Melanoma	1
24	NHL+HCL	1
25	Oral cancer	1
26	Stomach cancer	1
27	Total Childhood Cancer	1
28	Tyroid cancer (distant/regional)	1
29	Tyroid cancer (localized)	1
30	Tyroid cancer (primary)	1

Ook nu zien we duidelijk dat NHL met kop en schouder de meest onderzochte soort kanker is. Waarom dat zo is, is mij niet direct duidelijk, maar het zou goed kunnen zijn dat er tijdens de studieopzet gewoonweg een keuze is gemaakt naar het soort kankers dat men zou onderzoeken. Of men verwacht hier de meest duidelijke relatie met glyfosaat te vinden. Of dit was het soort kanker waar de meeste data voorhanden was. Hoe en waarom er voor deze soort zo vaak is gekozen blijft onduidelijk, maar het is wel een belangrijke bevinding als we straks naar de dierstudies uitwijken.

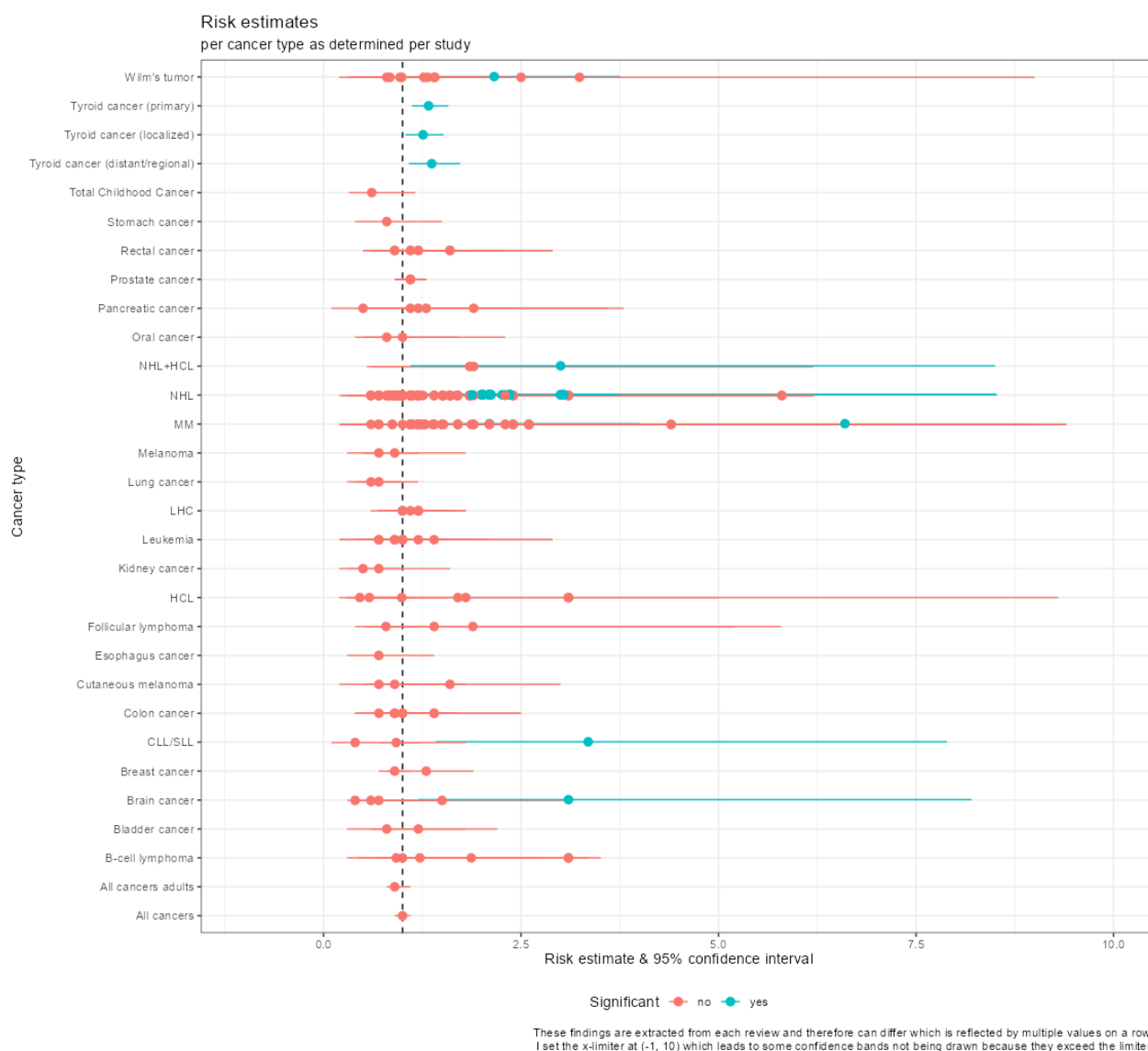
In **Figuur 9** kunnen we alle gevonden schattingen per kankersoort zien en of deze statistisch significant zijn of niet. Ook hier is duidelijk te zien dat er voor sommige kankersoorten veel schattingen zijn en voor andere soorten een stuk minder. Het gros van de schattingen is trouwens niet statistisch significant. Wat verder opvalt is dat veel schattingen aan de linkerkant van de lijn-van-geen-effect liggen⁴⁰. Dat lijkt te wijzen op een beschermend effect van glyfosaat en over die discussie heb ik al eerder [gerapporteerd](#). Maar om eerlijk te zijn wijst het in mijn optiek eerder op de inherente problemen van het schatten van risico's door middel van omgevingsepidemiologie, maar daarover in de discussie meer.

Kenmerken van de reviews

Zoals al eerder genoemd heb ik er deze keer voor gekozen om niet expliciet in de kenmerken van de meta-analyses te duiken. Dat is niet omdat ik geen interesse heb in de meta-analyse per se, maar omdat de kracht ligt in het werken met de studies die ze hebben meegenomen.

⁴⁰ Bij een *odds-ratio* of *risk-ratio* ligt de lijn-van-geen-effect op de 1 – een ratio van 1 betekent namelijk geen verschil tussen groepen.

In tegenstelling tot mijn onderzoek naar risicofactoren voor de ziekte van Parkinson heb ik nu dus niet naar het type meta-analyse gekeken en of er gecorrigeerd is voor bepaalde onderdelen. Dat komt omdat deze factoren uiteindelijk niet van belang leken en ik had in die studie heel wat meer data voorhanden dan nu. Deze bevinding sterkt mij in de keuze om uitsluitend naar de kenmerken van de studies te kijken en de verschillen tussen reviews te beschrijven.



Figuur 9. Risico schattingen per kankersort en of een schatting statistisch significant is of niet.

Kenmerken van de studies

Beter is het dus om in de kenmerken van de studies zelf te duiken. Dat doet we eerst door de gegevens in tabel format tot ons te nemen. Daarna kunnen we een start maken door de data grafisch door te spitten. Laten we eerst eens kijken waar de studies allemaal vandaan komen. In 20 reviews en 42 publicaties zijn in totaal 12 verschillende landen opgenomen waarvan het merendeel in USA(52-58, 60-63, 65, 72, 73, 75, 79, 84, 86, 87, 92):

```
# Groups:   study_country [12]
  study_country      n
  <chr>            <int>
1 USA                20
2 Canada             5
3 Sweden             4
4 International      3
5 North-America     2
6 Noth America      2
7 Brazil            1
8 France            1
9 Italy              1
10 Thailand          1
11 UK                1
12 West-Germany     1
```

We hebben geen studies uit Nederland opgenomen in deze reviews, omdat die er gewoonweg niet zijn⁴¹. In totaal zijn er twee verschillende soorten *study designs* opgenomen in de volgende verdeling:

```
study_design      n
<chr>            <int>
1 Case-control    30
2 Cohort          12
```

Het gros van de studies heeft, zoals verwacht, gebruikt maakt van een *case-control* design. Dan volgt het *cohort* design. Beiden zijn observationele studies(93) en echt experimentele designs zien we hier niet⁴².

We hebben eerder al gezien hoeveel studies we per kankersoort hebben. Nu is het tijd om te zien hoeveel kankersoorten er per studie werden onderzocht. Die top 10 ziet er als volgt uit:

```
# A tibble: 42 x 2
# Groups:   study_author_year [42]
  study_author_year      n
  <chr>                <int>
1 De Roos,2005         16
2 Orsi,2009            6
3 Leon,2019            5
4 Eriksson,2008       4
5 Harde11,2002        3
```

⁴¹ Wat deels haaks staat op de grote media-aandacht voor glyfosaat in Nederland.

⁴² Dit in tegenstelling tot de dierstudies die in feite allemaal experimenteel van aard zijn.

6	Omidakhsh, 2022	3
7	Cocco, 2013	2
8	Lee (b), 2004	2
9	Lee, 2007	2
10	Alavanja, 2003	1

Het lijkt er sterk op dat er één enkele studie (De Roos, 2005)(60) is die meer dan de helft (16/30) van alle onderzochte kankersoorten tot zich neemt. Dit betekent waarschijnlijk ook dat van veel kankersoorten er maar heel weinig studies zijn en in **Figuur 9** zien we dat ook wel. Laten we daarom eens dieper duiken in de studie van De Roos(60) en zien welke kankersoorten er precies zijn meegenomen:

```
# A tibble: 16 x 1
  disease
  <chr>
1 MM
2 LHC
3 NHL
4 All cancers adults
5 Bladder cancer
6 Colon cancer
7 Kidney cancer
8 Leukemia
9 Lung cancer
10 Melanoma
11 Oral cancer
12 Pancreatic cancer
13 Prostate cancer
14 Rectal cancer
15 All cancers
16 Cutaneous melanoma
```

Als we kijken naar de vermelde studiekeerperiode dan zien dat, op twee studies na, die informatie voorhanden is. Uit analyses blijkt dat deze periodes aardig van elkaar kunnen verschillen en ook dat de reviews het niet altijd met elkaar eens zijn over de exacte studiekeerperiode:

```
A tibble: 42 x 2
  study_author_year      study_period
  <chr>                <chr>
1 Fear,NA              1962-1999
2 Flower,2004          1975-1998
3 De Roos,2003         1979-1986
4 Lee (a),2004         1980-1986
5 Brown,1990           1981-1984
6 Brown,1993           1981-1984
7 Cantor,1992          1981-1984
8 Olshan,1993          1984-1996
9 Sharpe,1995          1987-1989
10 Hardell,2002         1987-1992
11 Nordström,1998      1987-1992
12 Hardell & Eriksson,1999 1987-1995
13 Lee,2005             1988-1994
14 Lee (b),2004         1988-1994
15 Schuz,2001          1988-2000
16 Hohenadel,2011      1991-1994
17 Kachuri,2013        1991-1994
18 Karunanayake,2012   1991-1994
```

19	McDuffie, 2001	1991-1994
20	Pahwa, 2012	1991-1994
21	Tsai, 2006	1992-2006
22	Alavanja, 2003	1993-2001
23	De Roos, 2005	1993-2001
24	Dennis, 2010	1993-2001
25	Engel, 2005	1993-2001
26	Sorahan, 2015	1993-2001
27	Andreotti, 2009	1993-2002
28	De Roos, 2005	1993-2002
29	Lee, 2007	1993-2002
30	Kim, 2023	1993-2011
31	Leon, 2019	1993-2011
32	Leon, 2019	1993-2012
33	Andreotti, 2018	1993-2013
34	Carreon, 2005	1995-1997
35	Ruder, 2004	1995-1997
36	Kaufman, 2009	1997-2003
37	Cocco, 2013	1998-2004
38	Eriksson, 2008	1999-2002
39	Conney, 2016	1999-2007
40	Omidakhsh, 2022	1999-2012
41	Orsi, 2009	2000-2004
42	Meloni, 2021	2012-2016

Het gros van de studies keek naar een periode van ongeveer drie jaar. Sommigen daarentegen duurden meer dan 10 jaar met één uitschieter van 37 jaar:

period	n
<dbl>	<int>
1	3
2	8
3	6
4	2
5	9
6	4
7	5
8	12
9	18
10	7
11	13
12	14
13	19
14	20
15	23
16	37

Als we een verdere splitsing maken op basis van het soort studie, zien we de volgende indeling die toont dat de studie periode bij *cohort* studies meestal groter is dan bij *case-control* studies⁴³:

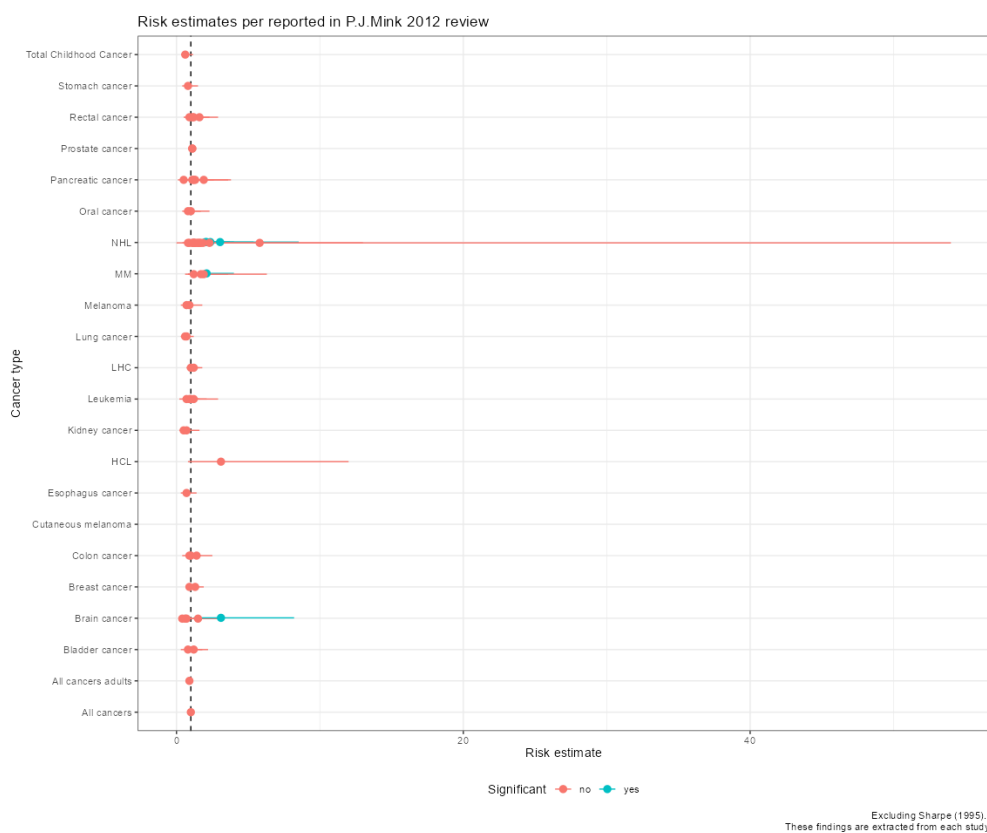
study_design	period	n
<chr>	<dbl>	<int>
1 Cohort	8	5
2 Cohort	9	3
3 Cohort	18	2
4 Cohort	19	1
5 Cohort	20	1
6 Cohort	23	1

⁴³ In principe geen vreemde bevinding als je bedenkt dat *case-control* studies retrospectieve momentopnames zijn, terwijl de meeste cohort studies prospectief en longitudinaal van opzet zijn. Wat betekent dat de *case-control* studie kijkt wat er NU aan ziekte voorhanden is, terwijl de cohort studie onderzoekt wat er in de toekomst KAN gebeuren op basis van mogelijke risico's.

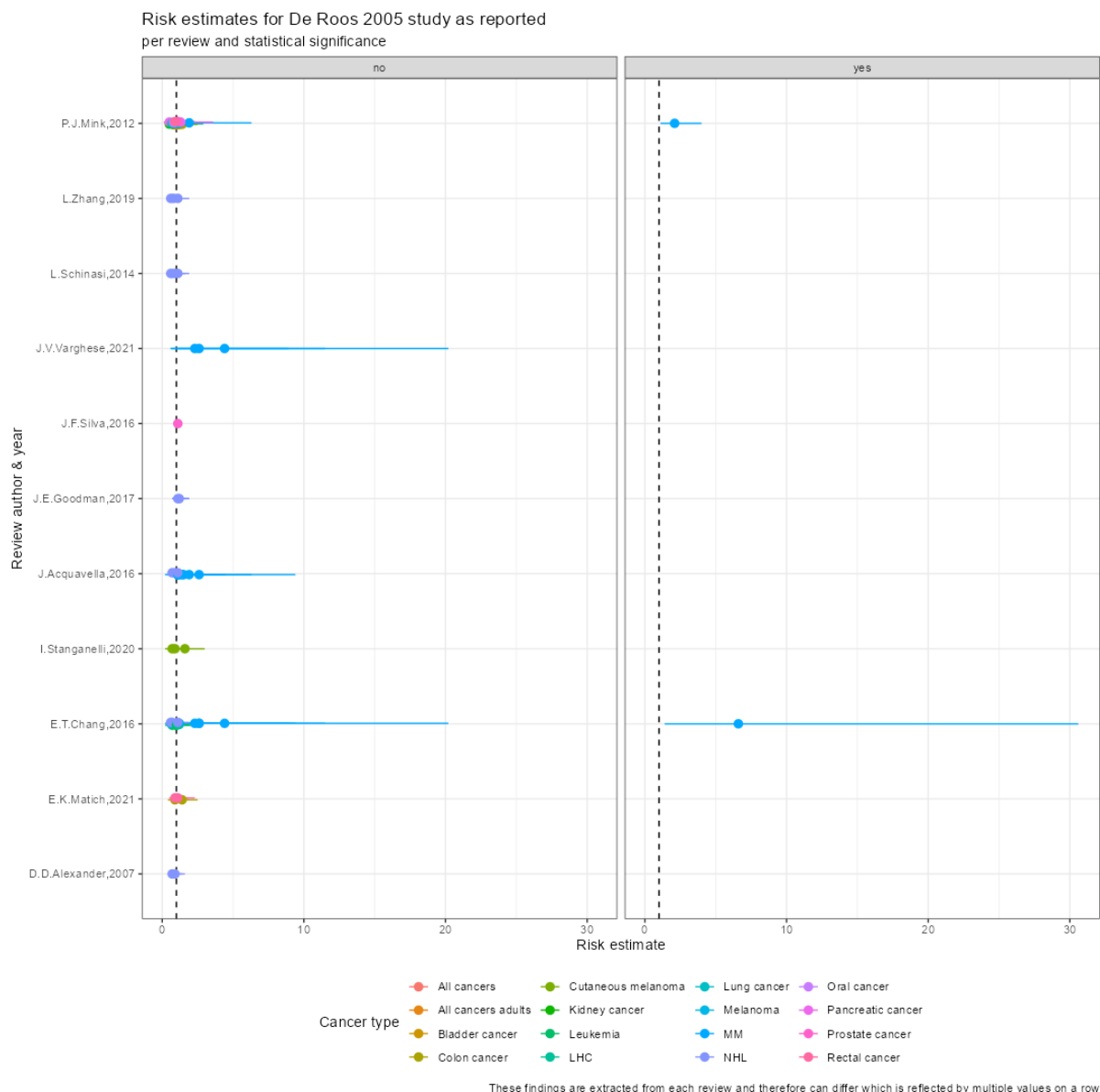
7	Case-control	2	3
8	Case-control	3	9
9	Case-control	4	2
10	Case-control	5	2
11	Case-control	6	5
12	Case-control	7	1
13	Case-control	8	2
14	Case-control	12	2
15	Case-control	13	1
16	Case-control	14	1
17	Case-control	37	1

Verschillen tussen reviews met dezelfde studie

We zagen al dat reviews kunnen verschillen in de informatie die ze over een bepaalde studie vermelden. Dit is heel belangrijk omdat het iets verteld over de betrouwbaarheid van de reviews en de bruikbaarheid van de informatie. Daarom is het goed dat we dieper in deze verschillen duiken en dat kunnen we wellicht het gemakkelijkst doen door te beginnen met de studie die het vaakst in een review is opgenomen. In onze eigen database hebben we de meeste informatie verkregen uit de review van Mink(32) (**Figuur 10**) en de studie van De Roos(60) (**Figuur 11**).



Figuur 10. Risicoschattingen per kankersoort (en per studie) in de review van Mink(32).



Figuur 11. Risicoschattingen uit de studie van De Roos(60) zoals gerapporteerd per review.

Een snelle analyse laat zien dat de studie van De Roos(60) in maar liefst 11 reviews is opgenomen. Daarmee voert deze studies de top 10 van meest opgenomen studies aan (50, 54, 56, 59-61, 64, 66, 76, 80):

study_author_year	n
<chr>	<int>
1 De Roos, 2005	11
2 Eriksson, 2008	10
3 Hardell, 2002	10
4 De Roos, 2003	9
5 Orsi, 2009	9
6 McDuffie, 2001	8
7 Cocco, 2013	6
8 Leon, 2019	4
9 Andreotti, 2018	3
10 Brown, 1993	3

Maar omdat we 30 verschillende kankersoorten hebben opgenomen kunnen we niet verwachten dat elk van de 16 kankersoorten die zijn opgenomen in de studie van De Roos(60) ook in elke review is opgenomen. We zullen dus nog specifiekere moeten zijn om te bepalen van welke studie welke kankersoort het vaakst is beschreven in de reviews.

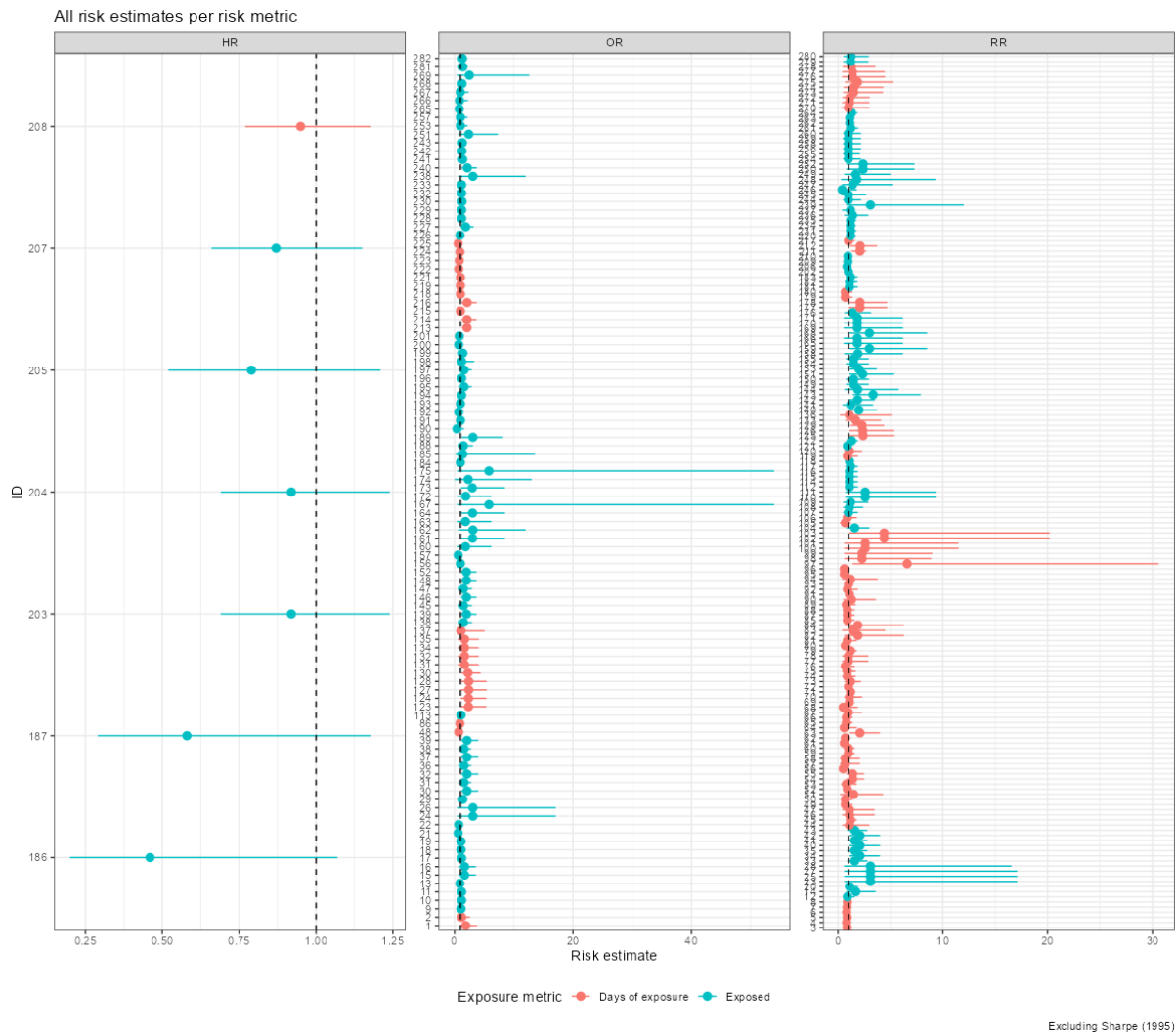
Uiteindelijk krijg ik de volgende lijst voor de studie van De Roos(60) wat betekent dat de NHL data uit De Roos in zeven reviews(32, 33, 36-38, 43, 47) is opgenomen:

disease <chr>	n <int>
1 NHL	7
2 MM	4
3 Colon cancer	2
4 LHC	2
5 Leukemia	2
6 Prostate cancer	2
7 Rectal cancer	2
8 All cancers	1
9 All cancers adults	1
10 Bladder cancer	1
11 Cutaneous melanoma	1
12 Kidney cancer	1
13 Lung cancer	1
14 Melanoma	1
15 Oral cancer	1
16 Pancreatic cancer	1

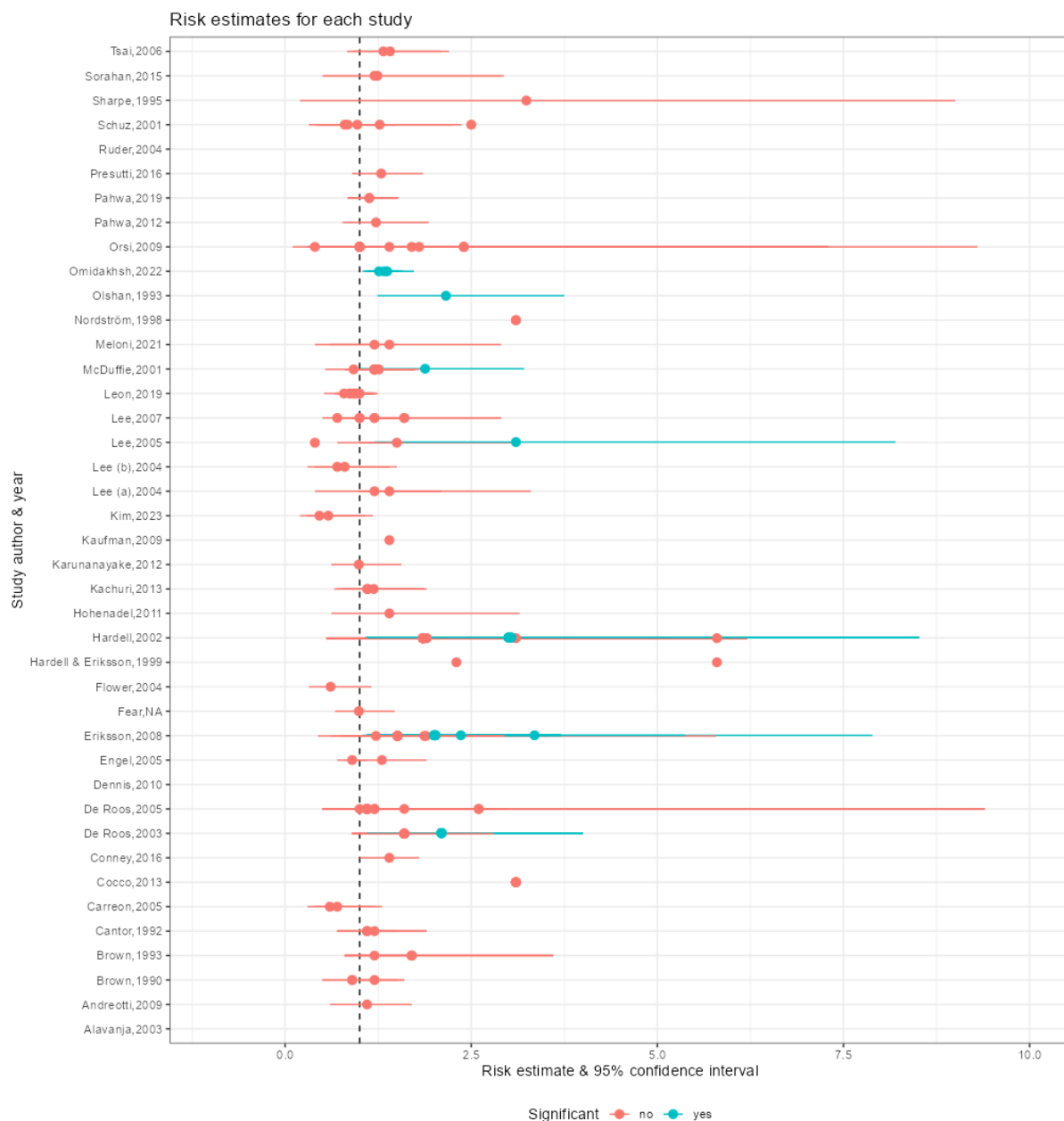
1 J.Acquavella,2016
2 L.Schinasi,2014
3 E.T.Chang,2016
4 D.D.Alexander,2007
5 P.J.Mink,2012
6 L.Zhang,2019
7 J.E.Goodman,2017

Ik kan dus nu aan de slag om te zien waar deze reviews verschillen in hun reportage van de gegevens voor NHL uit De Roos(60). Dit geeft mij belangrijke informatie over de kwaliteit van de rapportage én de betrouwbaarheid van de gerapporteerde gegevens. Maar voor ik deze gegevens toon, is het belangrijk om stil te staan bij twee andere kolommen, namelijk: *exposure metric* (**Figuur 12**) en *exposure category*. De eerste variabele zegt iets over de soort van blootstelling (*Exposed / Days of exposure*), terwijl de tweede iets zegt over de mate van blootstelling. Dat kan binair zijn (wel/niet blootgesteld – zie **Figuur 13** en **Figuur 14** voor een voorbeeld) of te maken hebben met het aantal dagen aan blootstelling (zie **Figuur 14**, **Figuur 15** en **Figuur 16** voor een voorbeeld)⁴⁴.

⁴⁴ Een andere onderzoeker zou de kolommen wellicht anders genoemd hebben, maar voor mij leken dit de meest zinvolle beschrijvingen om aan te tonen of de mate van blootstelling binair of niet-binair werd gemeten, en als niet-binair hoe deze vervolgens werd ingedeeld om een zinvolle uitspraak te doen over de invloed die de mate van blootstelling zou kunnen hebben op het risico voor kanker.



Figuur 12. Risicoschattingen per studie waarbij onderscheid wordt gemaakt tussen het type statistiek en type blootstelling.



Focused on exposed vs non-exposed. These findings are extracted from each review and therefore can differ which is reflected by multiple values on a row. I set the x-limiter at (-1, 10) which leads to some confidence bands not being drawn because they exceed the limiter.

Figuur 13. Risicoschattingen per studie waarbij onderscheid wordt gemaakt tussen wel of geen blootstelling.

De tweede variabele (*exposure category*) heeft te maken met het exacte aantal dagen waarnaar is gekeken⁴⁵. Als we kijken naar de studie van De Roos(60) dan zien we de volgende unieke combinaties voor beide variabelen:

exposure_metric	exposure_category
-----------------	-------------------

⁴⁵ Een andere onderzoeker zou de kolommen wellicht anders genoemd hebben, maar voor mij leken dit de meest zinvolle beschrijvingen om aan te tonen of de mate van blootstelling binair of niet-binair werd gemeten, en als niet-binair hoe deze vervolgens werd ingedeeld om een zinvolle uitspraak te doen over de invloed die de mate van blootstelling zou kunnen hebben op het risico voor kanker.

```

<chr> <chr>
1 Days of exposure 1-20 days
2 Days of exposure 21-56
3 Days of exposure 21-57
4 Days of exposure 337,2-18241
5 Days of exposure 57-2678
6 Days of exposure 79,6-337,1
7 Days of exposure Quartile 4
8 Days of exposure Tertile 1
9 Days of exposure Tertile 2
10 Days of exposure Tertile 3
11 Exposed NA
12 Days of exposure NA

```

Voor de variabele *exposure_metric* verwacht ik geen verdere informatie als er alleen maar gekeken is naar blootstelling in de vorm van wel/niet. Wanneer er wel werd gekeken naar de mate van blootstelling heb ik gezocht naar extra informatie en deze in de kolom *exposure_category* gezet. Die informatie was niet altijd voorhanden en soms ook niet echt behulpzaam, zoals wanneer er naar kwartielen of tertielen werd verwezen. Willen we zien of reviews verschillende waarden rapporteren dan zullen we dit moeten doen voor de combinatie studie, ziekte, *exposure metric* en *exposure category*. Dat ziet er voor de studie van De Roos(60) en de data voor NHL als volgt uit⁴⁶:

```

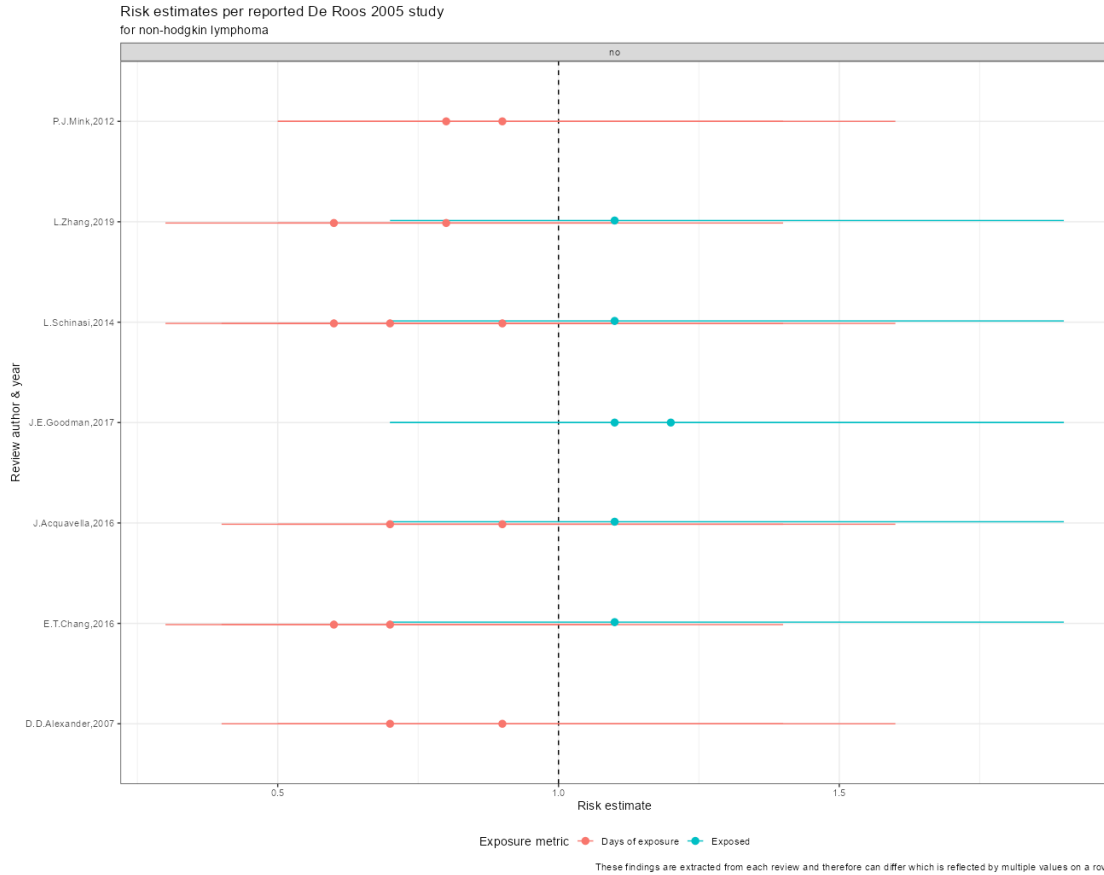
# A tibble: 19 x 6
# Groups:   exposure_metric, exposure_category [5]
  exposure_metric exposure_category review_author_year risk_estimate risk_estimate_low risk_estimate_high
  <chr>          <chr>          <chr>          <dbl>          <dbl>          <dbl>
1 Days of exposure 21-56 J. Acquavella, 2016 0.7 0.4 1.4
2 Days of exposure 21-56 L. Schinasi, 2014 0.7 0.4 1.4
3 Days of exposure 21-56 E. T. Chang, 2016 0.7 0.4 1.4
4 Days of exposure 21-56 D. D. Alexander, 2007 0.7 0.4 1.4
5 Days of exposure 337, 2-18241 P. J. Mink, 2012 0.8 0.5 1.4
6 Days of exposure 337, 2-18241 L. Zhang, 2019 0.8 0.5 1.4
7 Days of exposure 57-2678 P. J. Mink, 2012 0.9 0.5 1.6
8 Days of exposure 57-2678 J. Acquavella, 2016 0.9 0.5 1.6
9 Days of exposure 57-2678 L. Schinasi, 2014 0.9 0.5 1.6
10 Days of exposure 57-2678 D. D. Alexander, 2007 0.9 0.5 1.6
11 Days of exposure 79, 6-337, 1 L. Zhang, 2019 0.6 0.3 1.1
12 Days of exposure 79, 6-337, 1 L. Schinasi, 2014 0.6 0.3 1.1
13 Days of exposure 79, 6-337, 1 E. T. Chang, 2016 0.6 0.3 1.1
14 Exposed NA L. Zhang, 2019 1.1 0.7 1.9
15 Exposed NA J. Acquavella, 2016 1.1 0.7 1.9
16 Exposed NA L. Schinasi, 2014 1.1 0.7 1.9
17 Exposed NA E. T. Chang, 2016 1.1 0.7 1.9
18 Exposed NA J. E. Goodman, 2017 1.1 0.7 1.9
19 Exposed NA J. E. Goodman, 2017 1.2 0.7 1.9

```

We zien hierboven eigenlijk geen gekke zaken, maar om wat meer grip op de data te krijgen helpt het soms als we de cijfers ook afbeelden. Het resultaat is **Figuur 14**. We kunnen nu duidelijk zien dat waarden die te maken hebben met *days of exposure* zich allemaal aan de linkerkant van de verticale lijn bevinden, en risicoschattingen die te maken hebben met *exposed* aan de rechterkant te vinden zijn. Omdat de 95% betrouwbaarheidsintervallen van

⁴⁶ Het is belangrijk om hier toch nog even te vermelden wat ik ook al in de methode-sectie heb beschreven: als een review voor een bepaalde studie geen informatie beschreef, dan heb ik die informatie aangevuld, maar alleen wanneer de rest van de reviews exact dezelfde informatie toonde voor die studie. Dus als vier reviews exact dezelfde informatie aanboden voor de risicoschatting, dan heb ik de missende waarde voor de vijfde ingevuld op basis van die vier. Zodra er een verschil was tussen één van de vier heb ik dat niet meer gedaan. Op deze manier kon ik de lege cellen vullen met informatie die mij betrouwbaar leek, maar heb ik geen cellen blindelings overschreven. Het is aan de lezer om te controleren of dat naar behoren is gegaan.

haast elke schatting over die lijn heen gaat maakt het eigenlijk niet uit: ze zijn allemaal niet significant(94-97).

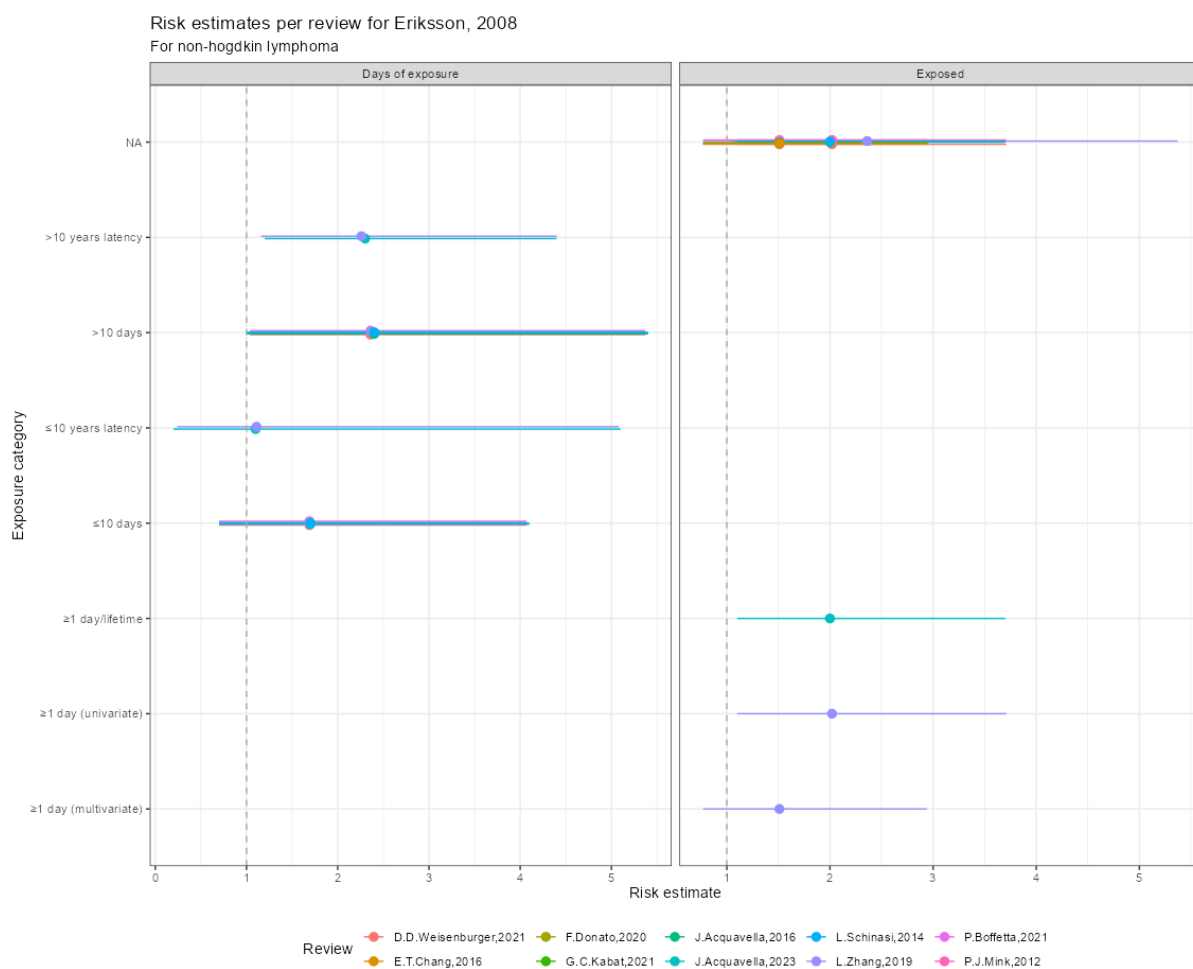


Figuur 14. Risicoschattingen voor de studie van De Roos(60), zoals gerapporteerd per review en per type blootstelling.

Eens kijken hoe het er bij de studie van Eriksson(64) uit ziet als we kijken naar alleen de data voor NHL. Hoewel deze informatie in minder reviews is opgenomen zien we een stuk meer informatie voor deze studie:

```
# A tibble: 29 x 6
# Groups:   exposure_metric, exposure_category [8]
  exposure_metric exposure_category review_author_year risk_estimate risk_estimate_low risk_estimate_high
  <chr>          <chr>          <chr>          <dbl>          <dbl>          <dbl>
1 Days of exposure >10 days D.D.Weisenburger,2021 2.36 1.04 5.37
2 Days of exposure >10 days P.J.Mink,2012 2.36 1.04 5.37
3 Days of exposure >10 days J.Acquavella,2023 2.4 1 5.4
4 Days of exposure >10 days L.Zhang,2019 2.36 1.04 5.37
5 Days of exposure >10 days J.Acquavella,2016 2.4 1 5.4
6 Days of exposure >10 days L.Schinasi,2014 2.4 1 5.4
7 Days of exposure >10 years latency J.Acquavella,2023 2.3 1.2 4.4
8 Days of exposure >10 years latency L.Zhang,2019 2.26 1.16 4.4
9 Days of exposure =10 days D.D.Weisenburger,2021 1.69 0.7 4.07
10 Days of exposure =10 days P.J.Mink,2012 1.69 0.7 4.07
11 Days of exposure =10 days J.Acquavella,2023 1.7 0.7 4.1
12 Days of exposure =10 days L.Zhang,2019 1.69 0.7 4.07
13 Days of exposure =10 days L.Schinasi,2014 1.7 0.7 4.1
14 Days of exposure =10 years latency J.Acquavella,2023 1.1 0.2 5.1
15 Days of exposure =10 years latency L.Zhang,2019 1.11 0.24 5.08
16 Exposed =1 day (multivariate) L.Zhang,2019 1.51 0.77 2.94
17 Exposed =1 day (univariate) L.Zhang,2019 2.02 1.1 3.71
18 Exposed =1 day/lifetime J.Acquavella,2023 2 1.1 3.7
19 Exposed NA D.D.Weisenburger,2021 1.51 0.77 2.94
20 Exposed NA D.D.Weisenburger,2021 2.02 1.1 3.71
21 Exposed NA P.J.Mink,2012 1.51 0.77 2.94
22 Exposed NA P.J.Mink,2012 2.02 1.1 3.71
23 Exposed NA F.Donato,2020 1.51 0.77 2.94
24 Exposed NA P.Boffetta,2021 1.51 0.77 2.94
25 Exposed NA L.Zhang,2019 2.36 1.04 5.37
26 Exposed NA J.Acquavella,2016 2 1.1 3.7
27 Exposed NA L.Schinasi,2014 2 1.1 3.7
28 Exposed NA G.C.Kabat,2021 1.51 0.77 2.95
29 Exposed NA E.T.Chang,2016 1.51 0.77 2.94
```

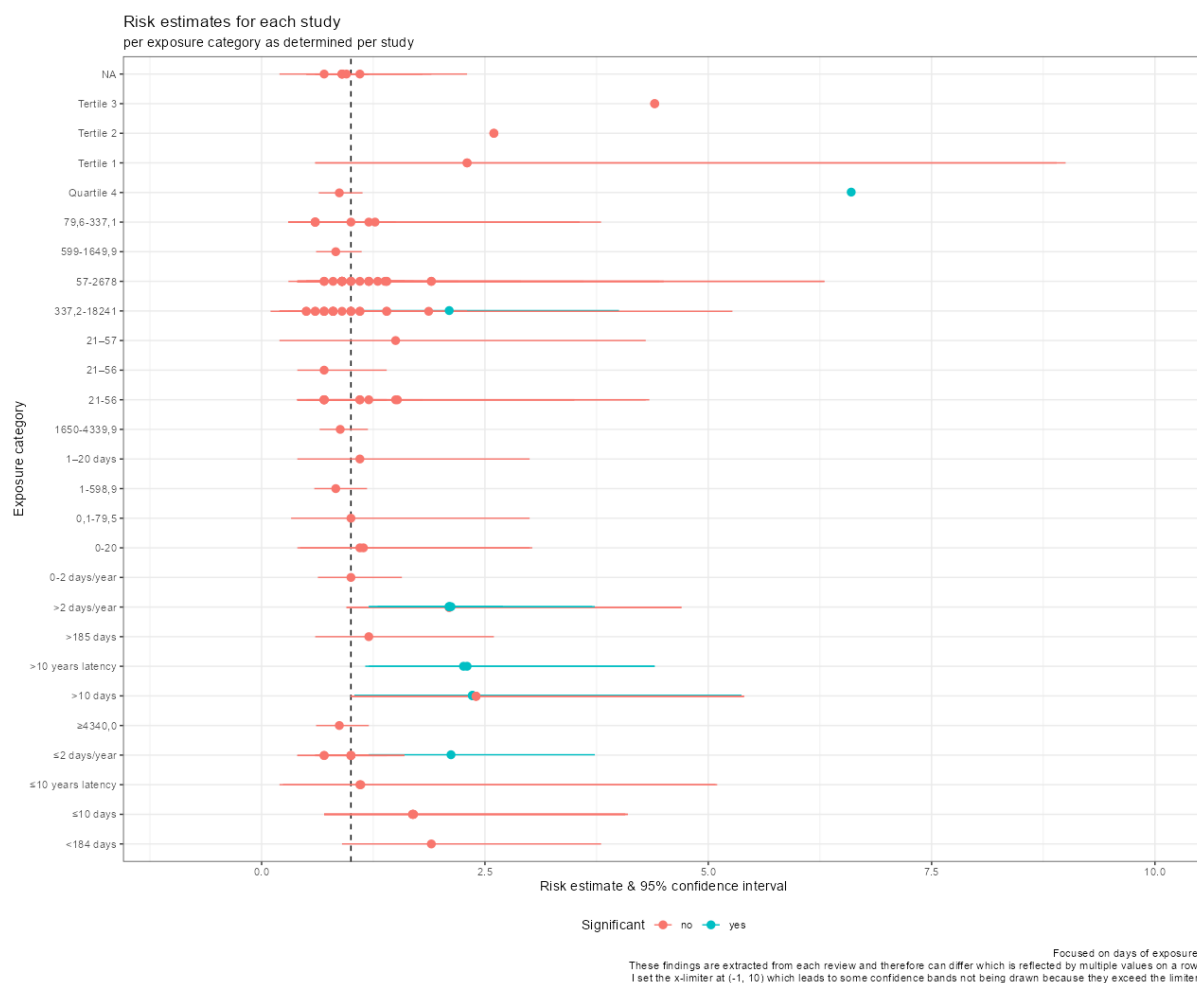
Ook nu is het wellicht beter om deze data grafisch weer te geven om zo te ontdekken waar de reviews verschillen in hun beschrijving. Het resultaat zien we in **Figuur 15** wat toont dat het gros van de reviews overeenkomt in hun beschrijving. Dat is alvast een goed uitgangspunt, maar pas in de fase van de meta-analyse kunnen we echt beoordelen in hoeverre de data voldoende overeenkomsten tonen.



Figuur 15. Schattingen voor de relatie tussen glyfosaat en non-hodgkin lymfoom zoals beschreven in elke review die de studie van Eriksson(64) heeft meegenomen.

We zien ook direct dat er best wat verschil is in wat de studies exact laten zien, maar ook wat de reviews hebben opgenomen. Zo is het niet dat elke review exact dezelfde mate van blootstelling heeft opgenomen (**Figuur 16**)⁴⁷.

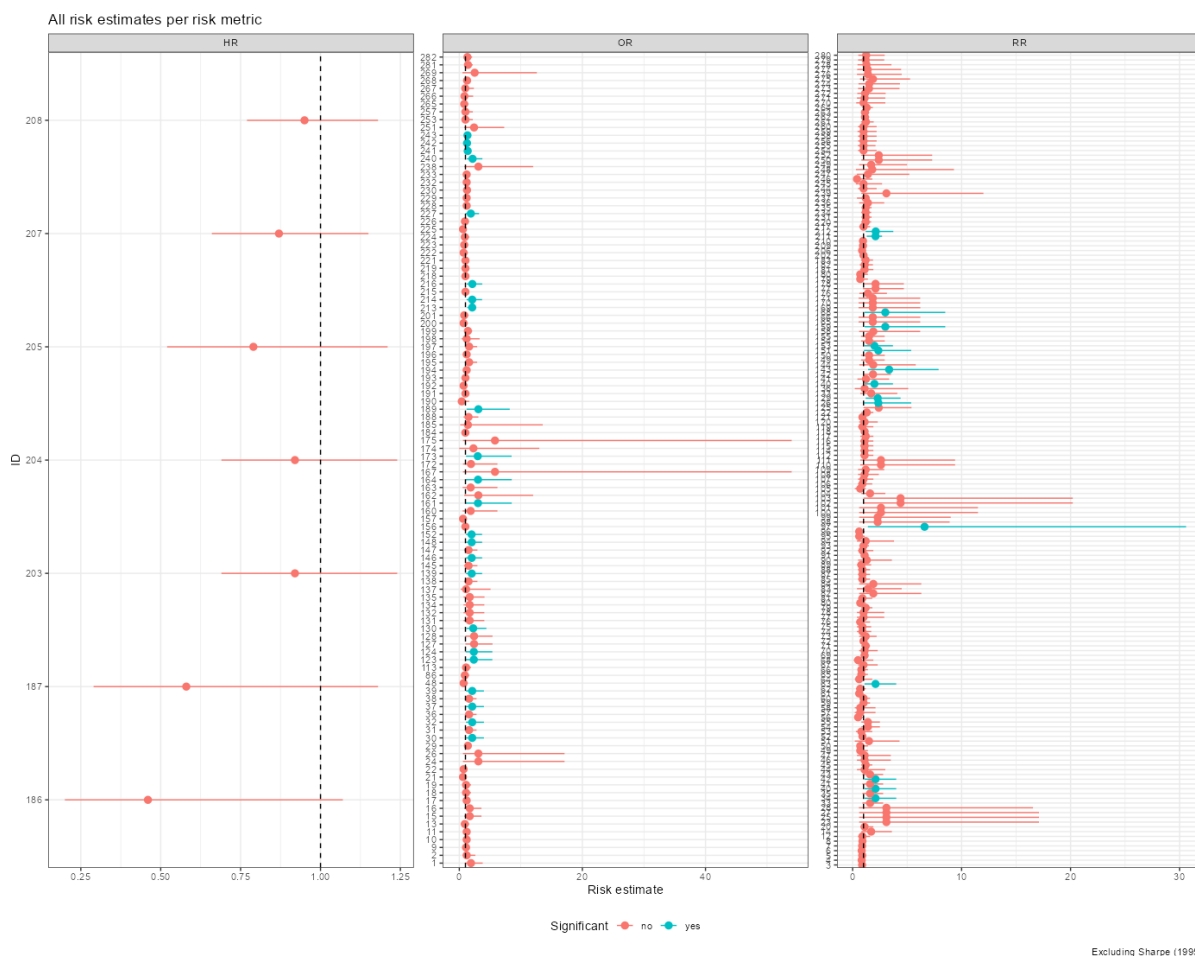
⁴⁷ Het kan natuurlijk heel goed zijn dat een review wel de ruwe data heeft opgevraagd en daar eigen analyses op heeft losgelaten, maar als ik kijk naar de overlap tussen reviews voor wat wel gepresenteerd wordt, lijkt het eerder dat er a priori keuzes zijn gemaakt in wat wel of niet wordt meegenomen. Dat hoeft natuurlijk niet gek te zijn, want een meta-analyse vraagt in principe om zoveel als mogelijk homogeniteit tussen de studies als het gaat om de uitkomst en hoe die wordt gepresenteerd. Met andere woorden: een meta-analyse is geen geautomatiseerde procedure maar het resultaat van menselijke afwegingen. Het is mijn taak hier om te zien of



Figuur 16. Risicoschattingen per studie en per hoeveelheid blootstelling. We kunnen zien dat er behoorlijk wat verschillen zijn in de mate van gemeten hoeveelheid blootstelling. Dit maakt het samenvoegen er niet makkelijker op.

Het is ook nog interessant om te zien of elke review exact dezelfde statistiek rapporteert. Omdat we *case-control* en *cohort* studies hebben ga ik er stiekem vanuit dat elke studie of een *odds-ratio* of een *risk-ratio* rapporteert. Over alle studies heen heb ik 291 schattingen en van die 291 schattingen zijn er 106 *odds-ratio's*, 171 *risk-ratio's*, 7 *hazard-ratio's* en van zeven andere schattingen heb geen informatie (**Figuur 17**). Het blijkt dat ik sowieso 4 schattingen mis en die missende waarden vind ik in drie studies: Alavanja(52) (N=1); Dennis(62) (N=1), en Ruder(84) (N=2).

ik die afwegingen (deels) kan achterhalen en hun invloed op het resultaat kan tonen. Dat doe ik door reviews én studies in reviews te vergelijken.



Figuur 17. Risicoschattingen per type statistiek en of ze significant zijn of niet. Het gaat hier om elke individuele schatting, ongeacht studie, review, of kankersoort.

Ik kan nu kijken of de reviews verschillen in de exacte statistiek die ze rapporteren. Ik ga er gemakshalve vanuit dat we echt kunnen spreken van een verschil als een review een andere statistiek rapporteert voor exact dezelfde kankersoort in exact dezelfde studie en dan exact dezelfde schattingen laat zien inclusief exact hetzelfde betrouwbaarheidsinterval. Het zal geen perfecte frequentie geven van het aantal divergenties, maar we krijgen wel een idee of er divergenties zijn en waar ze (ongeveer) optreden. Een eerste snelle analyse laat al zien dat statistieken verschillen tussen verschillende soorten *study design*:

study_design	risk_metric	n
<chr>	<chr>	<int>
1 Case-control	OR	92
2 Case-control	RR	73
3 Case-control	NA	2
4 Cohort	HR	7
5 Cohort	OR	14
6 Cohort	RR	98
7 Cohort	NA	5

Een volgende analyse laat zien hoe dat zit voor een bepaalde studie, bijvoorbeeld De Roos(60) of Eriksson(64):

```
# Groups: study_author_year, exposure_metric, exposure_category [14]
  study_author_year exposure_metric exposure_category risk_metric risk_estimate risk_estimate_low risk_estimate_high
  <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>
1 De Roos,2005 Days of exposure 79,6-337,1 RR 0.6 0.3 1.1
2 De Roos,2005 Days of exposure 79,6-337,1 NA 0.6 0.3 1.1
3 De Roos,2005 Days of exposure 21-56 OR 0.7 0.4 1.4
4 De Roos,2005 Days of exposure 21-56 RR 0.7 0.4 1.4
5 De Roos,2005 Days of exposure 21-56 NA 0.7 0.4 1.4
6 De Roos,2005 Days of exposure 337,2-18241 RR 0.8 0.5 1.4
7 De Roos,2005 Days of exposure 57-2678 RR 0.9 0.5 1.6
8 De Roos,2005 Days of exposure 57-2678 OR 0.9 0.5 1.6
9 Eriksson,2008 Days of exposure =10 years latency RR 1.1 0.2 5.1
10 De Roos,2005 Exposed NA RR 1.1 0.7 1.9
11 De Roos,2005 Exposed NA OR 1.1 0.7 1.9
12 Eriksson,2008 Days of exposure =10 years latency OR 1.11 0.24 5.08
13 De Roos,2005 Exposed NA RR 1.2 0.7 1.9
14 Eriksson,2008 Exposed =1 day (multivariate) OR 1.51 0.77 2.94
15 Eriksson,2008 Exposed NA OR 1.51 0.77 2.94
16 Eriksson,2008 Exposed NA RR 1.51 0.77 2.94
17 Eriksson,2008 Exposed NA RR 1.51 0.77 2.95
18 Eriksson,2008 Days of exposure =10 days OR 1.69 0.7 4.07
19 Eriksson,2008 Days of exposure =10 days RR 1.7 0.7 4.1
20 Eriksson,2008 Days of exposure =10 days OR 1.7 0.7 4.1
21 Eriksson,2008 Exposed =1 day/lifetime RR 2 1.1 3.7
22 Eriksson,2008 Exposed NA OR 2 1.1 3.7
23 Eriksson,2008 Exposed NA RR 2 1.1 3.7
24 Eriksson,2008 Exposed =1 day (univariate) OR 2.02 1.1 3.71
25 Eriksson,2008 Exposed NA OR 2.02 1.1 3.71
26 Eriksson,2008 Days of exposure >10 years latency OR 2.26 1.16 4.4
27 Eriksson,2008 Days of exposure >10 years latency RR 2.3 1.2 4.4
28 Eriksson,2008 Days of exposure >10 days OR 2.36 1.04 5.37
29 Eriksson,2008 Days of exposure >10 days RR 2.36 1.04 5.37
30 Eriksson,2008 Exposed NA RR 2.36 1.04 5.37
31 Eriksson,2008 Days of exposure >10 days RR 2.4 1 5.4
32 Eriksson,2008 Days of exposure >10 days OR 2.4 1 5.4
```

We kunnen duidelijk zien dat dezelfde schattingen anders opgenomen zijn in een review. Deze verschillen kunnen we niet zomaar over het hoofd zien want een *odds-ratio* is niet gelijk aan een *risk-ratio*. Alleen wanneer we spreken over een zeldzame ziekte kunnen we enigszins de schattingen van beide statistieken (dat wil zeggen vanuit een *odds-ratio* en een *risk-ratio*) combineren. Maar kanker is niet zo zeldzaam als bijvoorbeeld de ziekte van Parkinson. Het zomaar samenvoegen van beide statistieken introduceert daarmee haar eigen problemen is dus geen wenselijke optie. We zullen dus een keuze moeten maken over welke schattingen we waar meenemen. Als we bovenstaande tot een cijfer samenvatten dan zien we dat van de 32 'unieke rijen' er nog maar 22 overblijven en dat van de eerste schatting we drie statistieken hebben⁴⁸:

```
# Groups: study_author_year, exposure_metric, exposure_category, risk_estimate, risk_estimate_low, risk_estimate_high [21]
  study_author_year exposure_metric exposure_category risk_estimate risk_estimate_low risk_estimate_high n_risk_metrics
  <chr> <chr> <chr> <dbl> <dbl> <dbl> <int>
1 De Roos,2005 Days of exposure 21-56 0.7 0.4 1.4 3
2 De Roos,2005 Days of exposure 337,2-18241 0.8 0.5 1.4 1
3 De Roos,2005 Days of exposure 57-2678 0.9 0.5 1.6 2
4 De Roos,2005 Days of exposure 79,6-337,1 0.6 0.3 1.1 2
5 De Roos,2005 Exposed NA 1.1 0.7 1.9 2
6 De Roos,2005 Exposed NA 1.2 0.7 1.9 1
7 Eriksson,2008 Days of exposure >10 days 2.36 1.04 5.37 2
8 Eriksson,2008 Days of exposure >10 days 2.4 1 5.4 1
9 Eriksson,2008 Days of exposure >10 years latency 2.26 1.16 4.4 1
10 Eriksson,2008 Days of exposure >10 years latency 2.3 1.2 4.4 1
11 Eriksson,2008 Days of exposure =10 days 1.69 0.7 4.07 1
12 Eriksson,2008 Days of exposure =10 days 1.7 0.7 4.1 2
13 Eriksson,2008 Days of exposure =10 years latency 1.1 0.2 5.1 1
14 Eriksson,2008 Days of exposure =10 years latency 1.11 0.24 5.08 1
15 Eriksson,2008 Exposed =1 day (multivariate) 1.51 0.77 2.94 1
16 Eriksson,2008 Exposed =1 day (univariate) 2.02 1.1 3.71 1
17 Eriksson,2008 Exposed =1 day/lifetime 2 1.1 3.7 1
18 Eriksson,2008 Exposed NA 1.51 0.77 2.94 2
19 Eriksson,2008 Exposed NA 1.51 0.77 2.95 1
20 Eriksson,2008 Exposed NA 2 1.1 3.7 2
21 Eriksson,2008 Exposed NA 2.02 1.1 3.71 1
22 Eriksson,2008 Exposed NA 2.36 1.04 5.37 1
```

Het straks zomaar op een hoop gooien van gegevens is er dus niet bij. Ook is er geen sprake van 291 rijen aan unieke gegevens (Figuur 12). We zullen later heel goed moeten opletten

⁴⁸ Eigenlijk hebben we er maar twee (*odds-ratio* en *risk-ratio*) en is de derde NA.

om te zien wat het exact is dat we combineren én wat het gevolg is van gemaakte keuzes. Voor nu weten we dus al dat het allemaal niet zo eenvoudig is als soms wordt gecommuniceerd.

Risicoschattingen

Laten we nu eens kijken naar de risicoschattingen (de *estimates*) zoals deze in elke review zijn vermeld en door mij zijn verzameld (voor zover dat kon). Voordat ik begin met het tonen van de individuele resultaten is het, nogmaals, essentieel om te onthouden dat een ander *study design* vaak ook aanleiding geeft tot een andere statistiek (de test statistiek of *risk metric*⁴⁹)(93). Deze testen leiden weer tot een verandering in de interpretatie⁵⁰. We hebben ook al benoemd dat het vergelijken van bevindingen niet zomaar kan. Het moet minstens gaan over dezelfde ziekte en er moet eenzelfde soort blootstelling zijn geweest. Laten we eens kijken wat voor waarden we krijgen als we alle rijen aan gegevens meenemen die iets zeggen over NHL⁵¹:

	risk_metric	n	mean	sd	var	min	q10	q90	max
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	HR	1	0.95	NA	NA	0.95	0.95	0.95	0.95
2	OR	60	1.75	0.969	0.938	0.6	0.918	2.4	5.8
3	RR	72	1.45	0.650	0.423	0.6	0.834	2.35	3.1

Ik zie geen extreme waarden wat mij geruststelt in de extractie van de data. Nu ben ik benieuwd hoe de verdeling van de risicoschattingen eruit ziet als ik over alle kankersoorten heen de data samenvat:

	risk_metric	n	mean	sd	var	min	q10	q90	max
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	HR	7	0.784	0.191	0.0364	0.46	0.532	0.932	0.95
2	OR	106	2.84	12.4	153.	0.4	0.8	3.02	129.
3	RR	171	1.45	0.809	0.655	0.4	0.8	2.4	6.6

Nu ziet het er anders uit. Zo zie ik een extreme *odds-ratio* van 129 en deze blijkt afkomstig te zijn van de studie van Sharpe(85):

	study_author_year	risk_estimate
	<chr>	<dbl>
1	Sharpe, 1995	129.
2	De Roos, 2005	6.6

⁴⁹ Er bestaat geen alomvattende definitie en *test statistic* of *test metric* kun je in mijn optiek gerust door elkaar gebruiken, hoewel de purist hier wellicht aanstalten maakt tot bezwaar.

⁵⁰ Het grootste probleem met een meta-analyse is het verschuiven van het oorspronkelijk doel om toch tot een resultaat te komen. Dus wanneer het vergelijken van appels uiteindelijk resulteert in het vergelijken van appels met peren. In dat geval kun je eigenlijk alleen nog maar iets zeggen over fruit, terwijl je eigenlijk iets wilde zeggen over appels.

⁵¹ Ik verwijder in dit geval geen dubbelten.

3	Hardell,2002	5.8
4	Hardell & Eriksson,1999	5.8
5	De Roos,2005	4.4
6	De Roos,2005	4.4
7	Eriksson,2008	3.35
8	Sharpe,1995	3.24
9	Cocco,2013	3.1
10	Cocco,2013	3.1

Het is bij verre de meest extreme waarde in deze reeks, maar het is geen schrijffout. De waarde wordt daadwerkelijk opgenomen in de review van Khan(42) en heeft betrekking op nefroblastoom (*Wilm's tumor*): een zeldzame vorm van nierkanker die het meest voorkomt bij jonge kinderen(98). Hoewel deze waarde dus niet incorrect is, is hij dermate groot dat het elke volgende grafiek, samenvatting of analyse zal vertekenen. We kunnen er nu voor kiezen om de data op log schaal te presenteren, of om deze waarde uit toekomstige analyses te verwijderen. En hoewel ik liever geen data wil weglaten om het mijzelf makkelijker te maken is het in dit geval beter om dat wel te doen. Zo heb ik die data al uit **Figuur 12** en **Figuur 17** gehaald om het overzichtelijk te houden.

Laten we nu eens kijken naar de onzekerheden die gerapporteerd worden voor elke meting (in **Figuur 17** wederom zichtbaar). Dit zijn altijd 95% betrouwbaarheidsintervallen. Eerst zal ik de statistische samenvatting laten zien van de linkerkant van een 95% betrouwbaarheidsinterval. Vervolgens toon ik de rechterkant. In beide gevallen laat ik de studie van Sharpe(85) weg wat betekent dat ik twee rijen aan data niet meeneem:

review_author_year	disease	risk_estimate	risk_estimate_low	risk_estimate_high
<chr>	<chr>	<dbl>	<dbl>	<dbl>
1 A.Khan,2022	wilm's tumor	129.	6.4	2569
2 A.Khan,2022	wilm's tumor	3.24	0.2	9

De volgende tabellen laten de samenvattingen, per statistiek⁵², zien voor de onder -en bovenkant van een 95% betrouwbaarheidsinterval (zie ook **Figuur 12**):

risk_metric	n	mean	sd	var	min	q10	q90	max
1 HR	7	0.546	0.220	0.0484	0.2	0.254	0.722	0.77
2 OR	104	0.760	0.275	0.0755	0	0.4	1.1	1.3
3 RR	171	0.638	0.256	0.0653	0.1	0.33	1	1.42

risk_metric	n	mean	sd	var	min	q10	q90	max
1 HR	7	1.18	0.059	0.0035	1.07	1.12	1.24	1.24
2 OR	104	4.67	7.69	59.1	1.1	1.47	8.51	54
3 RR	171	3.87	4.11	16.9	1	1.2	7.89	30.6

⁵² Of metriek, wat je zelf wil hanteren, hoewel ik mij kan voorstellen dat voor de purist dit niet opgaat (metriek heeft meer te maken met wat je meet, statistiek met hoe je rekent).

Met het weghalen van Sharpe(85) lijken we alle extremen te hebben verwijderd. Nu is een *odds-ratio* of *risk-ratio* van 54 en 30.6, respectievelijk, nog steeds fors te noemen, maar het is niet onmogelijk⁵³.

Omdat we weten dat er verschillende statistieken zijn gebruikt bij verschillende *study designs* is het interessant om te zien of we daar noemenswaardige verschillen zien. Uiteraard zal de specifieke ziekte en de mate van blootstelling samenhangen met de gerapporteerde uitkomst, maar voor nu gaat ben ik vooral geïnteresseerd in eigenaardigheden. In principe ben ik dus in alles geïnteresseerd wat een eventuele meta-analyse makkelijker of moeilijker maakt om uit te voeren, zodat ik bij het interpreteren van de resultaten goed kan duiden waar we het exact wel of niet over kunnen hebben. Of, in sommige gevallen, waarom we beter geen meta-analyse kunnen doen. Ik ben dus op zoek naar ruis. De risicoschattingen per statistiek en *study design* ziet er als volgt uit:

	study_design	risk_metric	n	mean	sd	var	min	q10	q90	max
1	Cohort	HR	7	0.784	0.191	0.0364	0.46	0.532	0.932	0.95
2	Case-control	OR	90	1.71	0.921	0.848	0.4	0.836	3.00	5.8
3	Cohort	OR	14	1.13	0.370	0.137	0.61	0.7	1.6	1.9
4	Case-control	RR	73	1.70	0.681	0.464	0.4	1	2.88	3.35
5	Cohort	RR	98	1.26	0.849	0.721	0.5	0.7	1.96	6.6
1	Cohort	HR	7	0.546	0.220	0.0484	0.2	0.254	0.722	0.77
2	Case-control	OR	90	0.772	0.283	0.0803	0	0.4	1.1	1.3
3	Cohort	OR	14	0.68	0.202	0.0407	0.32	0.43	0.9	0.9
4	Case-control	RR	73	0.733	0.268	0.0721	0.1	0.408	1.1	1.42
5	Cohort	RR	98	0.567	0.222	0.0491	0.1	0.3	0.8	1.4
1	Cohort	HR	7	1.18	0.059	0.0035	1.07	1.12	1.24	1.24
2	Case-control	OR	90	5.09	8.18	66.9	1.1	1.50	8.87	54
3	Cohort	OR	14	1.95	0.791	0.626	1.1	1.23	2.9	3.8
4	Case-control	RR	73	4.56	3.71	13.8	1.4	1.7	8.38	17.1
5	Cohort	RR	98	3.36	4.33	18.7	1	1.17	6.3	30.6

Ik zie zelf geen noemenswaardige verschillen, maar wellicht dat het zich anders presenteert als we beginnen met het grafisch afbeelden van de data die we hebben. Daar zullen we straks nog uitgebreid bij stil staan (en wie niet kan wachten kan alvast door naar **Figuur 19**, **Figuur 20** en **Figuur 21**). Maar voordat we verder gaan wil ik bovenstaande exercitie lichtjes herhalen door te kijken naar de schattingen per kankersoort en de mate van blootstelling (zie ook **Figuur 9**). Wanneer ik alleen per kankersoort kijk dan zie ik de volgende data voor de risicoschattingen:

disease	n	mean	sd	var	min	q10	q90	max
1 NHL	135	1.57	0.822	0.675	0.6	0.846	2.36	5.8
2 MM	50	1.76	1.06	1.12	0.6	0.987	2.6	6.6
3 Wilms' tumor	10	1.37	0.560	0.314	0.8	0.836	2.19	2.5

⁵³ Waarschijnlijk is het een combinatie van een kleine studiegrootte en een miniem verschil in aantallen die wel of niet kanker hebben gekregen in de risicogroep en de controlegroep.

4	Leukemia	9	0.978	0.222	0.0494	0.7	0.7	1.24	1.4
5	HCL	8	1.85	1.13	1.29	0.46	0.544	3.1	3.1
6	Rectal cancer	8	1.2	0.273	0.0743	0.9	0.9	1.6	1.6
7	Colon cancer	7	1.04	0.264	0.0695	0.7	0.82	1.4	1.4
8	LHC	7	1.1	0.1	0.0100	1	1	1.2	1.2
9	B-cell lymphoma	6	1.87	1.01	1.02	0.92	0.96	3.1	3.1
10	Brain cancer	5	1.26	1.11	1.23	0.4	0.48	2.46	3.1
11	Pancreatic cancer	5	1.2	0.5	0.25	0.5	0.74	1.66	1.9
12	CLL/SLL	3	1.56	1.57	2.48	0.4	0.504	2.86	3.35
13	Cutaneous melanoma	3	1.07	0.473	0.223	0.7	0.74	1.46	1.6
14	Follicular lymphoma	3	1.36	0.551	0.304	0.79	0.912	1.79	1.89
15	NHL+HCL	3	2.25	0.65	0.422	1.85	1.86	2.78	3
16	Prostate cancer	3	1.1	0	0	1.1	1.1	1.1	1.1
17	Bladder cancer	2	1	0.283	0.0800	0.8	0.84	1.16	1.2
18	Breast cancer	2	1.1	0.283	0.08	0.9	0.94	1.26	1.3
19	Kidney cancer	2	0.6	0.141	0.0200	0.5	0.52	0.68	0.7
20	Lung cancer	2	0.65	0.0707	0.0050	0.6	0.61	0.69	0.7
21	Melanoma	2	0.8	0.141	0.0200	0.7	0.72	0.88	0.9
22	Oral cancer	2	0.9	0.141	0.0200	0.8	0.82	0.98	1
23	All cancers	1	1	NA	NA	1	1	1	1
24	All cancers adults	1	0.9	NA	NA	0.9	0.9	0.9	0.9
25	Esophagus cancer	1	0.7	NA	NA	0.7	0.7	0.7	0.7
26	Stomach cancer	1	0.8	NA	NA	0.8	0.8	0.8	0.8
27	Total Childhood Cancer	1	0.61	NA	NA	0.61	0.61	0.61	0.61
28	Tyroid cancer (distant/regional)	1	1.37	NA	NA	1.37	1.37	1.37	1.37
29	Tyroid cancer (localized)	1	1.26	NA	NA	1.26	1.26	1.26	1.26
30	Tyroid cancer (primary)	1	1.33	NA	NA	1.33	1.33	1.33	1.33

Wat direct opvalt is dat we van maar 11 kankersoorten vijf of meer schattingen hebben, waarbij NHL de kroon spant met 135 schattingen⁵⁴. Dit is al een interessante bevinding, omdat het ons zal beperking in welke meta-analyses we wel en niet kunnen uitvoeren. Ook blijkt dat eigenlijk maar voor drie kankersoorten de data echt interessant genoeg is om bruikbaar te zijn voor verdere analyses (zie ook **Figuur 43**). Om de invloed van *study design*, de mate van blootstelling en de gebruikte statistiek zo goed als mogelijk te kunnen duiden zal ik mij nu voornamelijk richten op NHL. Daar zijn gewoonweg de meeste schattingen van. Hieronder zien we de resultaten bij een steeds verdere uitsplitsing voor NHL:

study_design	risk_metric	n	mean	sd	var	min	q10	q90	max	
<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Case-control	OR	57	1.80	0.973	0.946	0.6	1	2.4	5.8
2	Case-control	RR	49	1.71	0.629	0.396	1	1	2.52	3.1
3	Cohort	HR	1	0.95	NA	NA	0.95	0.95	0.95	0.95
4	Cohort	OR	3	0.9	0.2	0.0400	0.7	0.74	1.06	1.1
5	Cohort	RR	23	0.895	0.159	0.0253	0.6	0.7	1.1	1.2

exposure_metric	study_design	risk_metric	n	mean	sd	var	min	q10	q90	max	
<chr>	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Days of exposure	Case-control	OR	21	1.57	0.643	0.414	0.6	0.8	2.36	2.4
2	Days of exposure	Case-control	RR	8	1.88	0.560	0.314	1	1.07	2.37	2.4
3	Days of exposure	Cohort	HR	1	0.95	NA	NA	0.95	0.95	0.95	0.95
4	Days of exposure	Cohort	OR	2	0.8	0.141	0.0200	0.7	0.72	0.88	0.9
5	Days of exposure	Cohort	RR	15	0.805	0.106	0.0113	0.6	0.64	0.9	0.9
6	Exposed	Case-control	OR	36	1.93	1.11	1.23	0.92	1.1	3.04	5.8
7	Exposed	Case-control	RR	41	1.68	0.643	0.413	1	1	3	3.1
8	Exposed	Cohort	OR	1	1.1	NA	NA	1.1	1.1	1.1	1.1
9	Exposed	Cohort	RR	8	1.06	0.0876	0.00768	0.95	0.95	1.13	1.2

exposure_metric	exposure_category	study_design	risk_metric	n	mean	sd	var	min	q10	q90	max	
<chr>	<chr>	<chr>	<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
1	Exposed	=1 day/lifetime	Case-control	RR	12	1.78	0.549	0.302	1	1.22	2.1	3.1
2	Exposed	=1 day/lifetime	Case-control	OR	7	1.89	0.267	0.0714	1.6	1.6	2.1	2.1
3	Days of exposure	>10 days	Case-control	OR	4	2.38	0.0231	0.000533	2.36	2.36	2.4	2.4
4	Days of exposure	=10 days	Case-control	OR	4	1.69	0.00500	0.000250	1.69	1.69	1.70	1.7
5	Days of exposure	=2 days/year	Case-control	OR	4	1.28	0.56	0.314	1	1	1.78	2.12
6	Days of exposure	57-2678	Cohort	RR	3	0.9	0	0	0.9	0.9	0.9	0.9

⁵⁴ Waarvan we weten dat ze niet allemaal uniek zijn. Het betreft hier schattingen en niet studies, maar daarover later meer.

7 Days of exposure 21-56	Cohort	RR	2	0.7	0	0	0.7	0.7	0.7	0.7
8 Days of exposure 337,2-18241	Cohort	RR	2	0.8	0	0	0.8	0.8	0.8	0.8
9 Days of exposure 79,6-337,1	Cohort	RR	2	0.6	0	0	0.6	0.6	0.6	0.6
10 Days of exposure >10 days	Case-control	RR	2	2.38	0.0283	0.000800	2.36	2.36	2.40	2.4

Bij de laatste tabel zien we wat er gebeurt als we te ver uitsplitsen: we houden te weinig data over om zinvolle samenvattingen te maken. Dit betekent dat we niet alle meta-analyses kunnen doen die we graag zouden willen doen, maar deze ontdekking is op zichzelf al een bijzonder interessante bevinding. Het maakt namelijk dat de bewijslast voor de relatie tussen glyfosaat en kanker anders is per kankersoort. Daarom kunnen we eigenlijk niet eens van 'kanker' als overkoepelende term praten, maar zullen we specifiek moeten zijn en exact duiden waar we het over hebben.

Correctie voor confounders

Het is nu tijd om te kijken voor welke variabelen er gecorrigeerd is, als er al gecorrigeerd is. Omdat elke review verschilt in de beschrijving van dezelfde studie zal ik om te beginnen de data per regel beschrijven (met 291 regels in totaal). Laten we eerst eens kijken hoeveel regels aan gegevens we hebben over het toepassen van correctiemethoden:

No	Yes
104	187

Dit betekent dat een deel van de regels (104/291 = 35.7%) geen informatie beschikt over mogelijke correctiefactoren⁵⁵. Voor die regels waarin we wel informatie terugvinden zien we volgende top 20⁵⁶:

```
# A tibble: 97 x 2
  confounder          n
  <chr>             <int>
1 age                125
2 education           72
3 alcohol consumption in the past year 44
4 family history of cancer 44
5 family history of cancer in first-degree relatives 44
6 pack-years of cigarette smoking 44
7 smoking            34
8 age at enrollment  30
9 medical history    23
10 and state of residence 22
11 and state of residence. also adjusted for other pesticides 22
12 province of residence 22
13 sex                21
14 state              19
15 site               18
16 vital status       18
17 alcohol            17
```

⁵⁵ Maar dat wil niet zeggen dat we helemaal geen informatie hebben per studie. Elke regel is een schatting van een studie zoals opgenomen door een review en reviews verschillen in wat ze publiceren.

⁵⁶ Het is belangrijk om te blijven onthouden dat ik de bewoording heb gesimplificeerd om enigszins te kunnen tellen met het aantal 'categorieën'. Toch heb ik sommige categorieën niet samengevoegd, zoals 'sex' en 'gender' of 'race' en 'ethnicity'. Dit omdat ik ze verschillend vond. Een andere onderzoeker had het hoogstwaarschijnlijk anders gedaan en kan zelf aan de slag met de meegeleverde data en codes.

18	socioeconomic category	17
19	10 pesticides	13
20	county	13

Er lijkt in totaal voor 97 verschillende factoren gecorrigeerd te zijn waarvan het gros heeft gekeken naar 'leeftijd', 'opleiding', 'alcohol consumptie', 'familiegeschiedenis voor kanker' en 'roken'. Een snelle blik verraad dat ik sommige categorieën bij elkaar had kunnen zetten (zoals *state of residence* en *province of residence*, en *pack-years of cigarette smoking* en *smoking*), maar dan had ik de gegevens (in sommige gevallen) meer homogeen gemaakt dan ze daadwerkelijk zijn⁵⁷. Verder zijn deze samenvoegingen niet zonder risico's. Een variabele toevoegen aan de analyse, welke variabele dan ook, maakt dat de schatting verandert (en ook de betrouwbaarheid van die schatting). Je splitst de data namelijk op. Door de data te presenteren zoals ze zijn overgenomen uit de reviews zien we beter dat we te maken hebben met een diverse manier van corrigeren⁵⁸. Dit gegeven kan ik alleen al tonen door de verschillende correctiemethoden voor de studie van De Roos(60) op te sommen:

confounder	n
<chr>	<int>
1 education	43
2 age at enrollment	30
3 alcohol consumption in the past year	30
4 family history of cancer in first-degree relatives	30
5 pack-years of cigarette smoking	30
6 and state of residence	22
7 age	19
8 alcohol	13
9 smoking	13
10 state	12
11 10 pesticides	9
12 and state of residence. also adjusted for other pesticides	8
13 family history of cancer	7
14 family history	5
15 10 other pesticides most strongly correlated with glyphosate	4
16 family cancer history	4
17 sex	4
18 alcohol consumption	3
19 education level	3
20 smoking history	3
21 year of enrolment	3
22 same as above	2
23 and other pesticides	1
24 demographic factors	1
25 fully adjusted	1
26 lifestyle	1

⁵⁷ Ik heb dat toentertijd voor [mijn rapport](#) over risicofactoren voor Parkinson wel gedaan, maar ik had daar ook veel meer data, waardoor ik me meer comfortabel voelde in het samenvoegen.

⁵⁸ Of in het beschrijven van correctiefactoren, wat ook weer laat zien hoe lastig het is om met gepubliceerde informatie aan de slag te gaan.

De oplettende lezer zal zien dat 'same as above' en 'fully adjusted'⁵⁹ ook zijn opgenomen als covariaten. Dit is natuurlijk helemaal niet informatief. Ook nu had ik er voor kunnen kiezen om verschillende soorten variabelen met elkaar te combineren (bijvoorbeeld *family history of cancer* en *family history, of alcohol* en *alcohol consumption*). Toch blijf ik van mening dat het beter is om een beeld krijgen van wat er allemaal verschilt tussen de reviews, dan om homogeniteit af te dwingen. Dit alles helpt mij in de latere verwerking van de gegevens. Het wordt nu sowieso tijd om de risicoschattingen verder uit te werken en te verenigen met de rest van de gegevens.

Welke factoren hebben invloed op een risicoschatting?

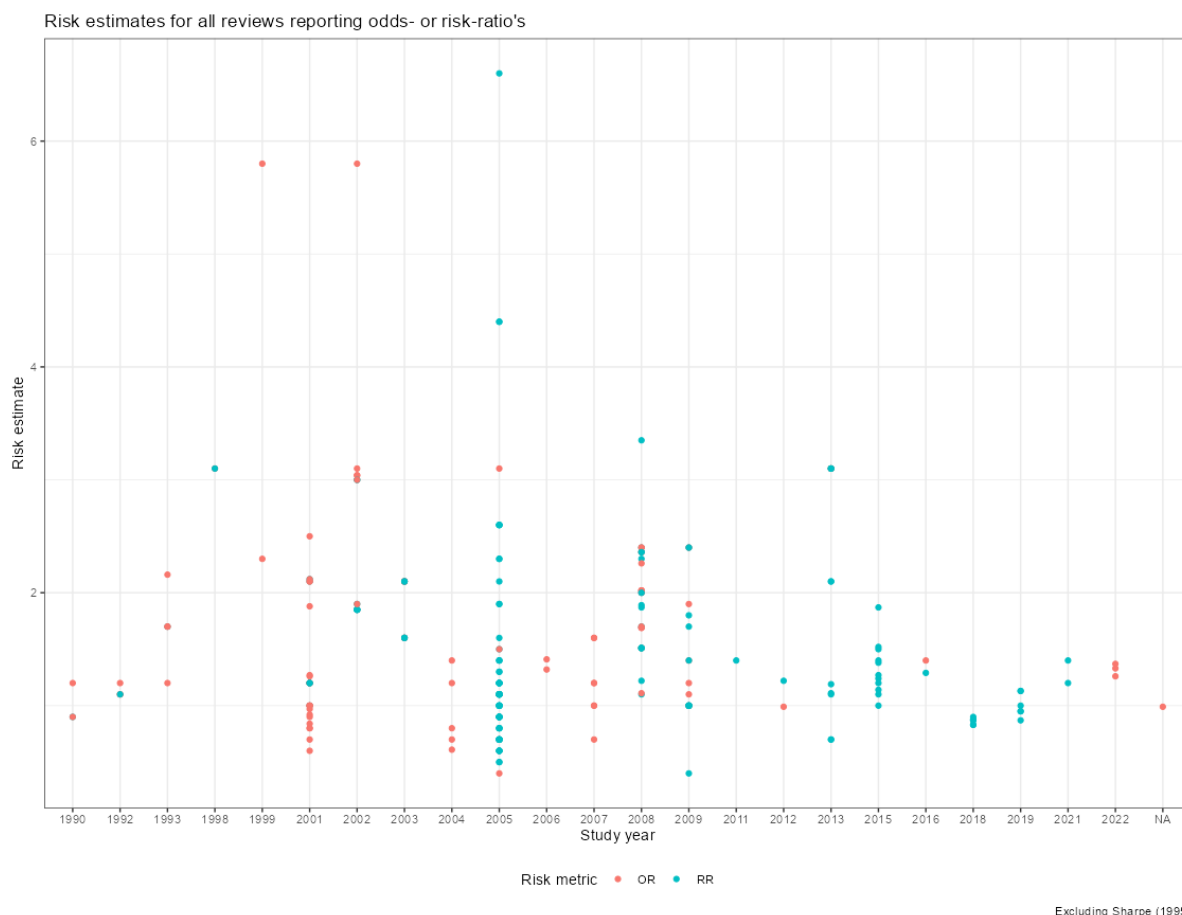
Zoals eerder vermeld beperk ik mij in deze analyses tot de *odds-ratio* en de *risk-ratio*. Ik zal de grafieken langzaam opbouwen door steeds meer informatie toe te voegen. De belangrijkste vraag bij elke toevoeging is of het toevoegen van informatie ook daadwerkelijk iets toevoegt aan het signaal of alleen maar ruis toevoegt⁶⁰. Ook heb ik al benoemt dat het identificeren van ruis juist een uitgangspunt van dit rapport is.

In **Figuur 18** zien we alle risicoschattingen over de jaren heen. Duidelijk is te zien dat de meeste schattingen onder de vier liggen met sommige uitschieters. Ook lijkt het erop alsof er door de jaren heen, vanaf het jaar 2000, ongeveer eenzelfde hoeveelheid aan studies is uitgevoerd⁶¹. In principe doe ik nu niets anders dan grafisch weergeven wat ik heb verzameld om een beeld te krijgen van het landschap.

⁵⁹ Een getal wat is gecorrigeerd noemt men ook wel vaker 'adjusted'. Een getal wat niet is gecorrigeerd heet 'raw'.

⁶⁰ Ik zal bij alle volgende analyses de studie van Sharpe, 1995 weglaten omdat deze studie een grote invloed heeft op de grafieken. Het gaat uiteindelijk om het weglaten van twee risicoschattingen voor nefroblastoom.

⁶¹ De punten die elk een gerapporteerde schatting moeten voorstellen liggen zo dicht op elkaar dat ze over elkaar heen worden getekend. De afbeelding geeft dus niet direct aan hoeveel metingen er zijn per jaar.



Figuur 18. De relatie tussen de gevonden risicoschattingen en het studie jaar waarin ze zijn gepubliceerd per soort meting.

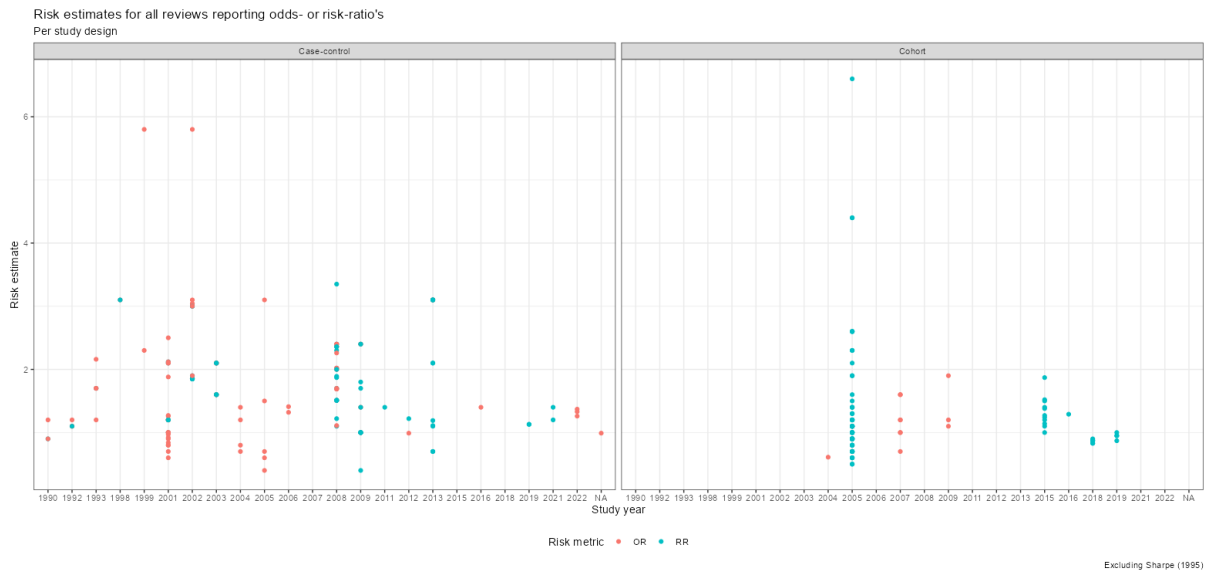
In **Figuur 19**, **Figuur 20**, **Figuur 21** en **Figuur 23** zien we het aantal schattingen per *study design*. Wat opvalt is dat de *case-control* studies haast net zo veel *odds-ratio's* als *risk-ratio's* rapporteren terwijl dit in de cohort studies vooral *risk-ratio's* zijn. In **Figuur 22** zien we de schattingen per studie en statistiek. Als we het publicatiejaar en *study design* weglaten, en gewoonweg kijken naar de verdeling van waarden, dan krijgen we **Figuur 24**. De stippellijn is de lijn van 'geen verschil'. Dit is de lijn waartegen statistische significantie wordt afgezet⁶². Ook deze figuur kunnen we opsplitsen in het type *study design* waardoor we **Figuur 25**⁶³ en **Figuur 26** krijgen. We zien direct dat het opbreken van de data invloed heeft op de curve die daarmee, zeker voor cohort studies, onwerkbaar wordt.

De stippellijn in **Figuur 24** en **Figuur 26** is eigenlijk niet genoeg om een splitsing te maken tussen bevindingen die wel significant zijn en bevindingen die niet significant zijn.

⁶² Statistische significantie heeft niks te maken met klinische relevantie. Een *odds-ratio* van 2 kan statistisch significant zijn, maar niet noodzakelijkerwijs klinisch relevant. Om een voorbeeld te noemen: de *odds-ratio's* voor de relatie tussen roken en longkanker liggen tussen de 50 en de 100, afhankelijk van type kanker en type persoon.

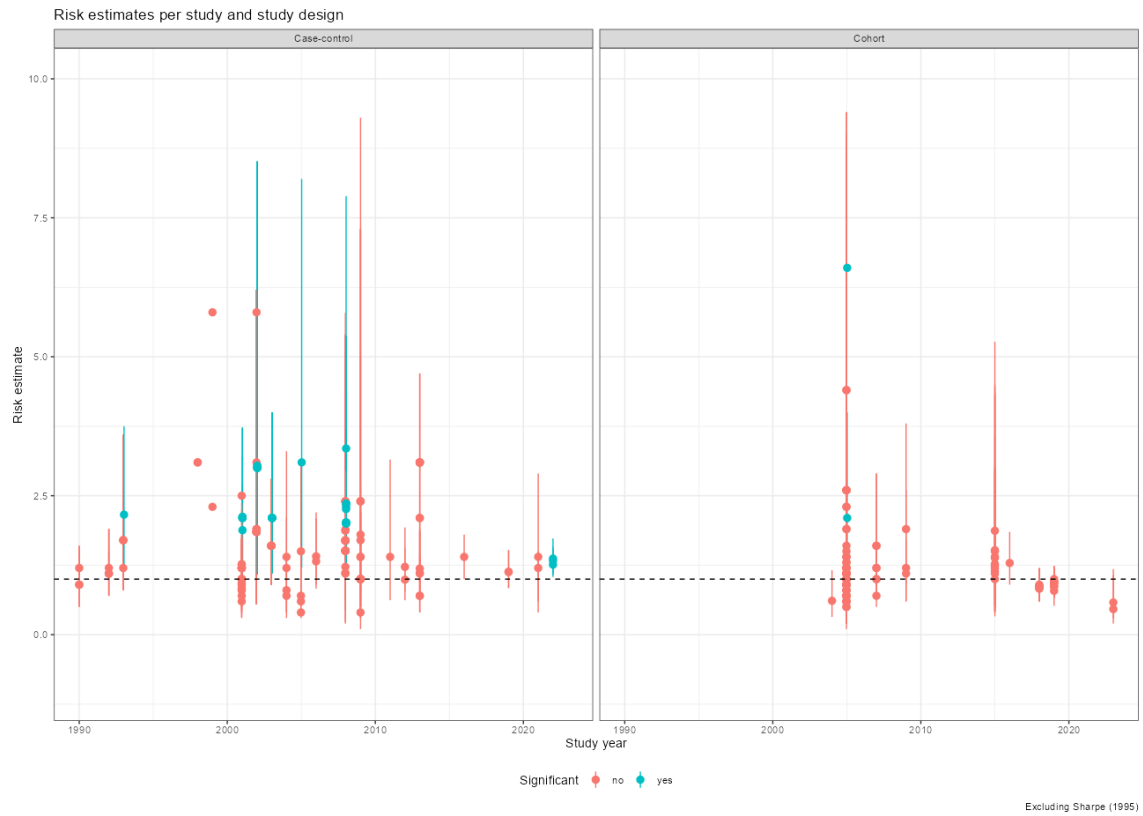
⁶³ Ik heb hier expres voor een *boxplot* gekozen omdat het oog niet wordt afgeleid door een soms onrealistische piek. Duidelijk wordt nu de lange staart bij de schattingen in cohort studies en de hogere risicoschatting bij *case-control* studies.

Daarvoor zullen we de onzekerheden van elke schatting moeten meewegen. Dit zijn de 95%-betrouwbaarheidsintervallen die we voor elke schatting hebben overgenomen vanuit de reviews⁶⁴.

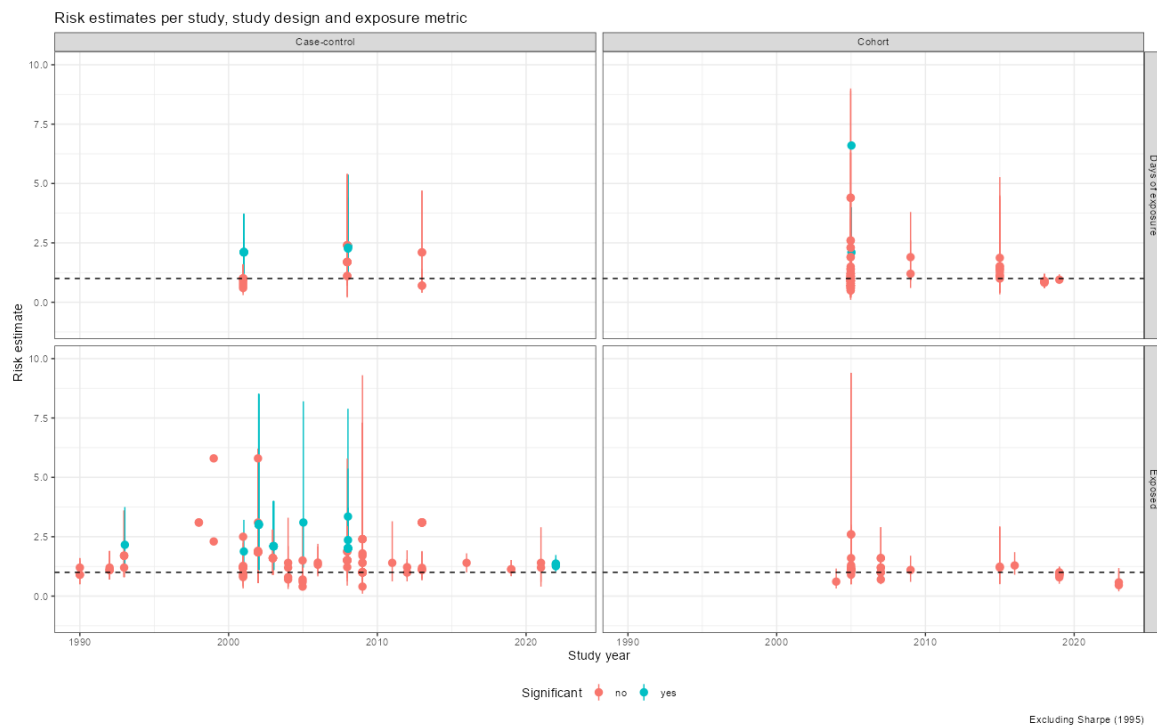


Figuur 19. De relatie tussen schattingen (*odds-ratio's* en *risk-ratio's*) en het jaar waarin de studies zijn gepubliceerd oer soort study design.

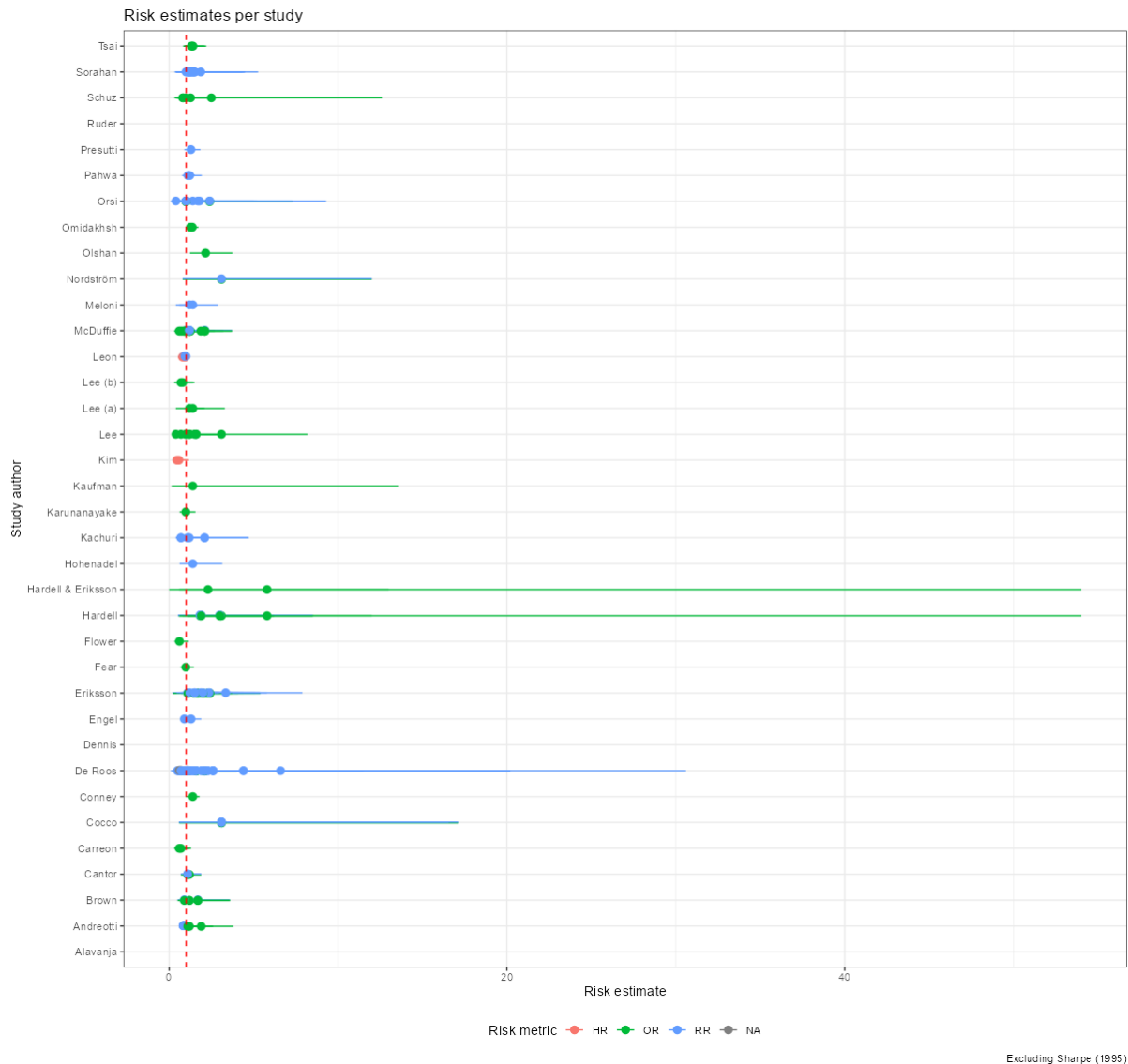
⁶⁴ En die soms ook echt kunnen verschillen tussen de reviews.



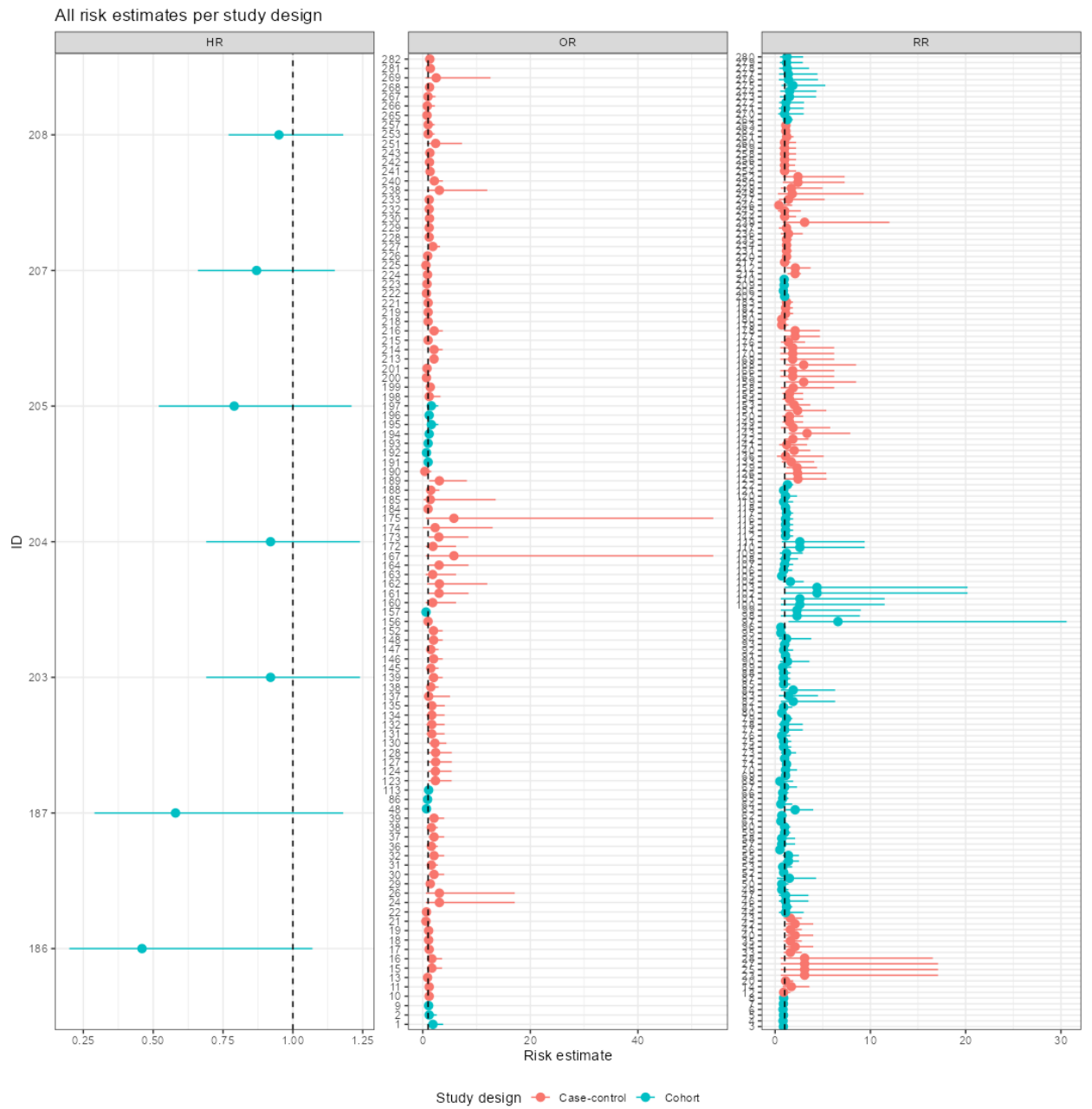
Figuur 20. Risicoschattingen per studie en *study design* waarbij de kleur aangeeft of een bevinding statistisch significant is of niet. De verticale lijnen laten de betrouwbaarheidsintervallen zien. De horizontale lijn is de lijn-van-geen-verschil.



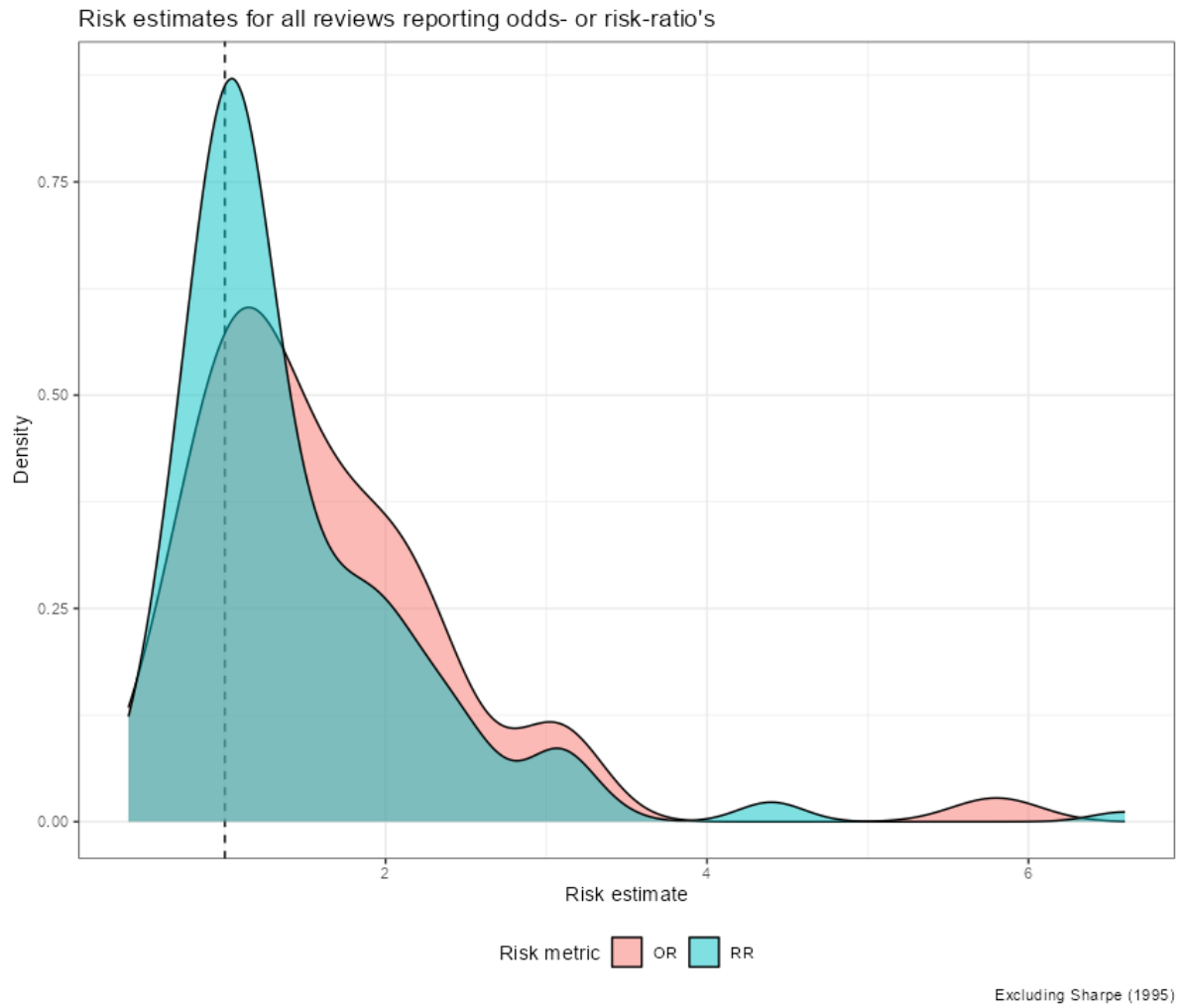
Figuur 21. Risicoschattingen per studie, *study design* en mate van blootstelling waarbij de kleur aangeeft of een bevinding statistisch significant is of niet. De verticale lijnen laten de betrouwbaarheidsintervallen zien. De horizontale lijn is de lijn-van-geen-verschil.



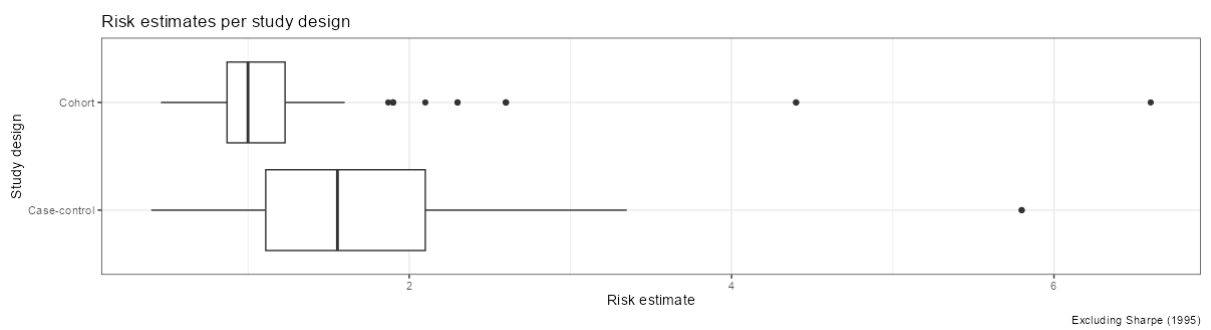
Figuur 22. Risicoschattingen per studie en statistiek. De verticale lijn is de lijn-van-geen-verschil.



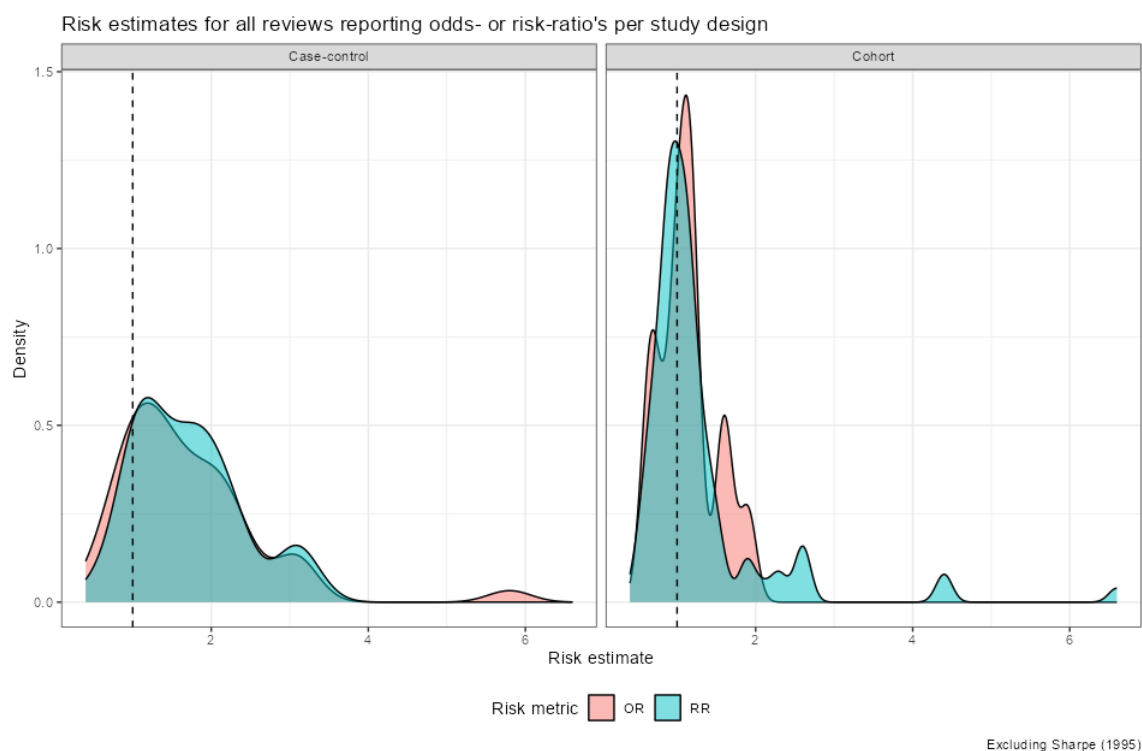
Figuur 23. Elke risicoschatting die is opgenomen in de database op basis van 20 reviews en 42 publicaties. De kolommen geven de specifieke statistiek weer en de kleur laat zien om welk soort *study design* het gaat.



Figuur 24. De verdeling van de *odds-ratio's* en *risk-ratio's* voor alle studies opgenomen in de database.



Figuur 25. De verdeling van de *odds-ratio's* en *risk-ratio's* voor alle studies opgenomen in de database per *study design*.



Figuur 26. De verdeling van de *odds-ratio's* en *risk-ratio's* per *study design* voor alle studies opgenomen in de database.

Om het een en ander te verduidelijken is het wellicht raadzaam om terug te keren naar de studie van De Roos(60) waarvan we de schatting tonen per review voor NHL en per soort blootstelling. De eerste figuur (**Figuur 27**) toont voor elke review de puntschatting (*estimate*) en de onzekerheid rondom die puntschatting. Omdat de onzekerheden rondom elke schatting in elke review de stippellijn raakt zijn ze allemaal niet significant bij een alfa-waarde van 0.05⁶⁵.

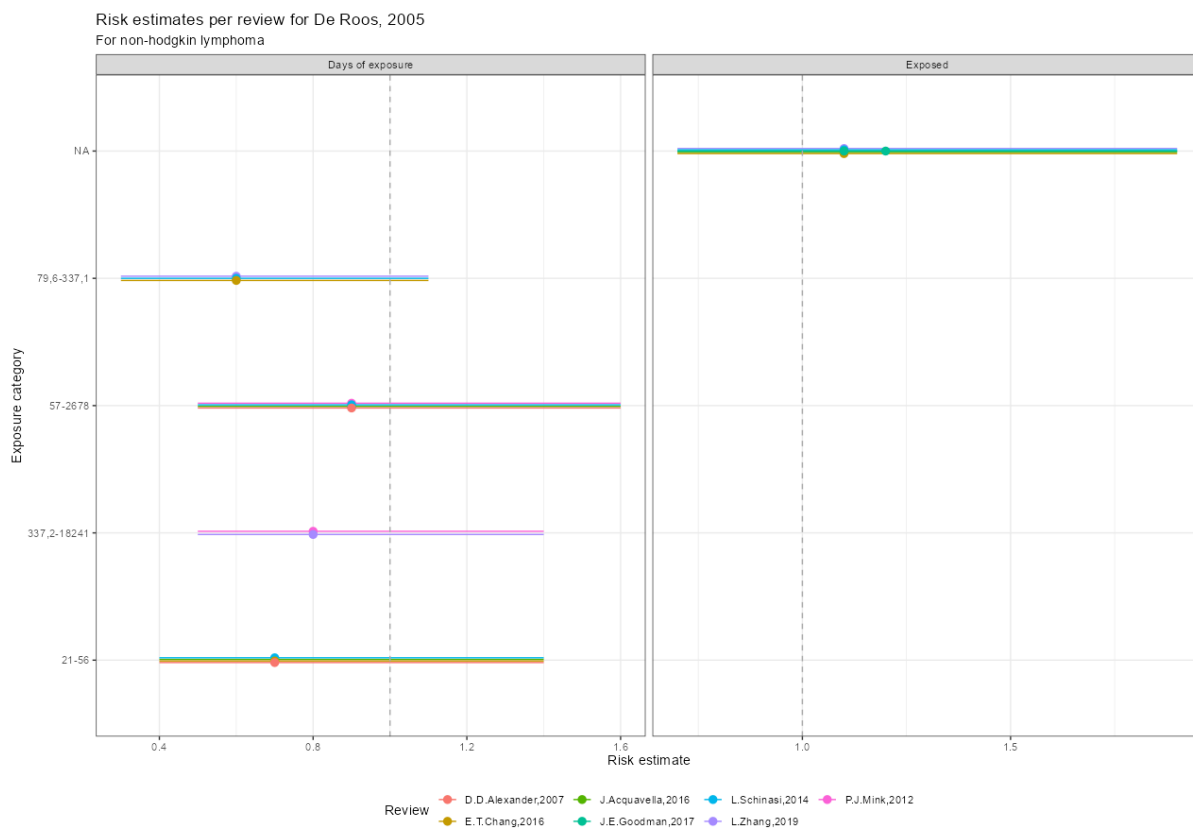
Wat verder opvalt is dat zowel de puntschattingen als de onzekerheden nagenoeg allemaal gelijk zijn, behalve bij de groep waarin we alleen kijken of er überhaupt blootstelling is geweest (de *exposed* groep). Dit heeft er mee te maken dat we wel kijken of er blootstelling is geweest, maar geen verder onderscheid kunnen maken in de mate van blootstelling. We zien dus alleen aan-of afwezigheid, maar geen mate van.

In de regel mag worden aangenomen dat hoe groter een betrouwbaarheidsinterval is, hoe groter ook de onzekerheid is⁶⁶. Nu is de reden dat we met meta-analyses werken voorkomt uit de wens om het signaal te versterken door studies samen te voegen, maar dit

⁶⁵ Zie mijn [rapport](#) over het toepassen van de eenzijdige toets voor een introductie in de Frequentistische statistiek.

⁶⁶ Of hoe kleiner de studiegrootte, maar die hangen in de regel samen.

voorbeeld alleen laat al zien hoe lastig samenvoegen het wel niet is als de rapportage van één enkele studie verschilt (hier in de *exposed* groep). We zagen dit ook al in **Figuur 15**.

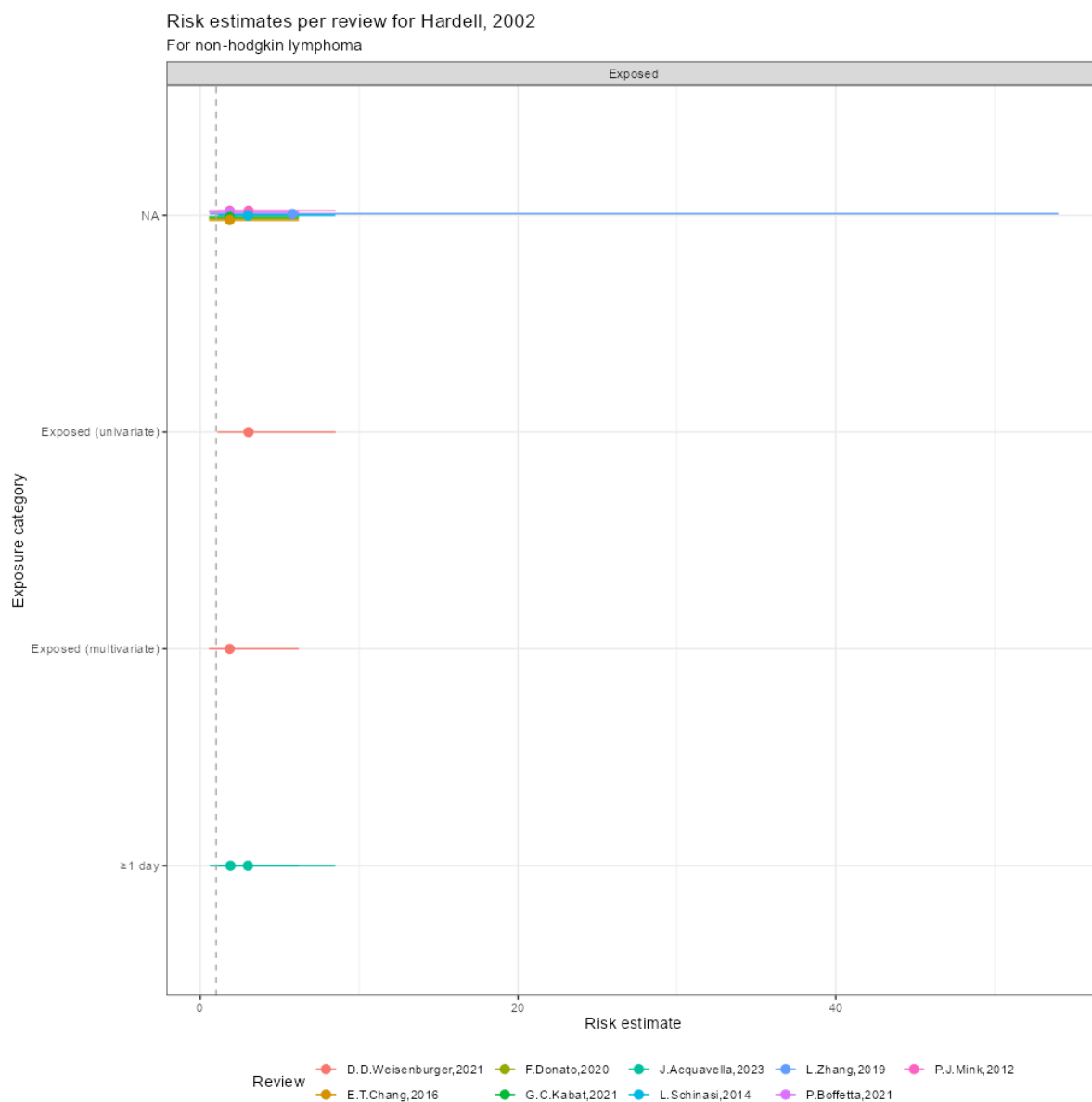


Figuur 27. Schattingen per mate van blootstelling zoals gerapporteerd in de studie van De Roos(60). De kleuren staan voor de informatie zoals aangeduid in elke review die deze studie heeft opgenomen.

De bevinding dat reviews niet hetzelfde rapporteren als ze over dezelfde studie schrijven is een hele belangrijke, want het betekent dat de data niet zomaar beschrijvend mag worden samengevoegd. Je moet namelijk wel weten wat een cijfer exact betekent en die informatie staat niet zomaar beschreven in elke review. Dit kunnen we duidelijk zien in **Figuur 28** waar we kijken naar de studie van Hardell (66).

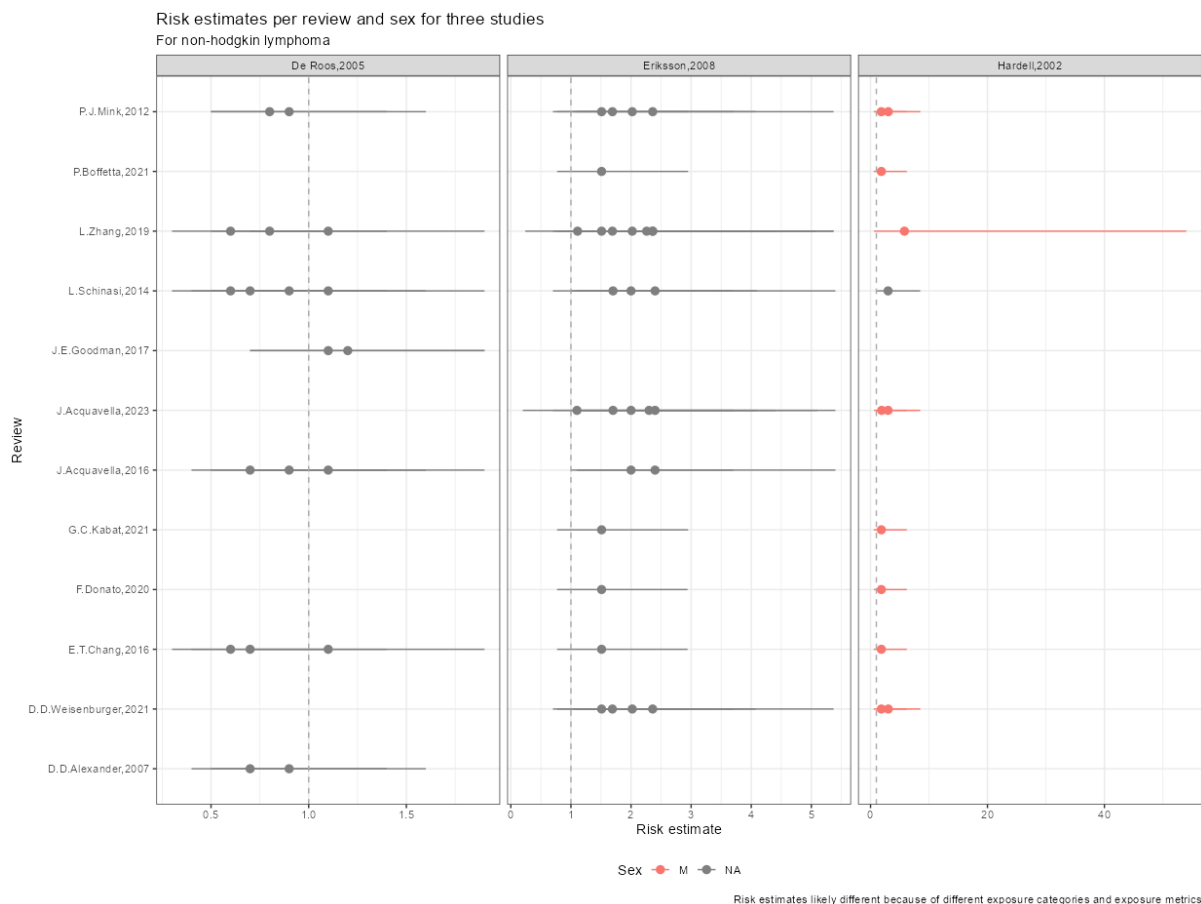
Sommige studies hebben ook gerapporteerd over het geslacht (*sex*) waar de studie betrekking op heeft, maar die informatie is niet altijd eenduidig te achterhalen (is er bijvoorbeeld sprake van geen geslachts-specifieke analyse als er niks over geslacht vermeld wordt?). Ook kan de mate van rapporteren verschillen per review. Dit kunnen we zien door naar die drie studies te kijken die het vaakst zijn meegenomen in reviews: De Roos(60), Eriksson(64) en Hardell(66). Het resultaat staat in **Figuur 29** en laat zien dat voor de studie

van Hardell(66) het niet in elke review duidelijk was om het nou alleen om mannen ging. Voor De Roos(60) en Eriksson(64) was meer duidelijk⁶⁷.



Figuur 28. Schattingen per mate van blootstelling zoals gerapporteerd in de studie van Hardell (66). De kleuren staan voor de informatie zoals aangeduid in elke review die deze studies heeft opgenomen.

⁶⁷ Nu verwacht ik sowieso geen verschil tussen geslacht, maar het gaat mij er meer om dat ik kan laten zien dat niet elke studie even duidelijk is in wat men rapporteert. Ook zien we in **Figuur 29** duidelijk dat de schattingen verschillen per studie en in een enkele studie op basis van verschillende factoren. Het samenvoegen van schattingen is dus geen eenvoudige klus en vereist het maken van keuzes die weer van invloed zullen zijn op het resultaat.



Figuur 29. Schattingen voor drie studies die de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL) hebben onderzocht. Per studie zien we de schattingen per review en of het duidelijk was of de schatting per geslacht is gedaan of niet.

Relatie tussen risicoschatting en de onzekerheid van die schatting

Bovenstaande voorbeelden laten zien hoe lastig het werken met een grote en heterogene database is. Die last komt voornamelijk voort uit de wens om gegevens zinvol te groeperen en de wens van de onderzoekers om met nieuwe bevindingen en relaties te komen. Echte replicatiestudies zien we dan ook niet in dit landschap. We kunnen er nu dus voor kiezen om te concluderen dat de studies waarschijnlijk allemaal te verschillend zijn om te groeperen óf we kunnen er voor kiezen om te ontdekken wat de relatie is tussen een risicoschatting en de onzekerheid zoals gevonden in de studie zelf. Het afbeelden en onderzoeken van deze mogelijke relatie is zinvol omdat het iets zegt over de betrouwbaarheid in én tussen studies. Het zegt uiteindelijk ook iets over de betrouwbaarheid van elke uitspraak die in dit dossier gemaakt wordt.

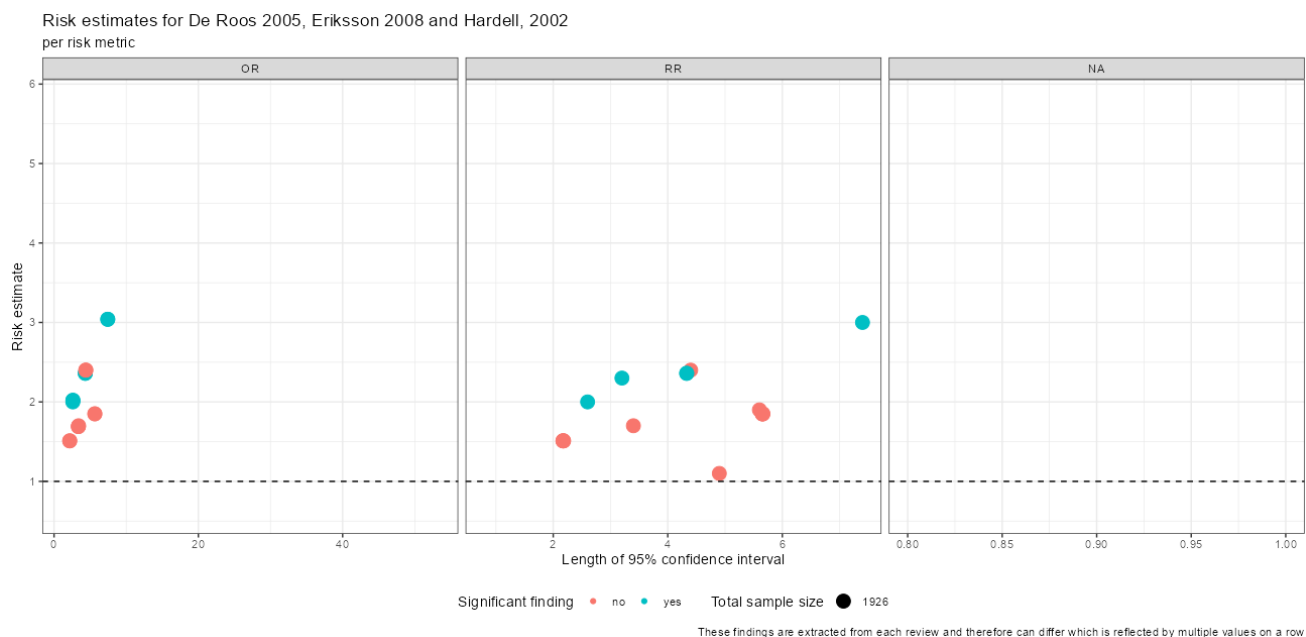
Om te laten zien hoe uitdagend het interpreteren van de studieresultaten is, kan ik aan de studies van De Roos(60), Eriksson(64) was en Hardeil(66) ook de totale studiegrootte

koppelen⁶⁸. Deze speelt in de traditionele meta-analyse een rol in het bepalen van het gewicht dat een studie krijgt wanneer deze wordt samengevoegd met andere studies. Dit komt omdat de studiegrootte een grote invloed heeft op de onzekerheid. Hoe minder onzeker, hoe groter dus het gewicht dat een studie krijgt in de meta-analyse wat betekent dat een grotere studie vaker een groter gewicht krijgt toegekend. Omdat we over dezelfde studie spreken (maar dan vanuit verschillende reviews) verwachten we eigenlijk dat alles hetzelfde is, maar dat blijkt dus niet zo te zijn. De vraag is nu of het überhaupt lukt om een soort van samengestelde waarde te maken voor deze studies?

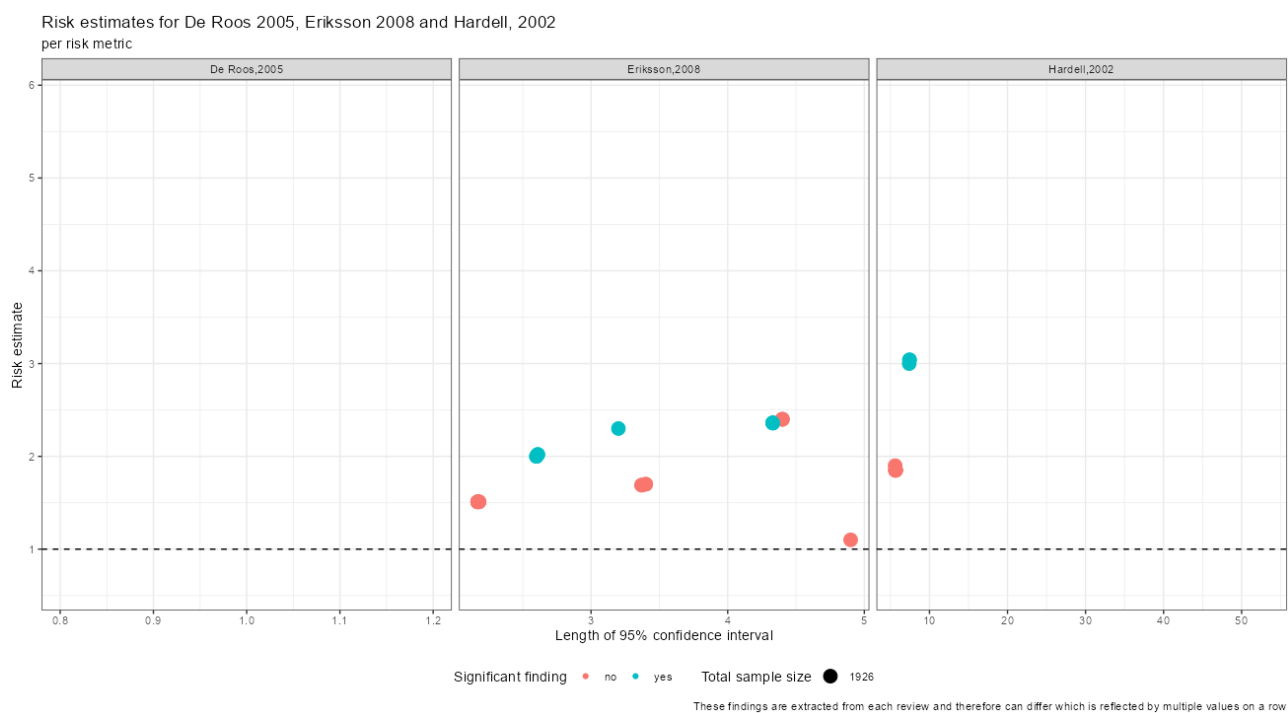
Om te beginnen zal ik voor alle drie de studies afbeelden hoe de relatie tussen de puntschatting, de onzekerheid én studiegrootte eruit ziet. Ik zal dit voor de verschillende statistieken apart doen en je kunt ook zien welke bevindingen wel als significant⁶⁹ worden bestempeld en welke niet (**Figuur 30**). Wat direct opvalt is het tekort aan informatie. Zo lijkt het alsof er geen onderscheid tussen de studies wordt gemaakt. Zo is de totale studiegrootte voor Eriksson(64) en Hardell(66) exact hetzelfde en ontbreekt die informatie voor De Roos(60) (**Figuur 31**). Omdat we niet kunnen differentiëren kunnen we ook geen relatie laten zien, maar het plaatje zelf is wel een bijzondere bevinding waar we later dieper op in moeten gaan. Het lijkt er namelijk sterk op dat verschillende publicaties met dezelfde data aan de slag zijn gegaan. Of ik ben ergens flink tekortgeschoten in het extraheren van de gegevens.

⁶⁸ Als men dit al enigszins heeft gerapporteerd, maar is natuurlijk alweer een bevinding op zichzelf.

⁶⁹ Op basis van 95% betrouwbaarheidsinterval.



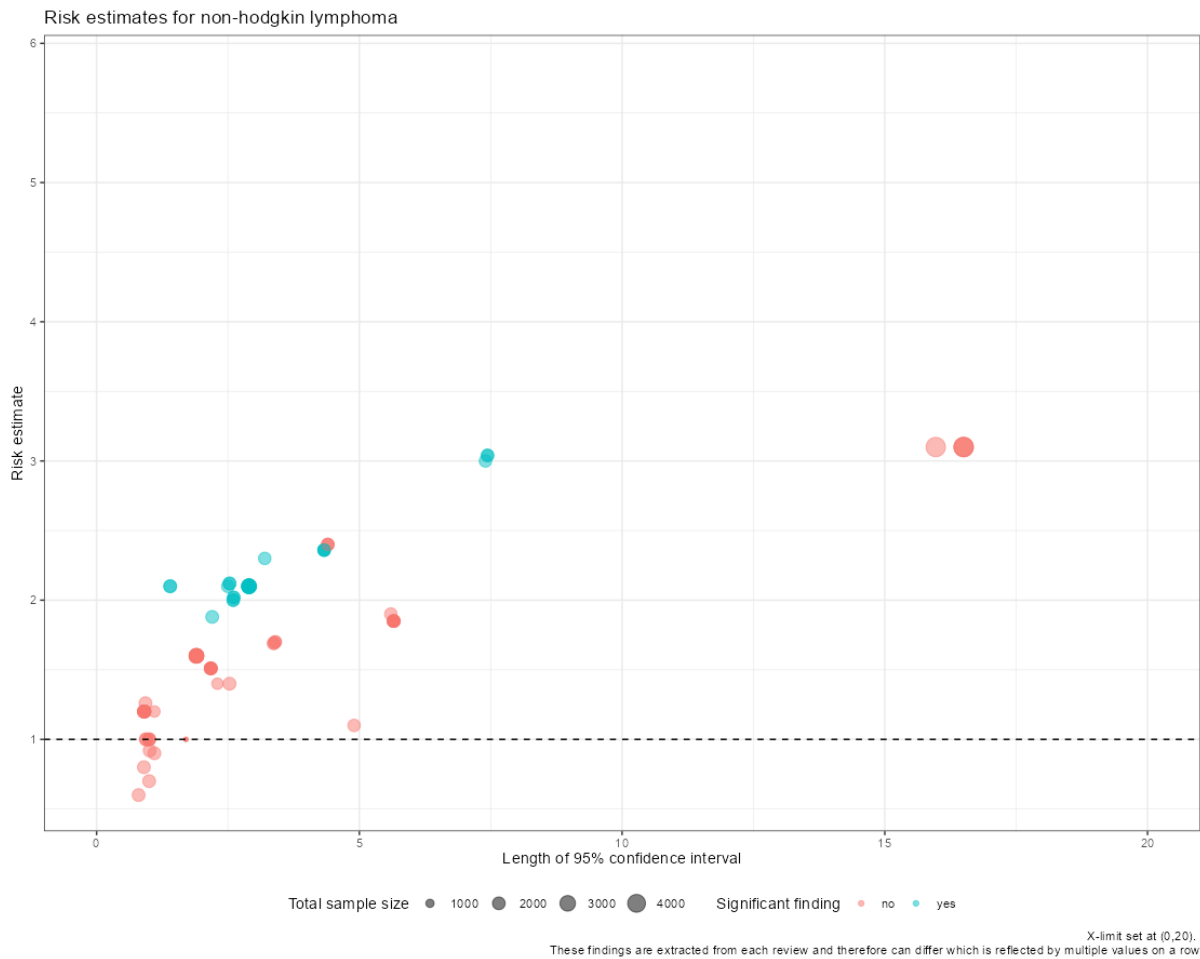
Figuur 30. Risicoschattingen voor De Roos(60), Eriksson (64) en Hardell (66) zoals opgenomen in verschillende reviews, en per statistiek.



Figuur 31. Risicoschattingen voor De Roos(60), Eriksson(64) en Hardell(66) zoals opgenomen in verschillende reviews.

Misschien helpt het als we deze figuur uitbreiden met daarin opgenomen alle studies die naar NHL hebben gekeken? Ik zal nu ook het onderscheid tussen de verschillende statistieken weglaten. Het resultaat zien we in **Figuur 32**. De relatie tussen de puntschatting

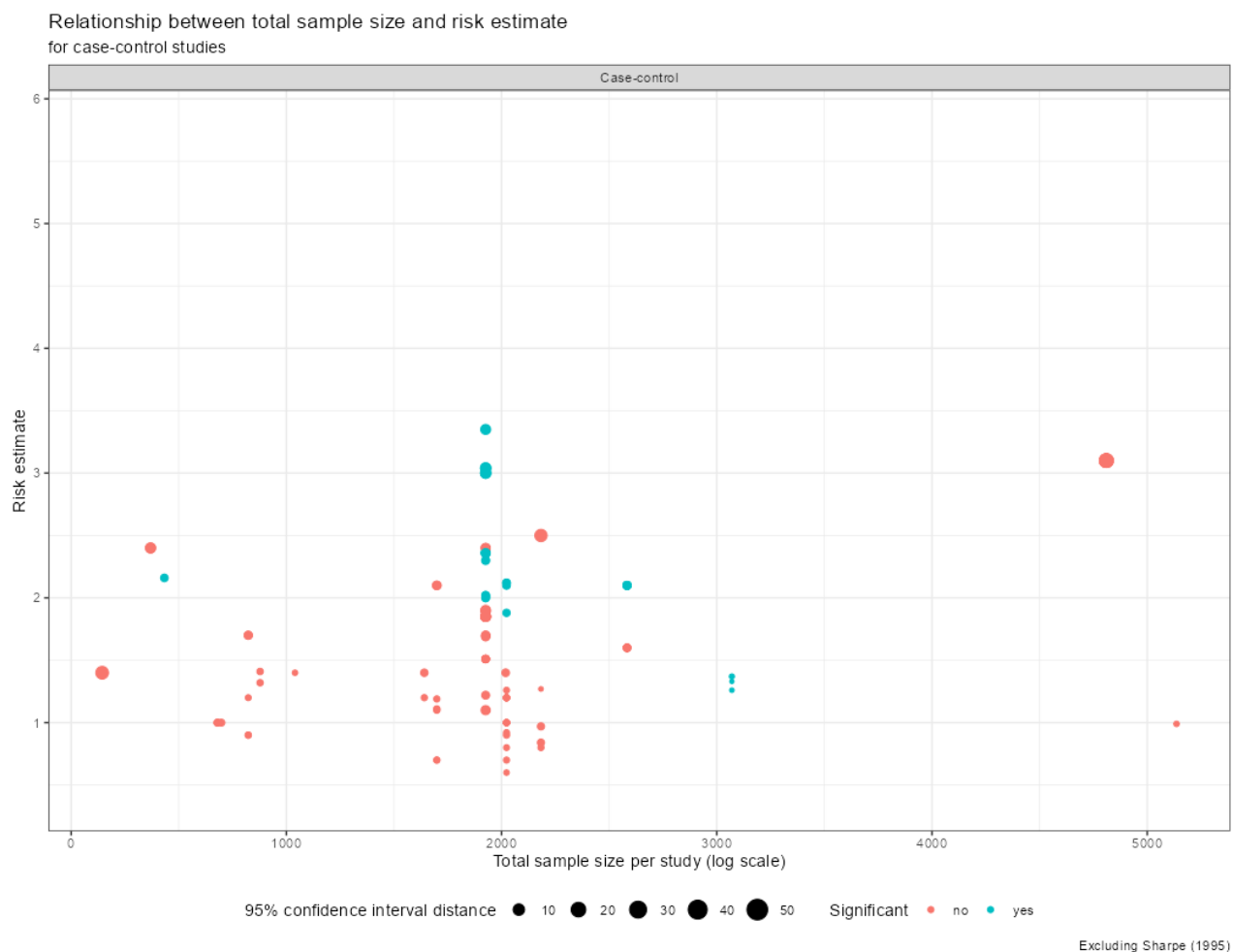
en de breedte van het 95% betrouwbaarheidsinterval lijkt een curve te tonen wanneer de bevinding niet-significant is, en lineair te zijn wanneer deze wel significant is. In hoeverre ik hier een conclusie uit kan trekken is mij niet duidelijk want ik heb geen directe verklaring paraat. Ook valt het wederom op hoe weinig informatie we hebben voor de ziekte het vaakst is onderzocht.



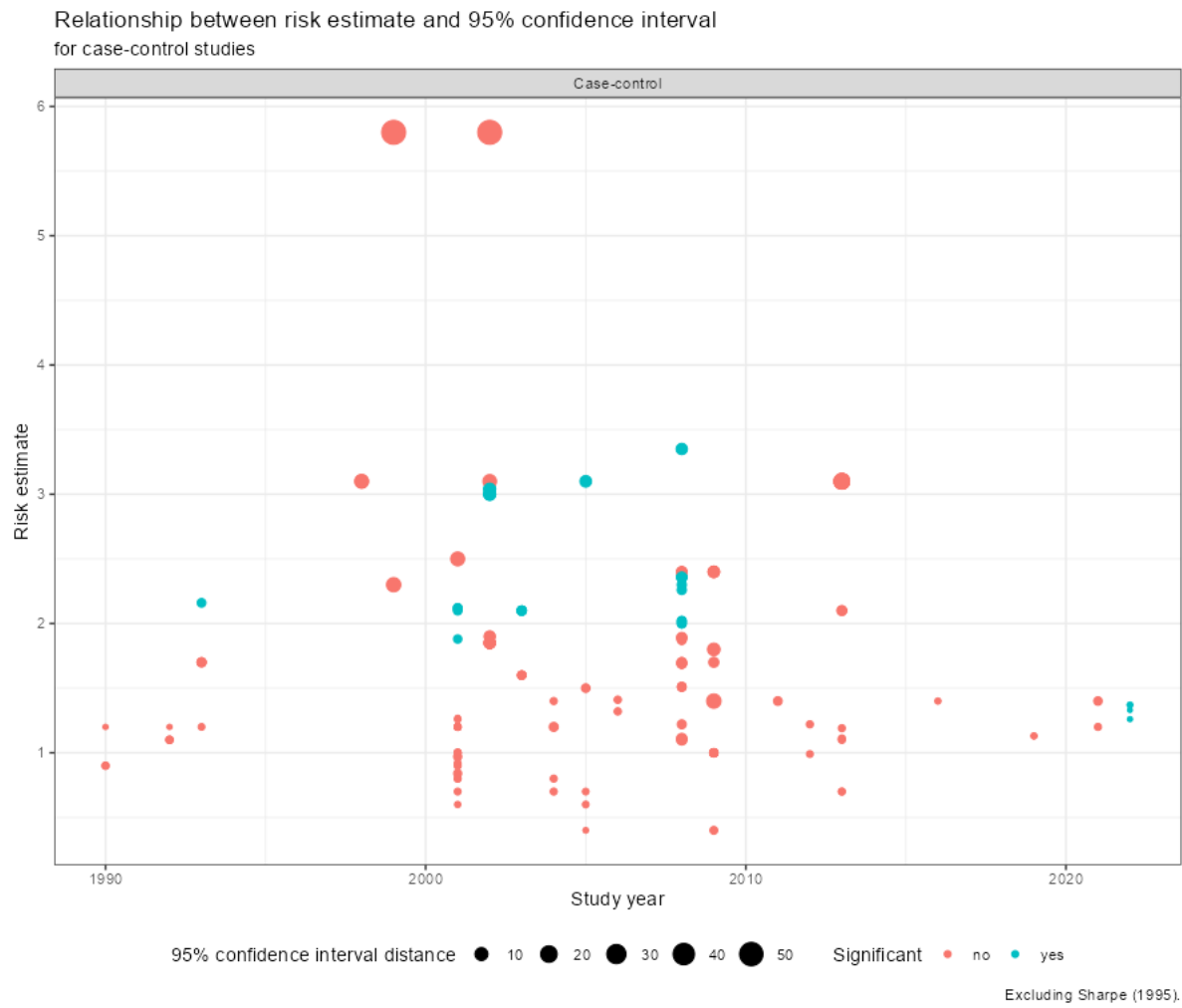
Figuur 32. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat op non-hodgkin lymfoom (NHL).

Het beeld wordt al iets anders als we alle schattingen voor alle case-control studies tonen (**Figuur 33** en **Figuur 34**). Er lijkt dus wat data voorhanden te zijn. We zien dit ook wanneer we alle schattingen voor elke kankersoort tonen (**Figuur 35**). We zien hier duidelijk een cluster linksonder, maar wanneer we studiegrootte meenemen blijkt maar weer eens hoe weinig informatie we hebben over deze belangrijke metriek (**Figuur 36**). De cluster blijft wel behouden, maar er lijkt nu ook een soort van relatie te ontstaan. Een gebrek aan data mag

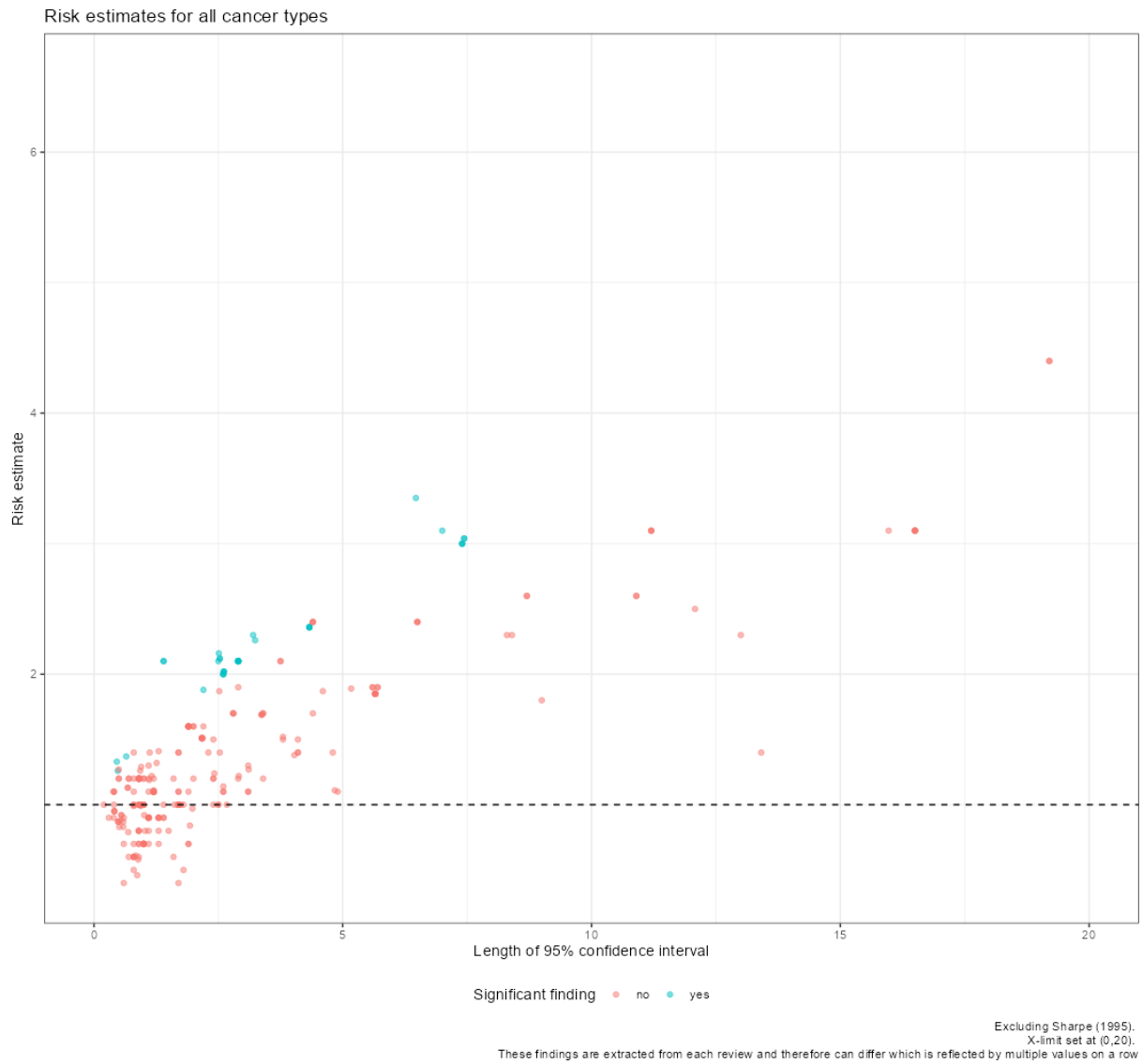
echter geen reden zijn om ineens wel een relatie te zien dus ik laat dit even voor wat het is. De relatie tussen risicoschattingen en onzekerheden per *study design* en statistiek zien we in **Figuur 37** en **Figuur 38**, respectievelijk.



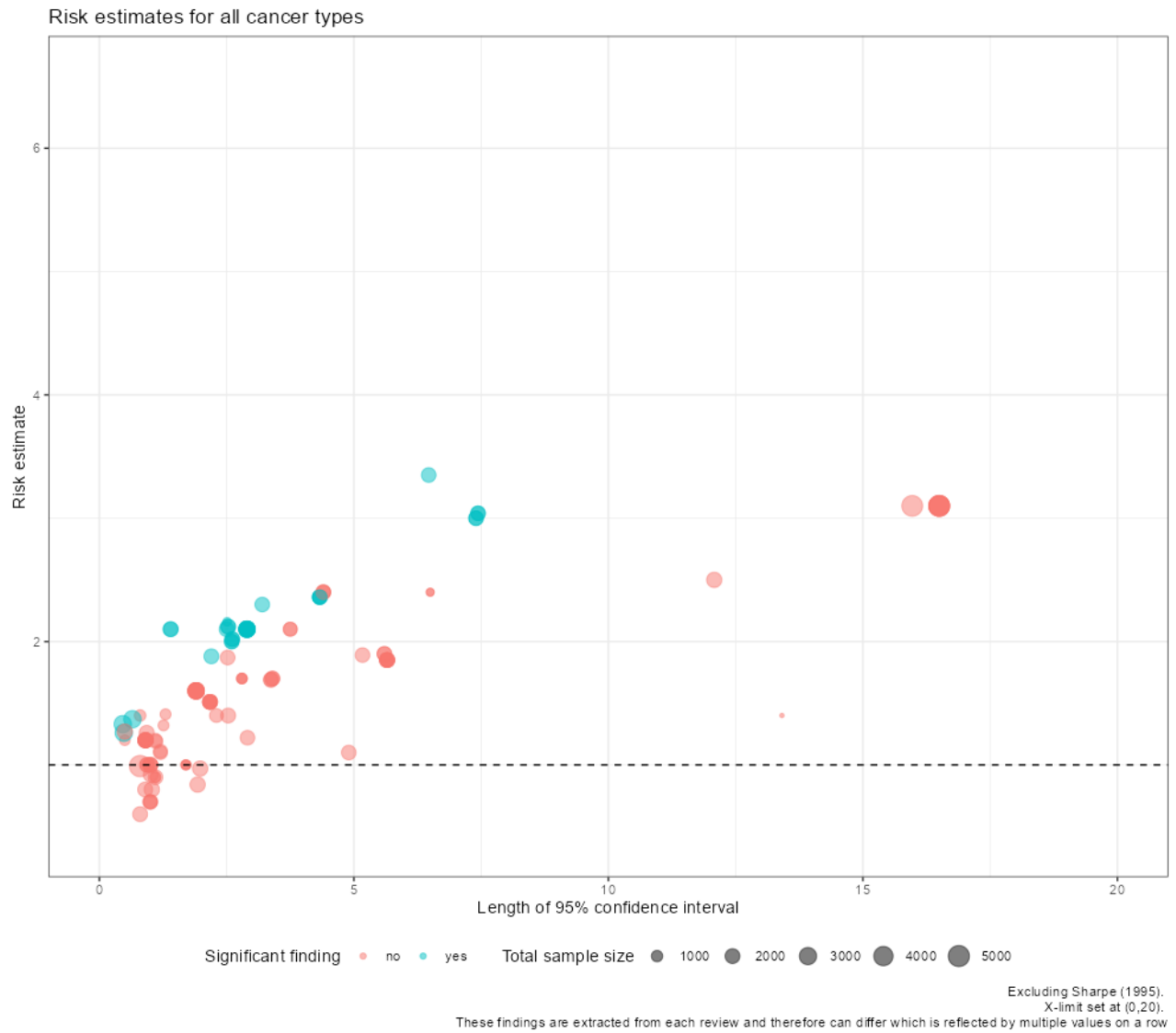
Figuur 33. Risicoschattingen voor alle *case-control* studies die hebben gekeken naar de invloed van glyfosaat. Ook getoond wordt de lengte van het 95% betrouwbaarheidsinterval en de totale studiegrootte (indient de data beschikbaar is).



Figuur 34. Risicoschattingen voor alle *case-control* studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval.



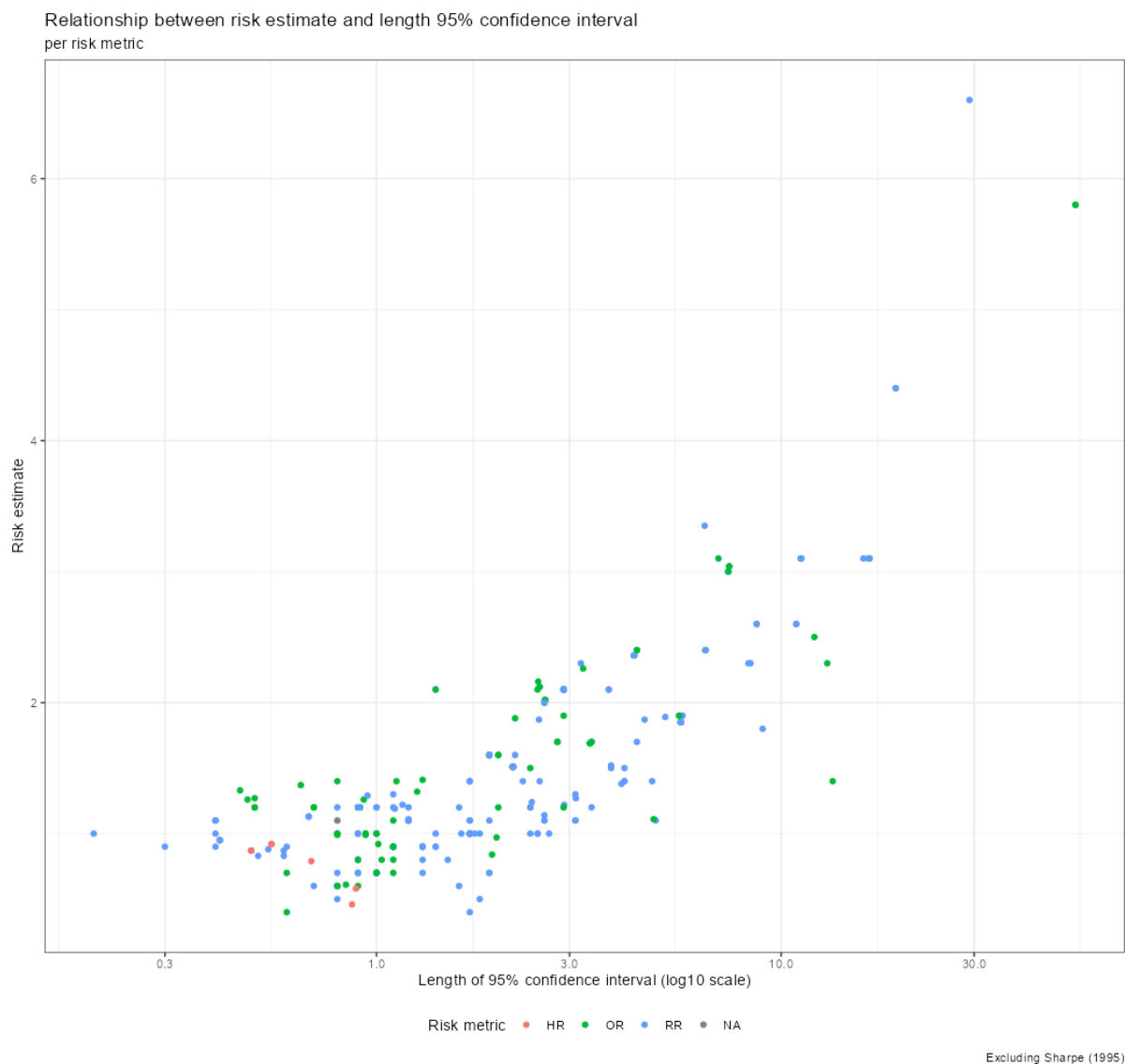
Figuur 35. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat met daarbij de lengte van het 95% betrouwbaarheidsinterval.



Figuur 36. Risicoschattingen voor alle studies met daarbij vermeld de totale studiegrootte en de lengte van het 95% betrouwbaarheidsinterval.



Figuur 37. Risicoschattingen voor alle studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval per *study design*.



Figuur 38. Risicoschattingen voor alle studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval per soort statistiek.

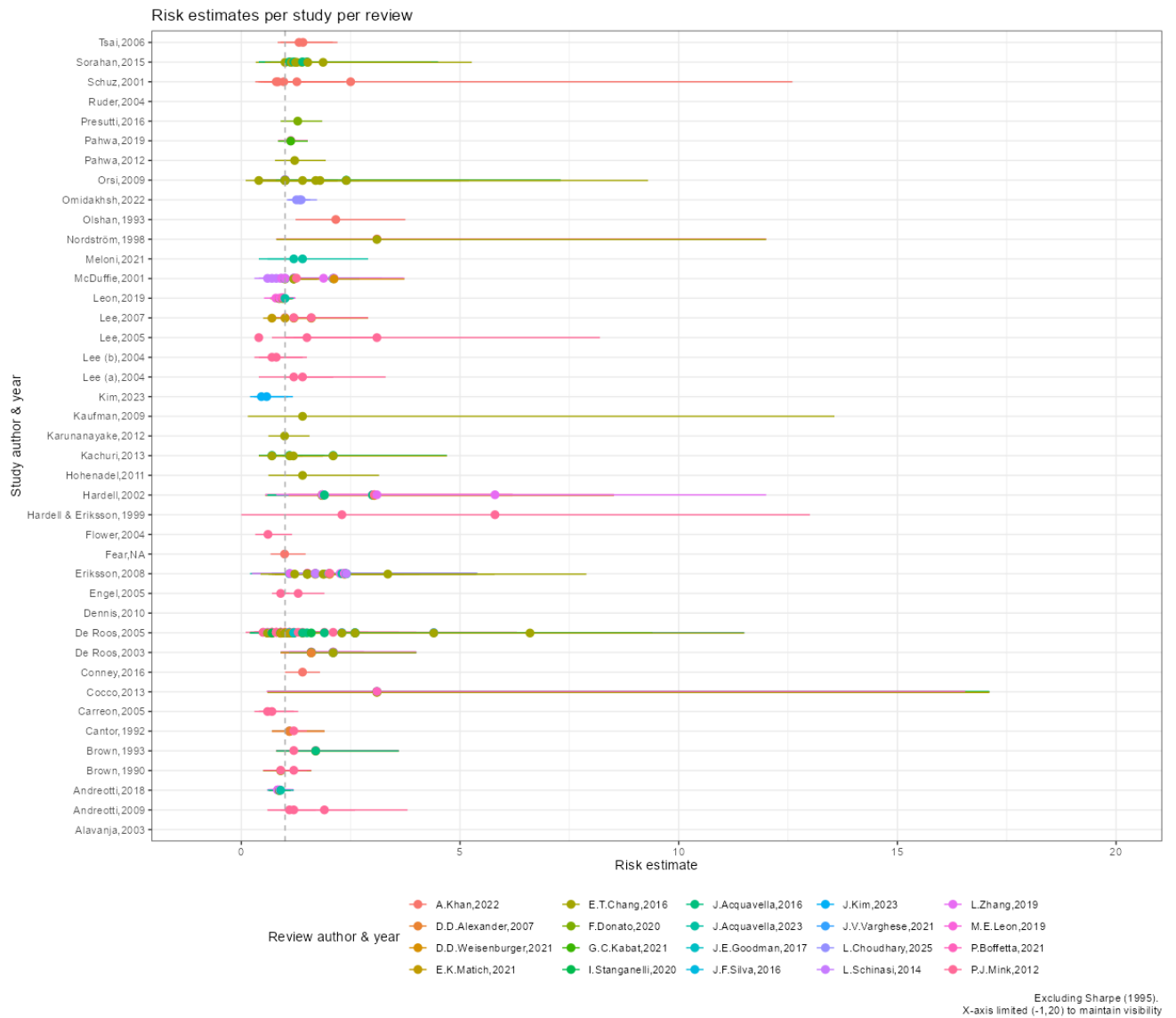
Verschillen tussen reviews in studies voor glyfosaat en kanker

Wat tot nu toe opvalt is dat de studies voor glyfosaat bovenal niet-significante resultaten laten zien (**Figuur 12** en **Figuur 13**) hoewel het voor nu onduidelijk is hoe dit komt. **Figuur 13** liet verder zien hoeveel significantie en niet-significante resultaten we hebben, en **Figuur 11** en **Figuur 14** lieten al zien dat voor een bepaalde studie (in **Figuur 15** was het bijvoorbeeld veel minder evident) er verschillen kunnen zijn tussen reviews (in wat ze rapporteren en ook de exacte waarde die men rapporteert). Verder splitsen per review zou daarom zinvol zijn, maar ook automatisch beteken dat we ook een splitsing moeten maken voor de mate van blootstelling en een splitsing voor de gehanteerde metriek/statistiek. Dan is er natuurlijk nog de splitsing voor de exacte tumorsoort en we zouden zelfs kunnen splitsen voor geslacht.

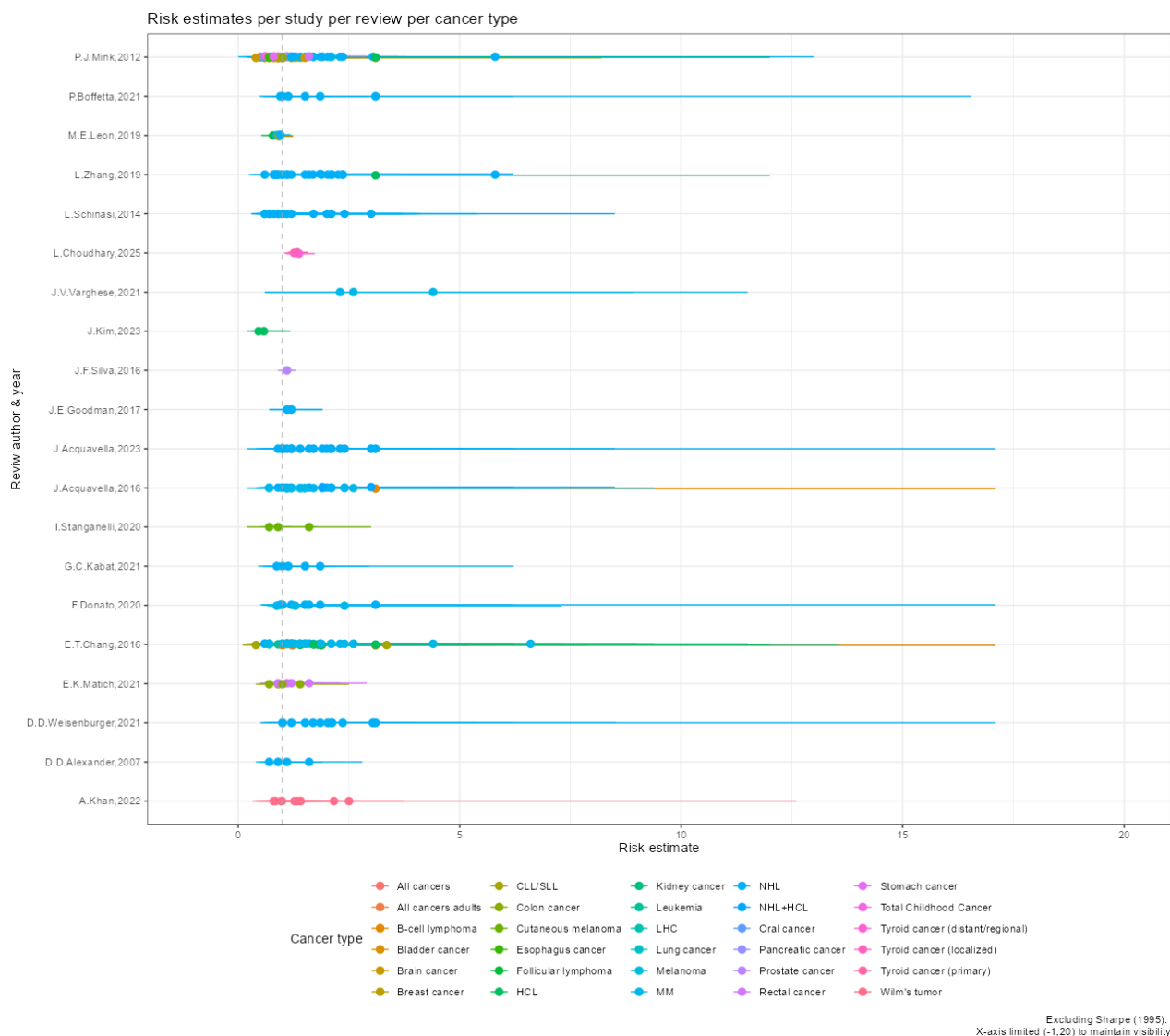
Uiteindelijk zouden we zo vaak splitsen dat we geen data meer overhouden om per onderdeel zinvol te duiden wat de relatie tussen glyfosaat en kanker is. We hebben in dit rapport al eerder gezien hoe snel we op dat punt uitkomen.

Toch wil (en moet) ik graag dieper blijven stil staan bij de verschillen tussen reviews omdat het een beeld geeft van de informatie zoals die (over de tijd) beschikbaar is. We kunnen ervoor kiezen om dit over kankersoorten heen te doen (zoals in **Figuur 9**), maar dan zien we niet goed hoe de schattingen verschillen per review. Het is daarom wellicht wijs om de data langzaam op te splitsen om zo te zien wat een splitsing doet met het beeld dat we hebben van de relatie tussen glyfosaat en kanker.

Een zinvolle start is het grafisch weergeven van alle risicoschattingen per review, met hun onzekerheden, zoals die zijn opgenomen per review (**Figuur 39**). We kunnen er ook voor kiezen om een kleur toe te voegen die aangeeft om welke kankersoort het gaat. Dan ziet het er allemaal iets anders uit (**Figuur 40**). We kunnen beide afbeeldingen ook weer versimpelen door ze als *boxplots* af te beelden (**Figuur 41**). Dit geeft een beter beeld van de spreiding.



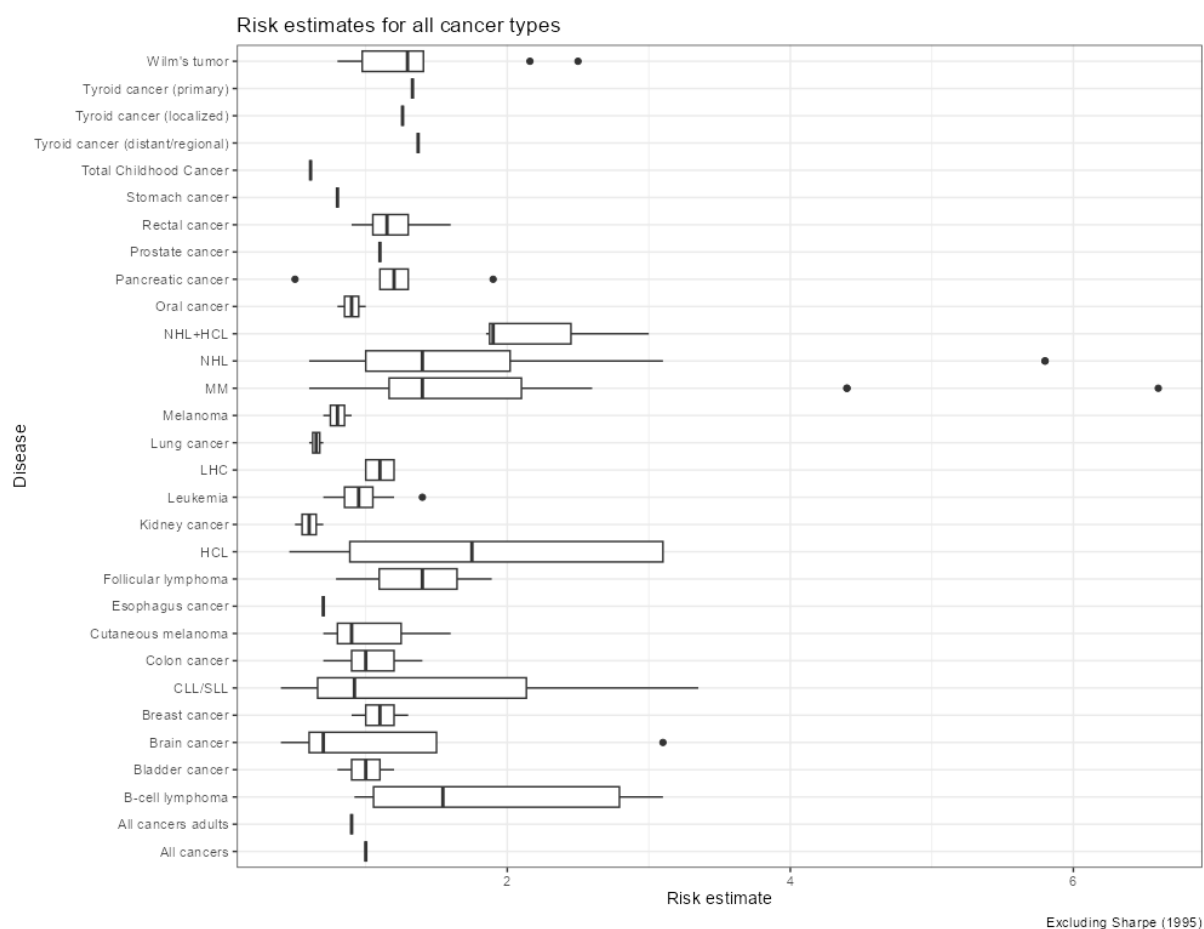
Figuur 39. Risicoschattingen per studie zoals opgenomen in de reviews.



Figuur 40. Risicoschattingen met 95% betrouwbaarheidsintervallen per studies zoals opgenomen in elke review.

Ik ben nu ook wel benieuwd hoe **Figuur 40** eruit komt te zien als we deze afbeelden zoals we **Figuur 35** hebben gedaan. Een dergelijke afbeelding laat namelijk zien of we van zogenaamde clusters kunnen spreken: regio's waarin risicoschattingen samenklonteren omdat er ergens een gedeelde noemer is. Ik gok dat die gedeelde noemer de kankersoort zelf is, maar het resultaat (**Figuur 42**) toont geen eenduidig beeld. Wellicht helpt het als we de focus verschuiven naar kankersoorten waarvan we weten dat er zowel niet-significante als significante resultaten zijn. We zagen in **Figuur 9** al dat het dan gaat om non-hodgkin lymfoom (NHL), hersenkanker, leukemie (CLL/SLL), en beenmergkanker (MM). Willen we zien of reviews verschillen dan zullen we een zoom-in op deze kankersoorten moeten doen. Het

kan dan heel goed zijn dat de reviews verschillen omdat ze verschillende studies hebben opgenomen (**Figuur 39**).

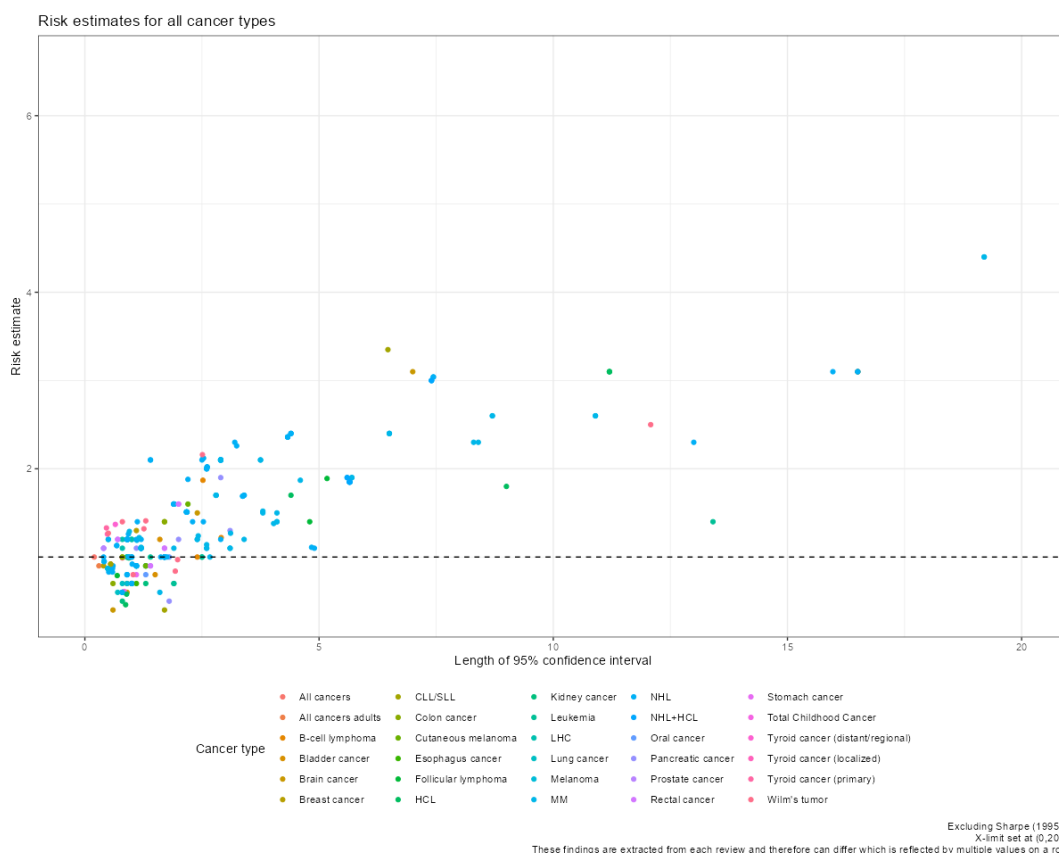


Figuur 41. Risicoschattingen voor elke kankersoort in *boxplot* formaat.

Uiteindelijk blijkt dat voor hersenkanker er maar één enkele review is, dus deze kunnen we niet meenemen in het vergelijk:

disease	review_author_year	significant	n
<chr>	<chr>	<chr>	<int>
1 Brain cancer	P.J.Mink,2012	no	6
2 Brain cancer	P.J.Mink,2012	yes	1
3 CLL/SLL	E.T.Chang,2016	no	1
4 CLL/SLL	E.T.Chang,2016	yes	1
5 CLL/SLL	M.E.Leon,2019	no	1
6 MM	E.T.Chang,2016	no	21
7 MM	E.T.Chang,2016	yes	1
8 MM	F.Donato,2020	no	3
9 MM	J.Acquavella,2016	no	17
10 MM	J.V.Varghese,2021	no	3
11 MM	M.E.Leon,2019	no	1
12 MM	P.J.Mink,2012	no	3
13 MM	P.J.Mink,2012	yes	1
14 NHL	D.D.Alexander,2007	no	4
15 NHL	D.D.Weisenburger,2021	no	7
16 NHL	D.D.Weisenburger,2021	yes	5
17 NHL	E.T.Chang,2016	no	10

18	NHL	E.T.Chang, 2016	yes	1
19	NHL	F.Donato, 2020	no	7
20	NHL	G.C.Kabat, 2021	no	5
21	NHL	J.Acquavella, 2016	no	8
22	NHL	J.Acquavella, 2016	yes	3
23	NHL	J.Acquavella, 2023	no	13
24	NHL	J.Acquavella, 2023	yes	5
25	NHL	J.E.Goodman, 2017	no	2
26	NHL	L.Schinasi, 2014	no	13
27	NHL	L.Schinasi, 2014	yes	4
28	NHL	L.Zhang, 2019	no	16
29	NHL	L.Zhang, 2019	yes	7
30	NHL	M.E.Leon, 2019	no	1
31	NHL	P.Boffetta, 2021	no	6
32	NHL	P.J.Mink, 2012	no	14
33	NHL	P.J.Mink, 2012	yes	4

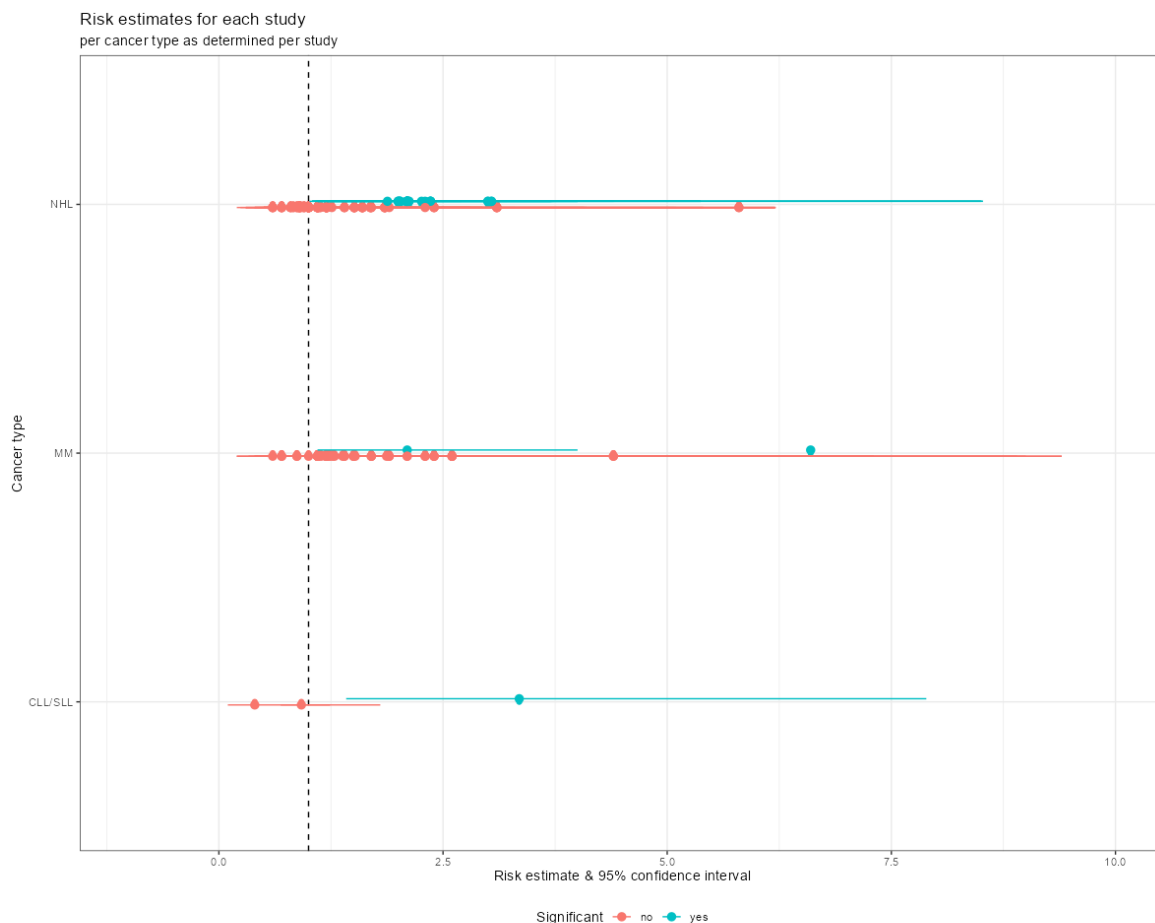


Figuur 42. Risicoschattingen en lengte van het 95% betrouwbaarheidsinterval per kankersoort.

Dit maakt dat we alleen nog kunnen kijken naar leukemie, beenmergkanker en non-hodgkin lymfoom. Als we opsommen hoeveel significante en niet significante bevindingen we hebben voor elk van deze kankersoorten dan krijgen we de volgende tabel en **Figuur 43**:

disease	significant	n	
<chr>	<chr>	<int>	
1	CLL/SLL	no	2
2	CLL/SLL	yes	1
3	MM	no	48
4	MM	yes	2
5	NHL	no	106
6	NHL	yes	29

Duidelijk wordt dat de aantallen voor deze drie kankersoorten verschillen. Ook zien we, wederom, dat het gros niet statistisch significant is. Alleen voor NHL zien we aardig wat significante schattingen, maar onduidelijk is of dit verschillen zijn tussen de reviews voor eenzelfde studie. Wederom blijkt dat we alleen voor NHL een eventueel zinvolle splitsing zouden kunnen maken, puur omdat er genoeg data is.

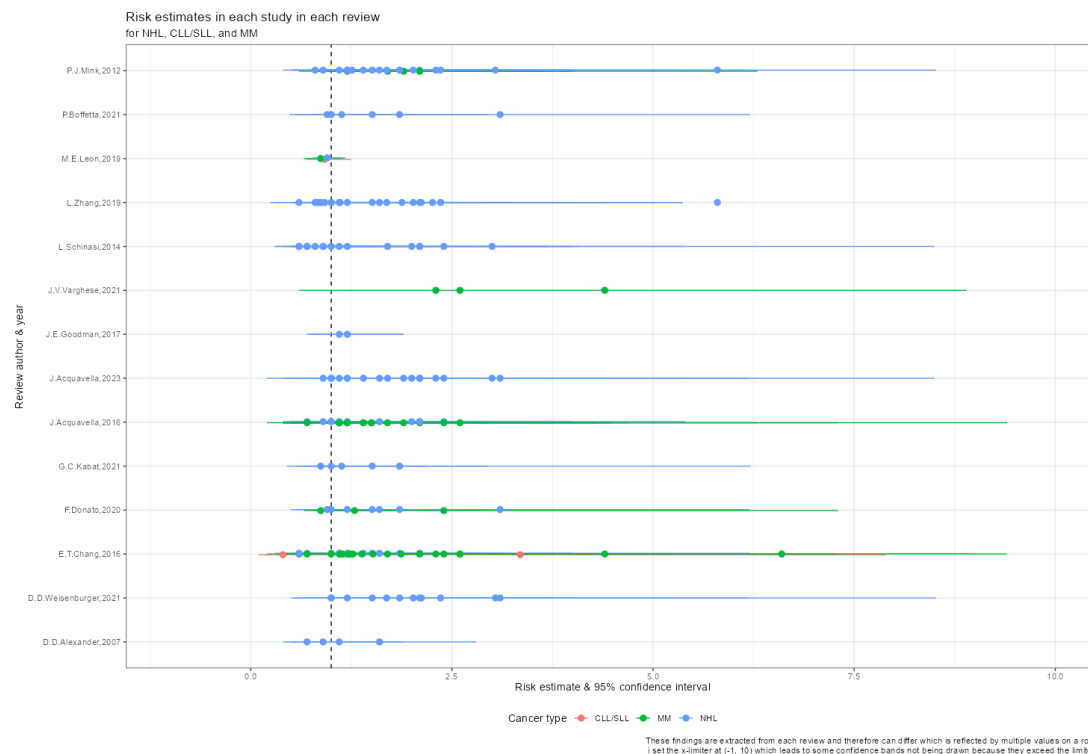


Figuur 43. Risicoschattingen en 95% betrouwbaarheidsintervallen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en leukemie (CLL/SLL). De kleur geeft statistische significantie aan.

Ik wil het toch nog een keer anders aanvliegen. Ik zou graag willen zien welke reviews welke studies hebben meegenomen voor elk van deze drie kankersoorten. Het blijkt dat veel reviews zich namelijk maar op één van deze drie kankersoorten hebben gericht met als uitzondering Chang(36) en Leon(50). In **Figuur 44** zien we elke risicoschatting voor elke studie in elke review die keek naar één of meer kankersoorten. We kunnen duidelijk zien dat

de schattingen in de reviews verschillen, maar ook de schattingen over de reviews heen. Verder zien we, wederom, behoorlijk wat betrouwbaarheidsintervallen met een flinke lengte.

review_author_year	n
<chr>	<int>
1 D.D.Alexander,2007	1
2 D.D.Weisenburger,2021	1
3 E.T.Chang,2016	3
4 F.Donato,2020	2
5 G.C.Kabat,2021	1
6 J.Acquavella,2016	2
7 J.Acquavella,2023	1
8 J.E.Goodman,2017	1
9 J.V.Varghese,2021	1
10 L.Schinasi,2014	1
11 L.Zhang,2019	1
12 M.E.Leon,2019	3
13 P.Boffetta,2021	1
14 P.J.Mink,2012	2



Figuur 44. Risicoschattingen en 95% betrouwbaarheidsintervallen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en leukemie (CLL/SLL) per review opgenomen in de database. De kleur geeft de specifieke kankersoort aan.

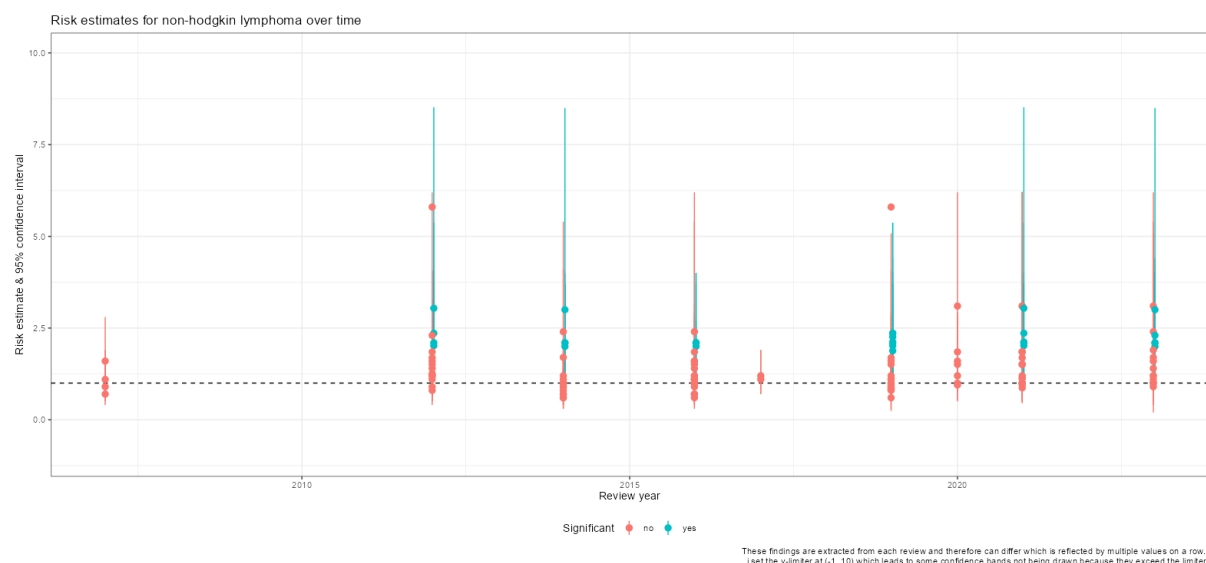
Als laatste wil ik alle schattingen laten zien voor alle reviews die naar NHL hebben gekeken. Het mag denk ik nu wel duidelijk zijn dat dit de enige kankersoort is waar we voldoende gegevens van hebben om eventuele splitsingen in de data te rechtvaardigen (en zinvol kunnen tonen). Het mag denk ik ook al duidelijk zijn dat in de latere uitstap naar de

dierstudies de primaire focus op deze kankersoort ligt (of moet liggen), maar dat punt kan en wil ik pas later adresseren.

In **Figuur 45** zien we het resultaat. Het wordt direct duidelijk dat niet elke studie in elke review voorkomt en dit gegevens hebben we al vaker gezien. Het wordt ook duidelijk dat niet elke review dezelfde resultaten zal tonen (ook dat hebben we vaker gezien) en dat wat er rondom dit dossier wordt gepresenteerd verschilt over de tijd. Om een idee te krijgen wat een onderzoeker over de jaren te zien krijgt, kunnen we alle schattingen over de jaren heen ordenen. We krijgen dan **Figuur 46** en zien dat de reviews niet echt verschillen in hun bevindingen terwijl ze wel degelijk andere studies opnemen. Een meer recente review kenmerkt zich dus niet door het opnemen van alle voorheen opgenomen studies plus nieuwe studies. Het is oprecht een divers veld te noemen. Wat ook opvalt is dat sommige reviews helemaal geen significante effecten hebben opgenomen, terwijl anderen een mix hebben.



Figuur 45. Risicoschattingen voor alle studies die hebben gekeken naar de relatie tussen glyfosaat en NHL of CLL/SLL of MM. De kleur geeft aan of een studie statistisch significant is of niet.



Figuur 46. Risicoschattingen met 95% betrouwbaarheidsintervallen per review jaar voor non-hodgkin lymfoom (NHL). De kleur geeft statistische significantie aan.

Nabootsen van één meta-analyse

Het wordt tijd om dieper in de bevindingen van de reviews en hun meta-analyses te duiken. Specifiek genomen zal ik trachten om te laten zien hoe de reviews tot hun *summary estimates* zijn gekomen én of het mij lukt om deze na te bootsen. Voor deze exercities zal ik mij richten tot één meta-analyse en dan specifiek op de meest recente voor NHL. Op die manier krijg ik een indruk van wat mij te wachten staat als ik de rest van de reviews tot mij neem⁷⁰. En van NHL hebben we nu eenmaal de meeste data. Een overzicht van alle reviews die NHL hebben opgenomen staat hieronder:

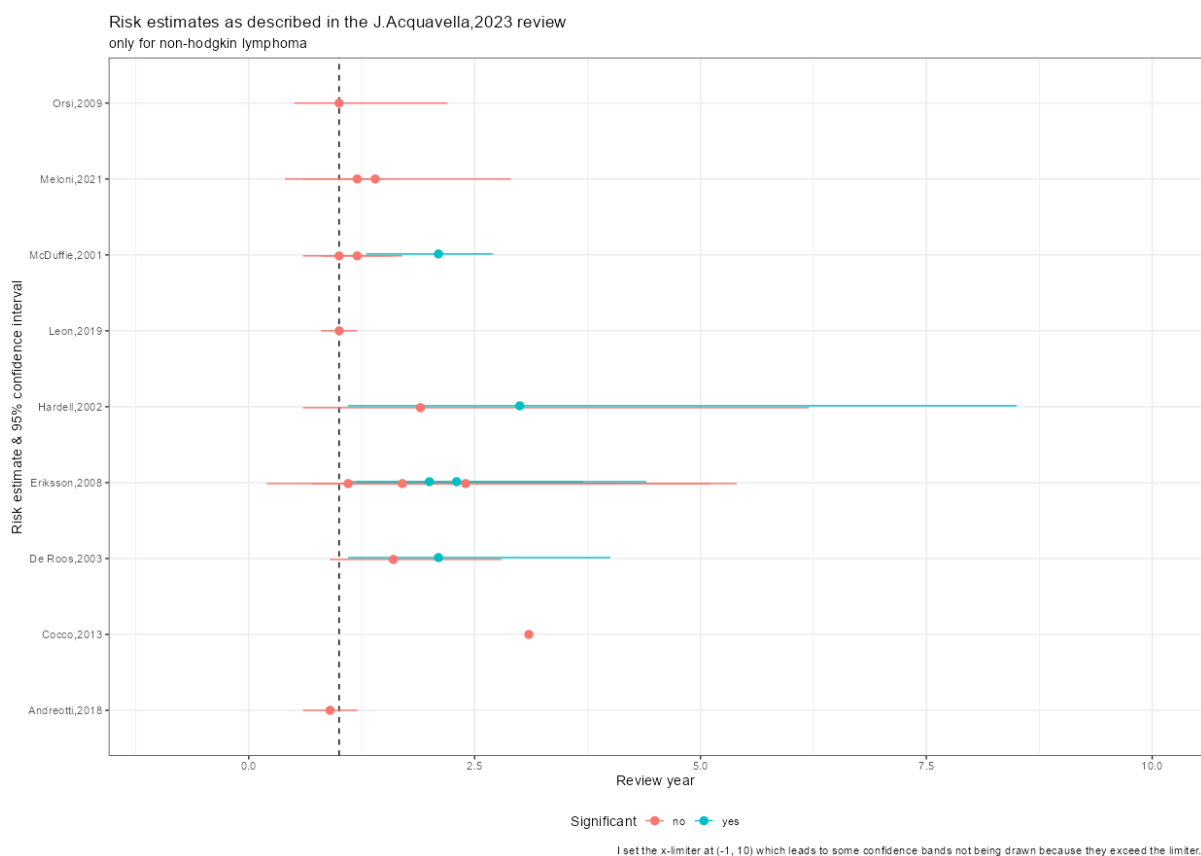
```
# A tibble: 13 x 2
  review_author_year review_year
<chr>              <dbl>
1 J.Acquavella,2023 2023
2 G.C.Kabat,2021    2021
3 D.D.weisenburger,2021 2021
4 P.Boffetta,2021   2021
5 F.Donato,2020     2020
6 L.Zhang,2019      2019
7 M.E.Leon,2019     2019
8 J.E.Goodman,2017  2017
9 E.T.Chang,2016    2016
10 J.Acquavella,2016 2016
11 L.Schinasi,2014   2014
12 P.J.Mink,2012     2012
13 D.D.Alexander,2007 2007
```

⁷⁰ Mijn werk rondom de meta-analyse van Portier liet mij al zien hoe lastig het is om de bevindingen van een ander te repliceren als je niet de codes van die ander hebt. Er is ook maar zoveel wat je kunt doen om te trachten dit wel voor elkaar te krijgen.

Als ik echter kijk naar welke review de meeste informatie voor NHL heeft, dan zie ik dat het beter is om een andere review te nemen:

review_author_year	n
<chr>	<int>
1 L.Zhang,2019	23
2 J.Acquavella,2023	18
3 P.J.Mink,2012	18
4 L.Schinasi,2014	17
5 D.D.weisenburger,2021	12
6 E.T.Chang,2016	11
7 J.Acquavella,2016	11
8 F.Donato,2020	7
9 P.Boffetta,2021	6
10 G.C.Kabat,2021	5
11 D.D.Alexander,2007	4
12 J.E.Goodman,2017	2
13 M.E.Leon,2019	1

Maar omdat de review van Acquavella(35) de op één na grootste review is, ga ik toch eerst met deze review aan de slag. In **Figuur 47** kunnen we alle studies zien en alle risicoschattingen die voor NHL zijn opgenomen in die review. Ook zien we welke schattingen wel of niet significant zijn.



Figuur 47. Risicoschattingen met 95% betrouwbaarheidsinterval per studie opgenomen in de review van Acquavella(35) . Het betreft hier alleen non-hodgkin lymfoom (NHL).

Het wordt nu tijd om in de review zelf te duiken. Als ik de review erbij pak en zoek naar de *summary estimate* voor NHL dan zie ik dat die er niet zijn. Wel lees ik een beschrijving over wat andere meta-analyses gedaan hebben, maar de review van Acquavella(35) is dus eigenlijk geen meta-analyse zo blijkt nu⁷¹. De reden voor het niet uitvoeren van een meta-analyse staat, hoewel niet expliciet zo aangeduid, wellicht hieronder:

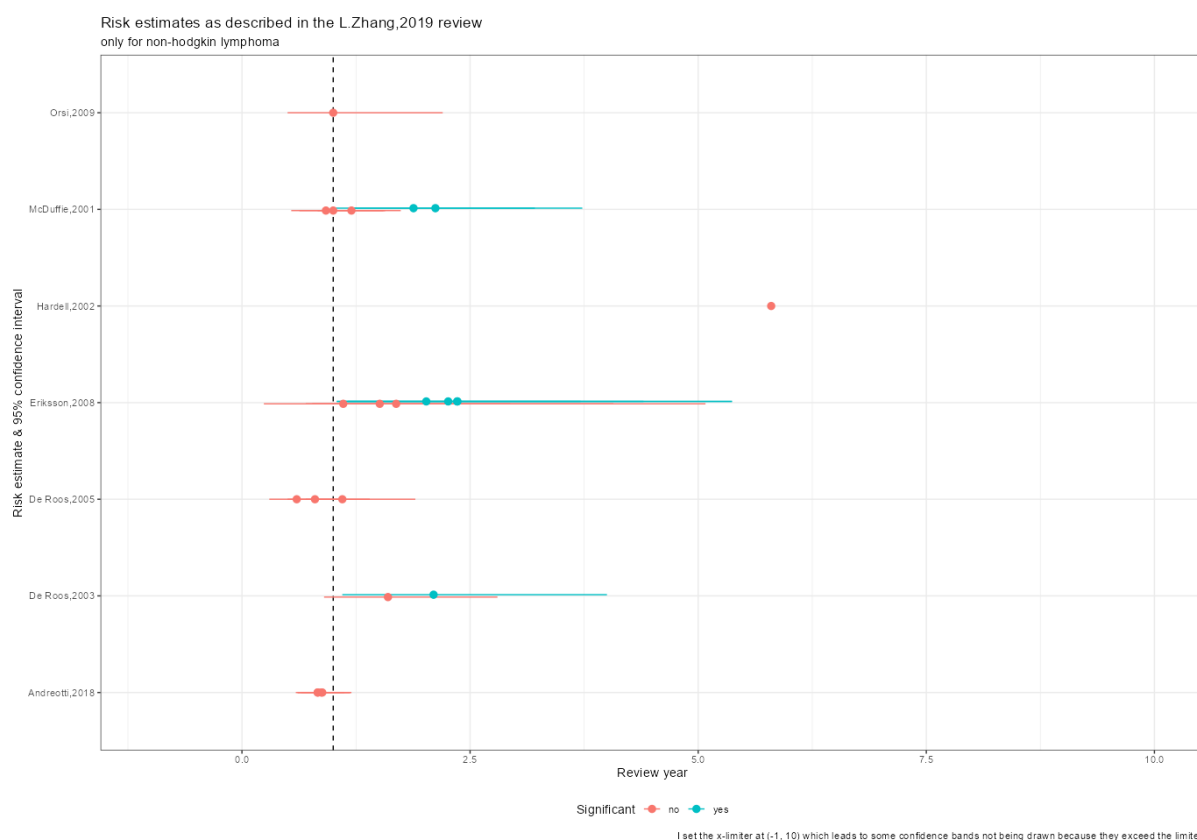
A number of meta-analyses have been published regarding glyphosate and NHL including Schniasi & Leon, Chang & Delzell, Zhang et al., Donato et al., Boffetta et al., and Kabat et al. The different authors used different conventions in their analyses. These meta-analyses are not primary data, but efforts to take a weighted average of the primary data to increase statistical precision and learn something about between study variability. Given the variability across studies in number of days of glyphosate use, possible recall bias in the case control studies, possible selection bias in the case control studies, variable amounts of proxy information in the case control studies, incorrect analytic approaches (defining unexposed as not having exposure to any pesticide), approach to exposure assessment, and little to no control for confounding by other pesticide exposures, these meta-analyses are averaging results of questionable similarity and validity. Fixed and random effects models, as implemented by these authors, cannot compensate for the underlying variability in exposure scenarios and the limitations of many of the primary studies. As Greenland and O'Rourke counsel: averaging results should be limited to results that could be reasonably expected to be similar across studies. Further, they counsel that if systematic variation exists between studies, as it clearly does in this literature, it violates the assumptions of random and fixed effects models.

Dit zijn fikse woorden geplaatst in de laatste review die in dit onderzoeksgebied voorhanden is. We zullen met de review dus niet verder uit de voeten kunnen komen. Toch wil ik een poging wagen om een meta-analyse na te bootsen en wijk dus uit naar de review van Zhang(33) waar ik de volgende studies en resultaten voor NHL vindt (zie **Figuur 48**). Als ik in deze review op zoek ga naar de *summary estimates* dan stuit ik op **Figuur 49**. Los van de schattingen lees ik nog iets anders interessant, namelijk het verschil in resultaat als we de 'nieuwste' data uit het AHS cohort meenemen. Het blijkt dat AHS staat voor [Agricultural Health Study](#), een langlopende cohort studie die, onder andere, kijkt naar de invloed van agrarische activiteiten op gezondheid. De relatie tussen glyfosaat en kanker is maar een klein onderdeel van wat je met deze data kunt doen, en omdat er om de zoveel jaar voldoende

⁷¹ In principe zou deze review dus niet eens opgenomen mogen worden in dit rapport, omdat ik mijn eigen bevindingen nu niet kan vergelijken met de schattingen die in de originele review zouden moeten staan. Toch beslis ik om de review wel te houden. Het feit dat deze review ervoor koos om geen meta-analyse uit te voeren is al een bevinding op zichzelf. Het is ook tekenend voor een veld wat vaak niet verder komt dan het narratief samenvoegen van gegevens.

nieuwe data is verzameld⁷² wordt de relatie ook opnieuw uitgerekend. Zo zien we dat de vakere aangehaalde studie van De Roos uit 2005 (60) ook gebruik maakt van de AHS data. Ook zitten er andere studies in dit rapport die allemaal gebruik maken van de AHS data⁷³ en dat zijn er in totaal acht ((52-54, 60, 63, 65, 75, 86). Dat betekent dat een aardig deel van de geïnccludeerde schattingen afkomstig zijn van één en dezelfde cohort. Sterker nog, van de 291 rijen aan data zijn er 111 afkomstig van de AHS cohort studie: dat is 38%.

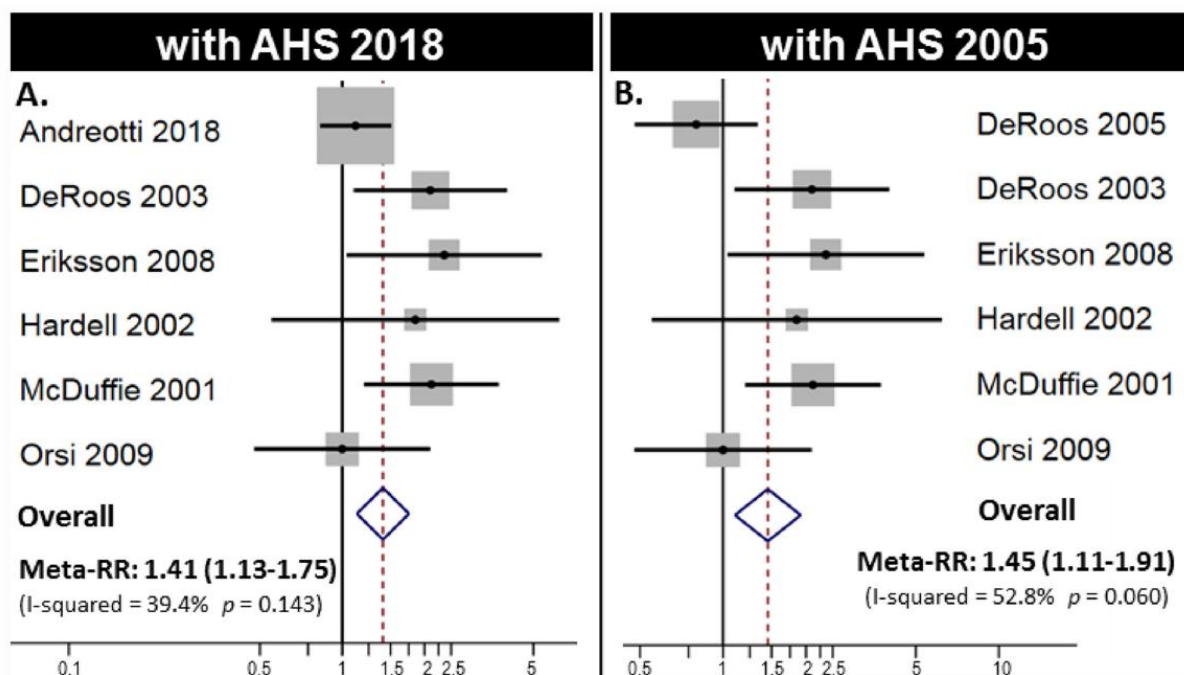
Omdat in het figuur van de Zhang review(33) exact staat beschreven welke studies zijn opgenomen kunnen we beide meta-analyses herhalen door de juiste studies te selecteren.



Figuur 48. Risicoschattingen met 95% betrouwbaarheidsinterval per studie opgenomen in de review van Zhang(33). Het betreft hier alleen non-hodgkin lymfoom (NHL).

⁷² Voldoende in de zin van aantallen, maar niet per se voldoende in de zin dat een relatie met voldoende zekerheid bevestigd of ontkracht kan worden.

⁷³ Voor een compleet overzicht van alle publicaties kun je [hier](#) terecht.



Figuur 49. Resultaten uit Zhang(33) zoals daar gepubliceerd. Het betreft de resultaten van twee meta-analyses: elk met een andere set aan AHS data.

Omdat ik moet werken met gepubliceerde gegevens (inclusief de 95% betrouwbaarheidsintervallen, maar zonder de aantallen) zal ik allereerst van elke meting de puntschatting en de schatting van de variatie moeten bepalen. De puntschatting kan ik makkelijk aflezen, maar de schatting van de variatie moet ik zelf berekenen. Daar zijn [formules](#) voor en als ik die toepas mag ik ervan uit gaan dat ik het resultaat van Zhang(33) óf expliciet herhaal óf minstens heel dicht benader⁷⁴. Ik zal gemakshalve twee verschillende methodes met elkaar gaan vergelijken. De eerste formule staat in het [Cochrane Handbook of Systematic Reviews](#). De ander staat beschreven in de documentatie van het R pakket [metafor](#) en is in uitvoerbare vorm verwerkt in de functie [metafor::conv.wald\(\)](#). In formule vorm zeggen beiden methoden eigenlijk hetzelfde en dat is dat v_i , de variatie van de schatting, als volgt te bepalen is (**Figuur 50**):

$$v_i = \left(\frac{ci.ub - ci.lb}{2 \times z_{crit}} \right)^2$$

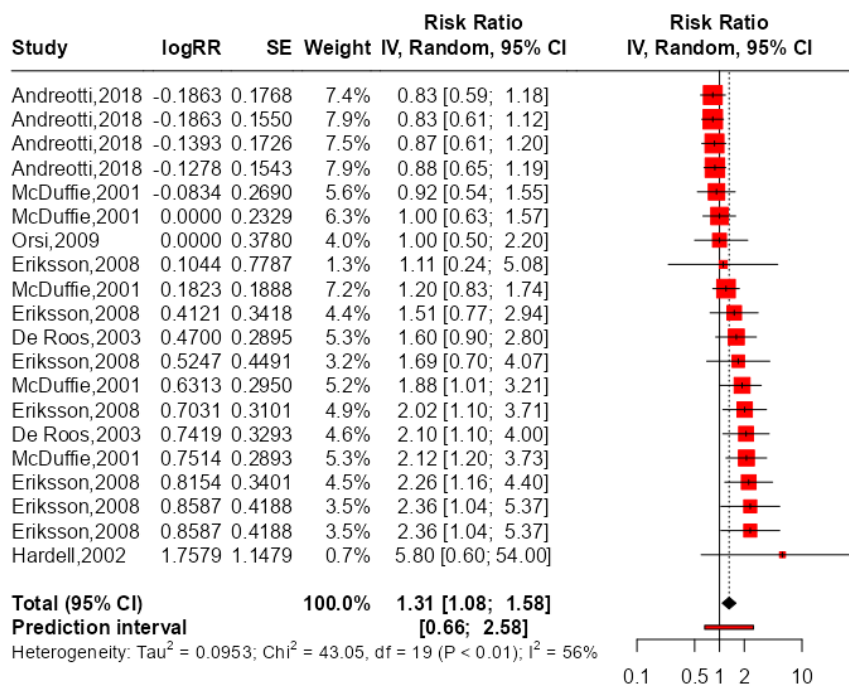
Figuur 50. Berekening van de variatie van een odds-ratio op basis van gerapporteerde betrouwbaarheidsintervallen

⁷⁴ Het kan soms buitengewoon moeilijk zijn om methoden te herhalen op basis van gepubliceerde gegevens. Het is daarom eigenlijk niet meer dan wenselijk dat onderzoekers hun datasets ter beschikking stellen. Helaas gebeurt dit zelden of onder extreme voorwaarden (zoals het niet mogen publiceren met die data).

Omdat ik deze exercitie al met voldoende zekerheid heb uitgevoerd in mijn [tweede rapport](#) zal ik de test hier nu niet herhalen⁷⁵. Hoe dan ook zullen we straks uit de resultaten kunnen afleiden of dat wat ik schat ook exact datgene is wat geschat wordt in de meta-analyse van Zhang(33). Maar voordat ik mijn eigen schattingen naast die van Zhang(33) kan houden wil ik zien of ik wel exact dezelfde studies includeer. Een eerste meta-analyse laat namelijk het volgende resultaat zien (**Figuur 51**) en dat komt helemaal niet overeen met wat er is gerapporteerd. Zo includeer ik veel meer risicoschattingen per studie dan wat Zhang(33) doet. Ergens is daar dus een keuze gemaakt die ik moet volgen om eenzelfde resultaat te behalen. Het is daarom raadzaam om de mate van blootstelling bij elke schatting neer te zetten om zo te zien of daar de reden voor de afwijking ligt. We zien al snel dat we appels met peren aan het vergelijken zijn:

```
# A tibble: 20 x 4
  study_author_year risk_estimate exposure_metric exposure_category
  <chr>             <dbl> <chr>           <chr>
1 Andreotti,2018    0.87 Days of exposure =4340,0
2 Andreotti,2018    0.83 Days of exposure 1-598,9
3 Andreotti,2018    0.88 Days of exposure 1650-4339,9
4 Andreotti,2018    0.83 Days of exposure 599-1649,9
5 De Roos,2003      1.6 Exposed          =1 day/lifetime
6 De Roos,2003      2.1 Exposed          =1 day/lifetime
7 Eriksson,2008     2.36 Days of exposure >10 days
8 Eriksson,2008     2.26 Days of exposure >10 years latency
9 Eriksson,2008     1.69 Days of exposure =10 days
10 Eriksson,2008    1.11 Days of exposure =10 years latency
11 Eriksson,2008    1.51 Exposed          =1 day (multivariate)
12 Eriksson,2008    2.02 Exposed          =1 day (univariate)
13 Eriksson,2008    2.36 Exposed          NA
14 Harde11,2002     5.8 Exposed          NA
15 McDuffie,2001    2.12 Days of exposure >2 days/year
16 McDuffie,2001    1 Days of exposure 0-2 days/year
17 McDuffie,2001    0.92 Exposed          Glyphosate only
18 McDuffie,2001    1.88 Exposed          High exposure
19 McDuffie,2001    1.2 Exposed          NA
20 Orsi,2009        1 Exposed          NA
```

⁷⁵ Ook zal ik niet het probleem herhalen wat ik krijg met het terugrekenen omdat de betrouwbaarheidsintervallen niet symmetrisch zijn. Een uitgebreide uitleg over dat probleem valt te lezen in mijn rapport over risicofactoren voor de ziekte van Parkinson. Hoe dan ook, deze problemen blijven voortkomen zolang studies niet hun gehele dataset publiceren en niet op basis van *individual patient data* meta-analyses kunnen uitvoeren. Dit is één van de oorzaken waarom we überhaupt een replicatiecrisis hebben.



Figuur 51. Meta-analyse voor glyfosaat en non-hodgkin lymfoom (NHL) op basis van de studies van Zhang(33). Ik probeer de eerste meta-analyse uit Zhang(33) na te bootsen door de studie van Andreotti(54) toe te voegen en die van De Roos(61) eruit te laten.

In de review van Zhang(33) lees ik het volgende over de meta-analyse die is uitgevoerd:

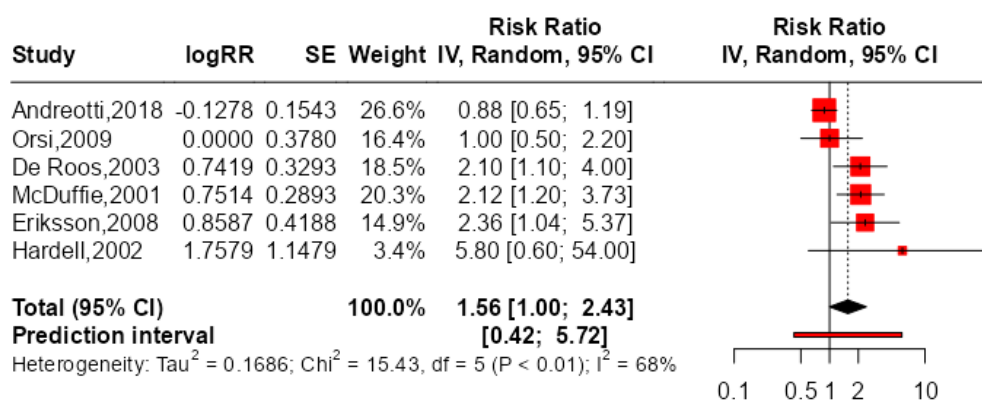
We prioritized highest cumulative exposure based on evidence of glyphosate's persistence in the environment and because chronic disease, including cancer, is usually the result of cumulative exposures. We selected the longest lag or latency because decades may be needed for the health effects of many environmental toxicants to manifest as detectable cancers. If no high exposure data were available, we used the ever-exposure estimate.

Het lijkt er dus op dat ik voor elke studie de hoogste mate van blootstelling moet kiezen indien mogelijk. Omdat we er stiekem van mogen uitgaan dat een langere blootstelling een hogere risicoschatting met zich meebrengt, kies ik nu voor om van elke studie gewoonweg de hoogste schatting te selecteren. Maar als ik dat doe krijg ik twee rijen voor de studie van Eriksson(64):

study_author_year	risk_estimate	exposure_metric	exposure_category
<chr>	<dbl>	<chr>	<chr>
1 Andreotti,2018	0.88	Days of exposure	1650-4339,9
2 De Roos,2003	2.1	Exposed	=1 day/lifetime
3 Eriksson,2008	2.36	Days of exposure	>10 days
4 Eriksson,2008	2.36	Exposed	NA
5 Hardell,2002	5.8	Exposed	NA

6	McDuffie, 2001	2.12	Days of exposure >2 days/year	
7	Orsi, 2009	1	Exposed	NA

Wanneer ik één van die waarden filter, eindig ook ik met zes schattingen. Elke schatting weerspiegelt de hoogste schatting voor het risico op NHL door glyfosaat. De meta-analyse die dan volgt zien we in **Figuur 52**, maar ik krijg niet het getal zoals zichtbaar in **Figuur 49**. Maar de puntschatting blijkt wel overeen te komen met de waarde zoals gepresenteerd in Tabel 5 van de publicatie wanneer ik kijk onder het kopje *random effects*. Dat is methode die ook ik heb toegepast(99). Het lijkt er dus op dat de publicatie van Zhang(33) alleen de *fixed effects* variant publiceert in de tekst en daar valt al wat op aan te merken⁷⁶. Als ik dezelfde methode toepas voor de tweede meta-analyse komt **Tabel 1** er als volgt uit te zien. De tweede door mij berekende waarde (1.64 [1.15; 2.43]) zie ik nu trouwens niet meer terug in **Tabel 5** uit de publicatie. Nu is het zo dat er verschillende manieren zijn om een *random-effects* meta-analyse uit te voeren en die keuzes zijn van invloed op het resultaat. Toch laat ik het verschil nu even voor wat het is, omdat de bevindingen ook weer niet zover uit elkaar liggen. Zo is het goed mogelijk dat mijn keuze om voor elke studie het grootste getal te kiezen niet is wat Zhang(33) heeft gedaan, maar wat Zhang exact heeft gedaan is niet zo makkelijk te achterhalen.



Figuur 52. Meta-analyse op basis van de review van Zhang, 2019. Het was de bedoeling om afbeelding A uit **Figuur 49** te reproduceren, maar dat is niet helemaal gelukt.

Wat wel duidelijk is, is dat de nieuwste berekening van Zhang(33) (de bovenste rij) twee keer een significant effect vindt, terwijl ik er één vind. Verder zijn alle gevonden *odds-ratio's* klein

⁷⁶ Deze manier van data samenvoegen kan eigenlijk alleen als elke studie een exacte replica is van een andere studie (of de andere studies). Wat betekent dat als er al verschillen zijn tussen de bevindingen in studie A en studie B deze voortkomen uit het nemen van een andere steekproef en daarmee geen 'echt' verschil vertegenwoordigen. Een bijzonder zware assumptie om te maken in observationeel onderzoek. Daarom past men deze *fixed effects* modellen eigenlijk nooit toe, maar wijkt men altijd uit naar *random effects* modellen.

te noemen. Ook kunnen er verschillen optreden door het specifieke statistiekpakket dat iemand hanteert en de keuzes die iemand maakt⁷⁷. Al deze facetten, en meer, maakt het lastig om de bevinding van een meta-analyse exact na te bootsen⁷⁸.

Review	Summary estimate gerapporteerd (fixed)	Summary estimate gerapporteerd (random)	Summary estimate berekend (random)
Zhang, 2019 (AHS 2018)	1.41 (1.13; 1.75)	1.56 (1.12; 2.16)	1.56 (1.00; 2.43)
Zhang, 2019 (AHS 2005)	1.45 (1.11; 1.91)	1.52 (1.00; 2.31)	1.64 (1.15; 2.43)

Tabel 1. Summary estimates voor Zhang(33) zoals gerapporteerd in de review én zoals berekend door mij.

Invloed van correctiefactoren

We hebben tot nu toe gekeken naar de reviews, de studies, wat de eigenschappen zijn van zowel de reviews als de studies, hoe die metingen eruit zien, de onzekerheden in een studie, en de onzekerheden tussen studies en tussen reviews (waaronder onzekerheden tussen dezelfde studies tussen reviews). We hebben gezien voor welke kankersoort welke reviews en studies beschikbaar zijn en hoe schattingen over tijd veranderen. Ook hebben we gekeken naar de mate van blootstelling en de specifieke designs die zijn gebruikt. De focus lag daarbij op de ratio's zoals de *odds-ratio* en de *risk-ratio*, waarbij ik wederom de serieuze kanttekening wil maken dat beide statistieken niet inwisselbaar zijn bij een ziekte zoals kanker⁷⁹. Als laatste heb ik geprobeerd om de gepubliceerde resultaten van een meta-analyse na te bootsen én ontdekt hoe lastig het is om exact tot hetzelfde getal te komen. Eigenlijk is het haast onmogelijk om, zonder codes, te weten wat iemand gedaan heeft.

Waar we nog niet naar hebben gekeken is of er verschil is in schattingen als er wordt gecorrigeerd voor bepaalde factoren (zogenaamde *confounders*). Het is goed om te weten dat bij *case-control* studies er bijna automatisch gecontroleerd wordt voor factoren zoals leeftijd en geslacht, omdat het *study design* vraagt om een specifieke vorm van *matching*. Bij *cohort* studies is dat niet het geval en zal men in het ontwerp van de studie, maar ook in de statistische analyse, expliciet moeten controleren voor variabelen die ook van invloed kunnen zijn op de relatie tussen glyfosaat en kanker. Een *confounder* is een variabele die van

⁷⁷ Ook dit gedeelte heb ik uitvoerig besproken in mijn [rapport](#) over risicofactoren voor de ziekte van Parkinson en zal ik daarom hier niet meer adresseren.

⁷⁸ Dat is een bevinding op zich, en benadrukt de noodzaak om review overstijgend te werk te gaan. Het is ook een bevestiging van de noodzaak om data en codes te delen met anderen onderzoekers. Dit gebeurt helaas zelden.

⁷⁹ Deze is niet zeldzaam genoeg om van een *odds-ratio* een *risk-ratio* te maken en omgekeerd.

invloed is op zowel glyfosaat én kanker en op die manier een gevonden relatie tussen glyfosaat en kanker kan verklaren. Corrigeren voor risicofactoren is één van verschillende manieren om via statistiek een signaal tussen ruis te vinden, maar het is verre van perfect en de eerdergenoemde matching in case-control studies is eerder een 'moetje'.

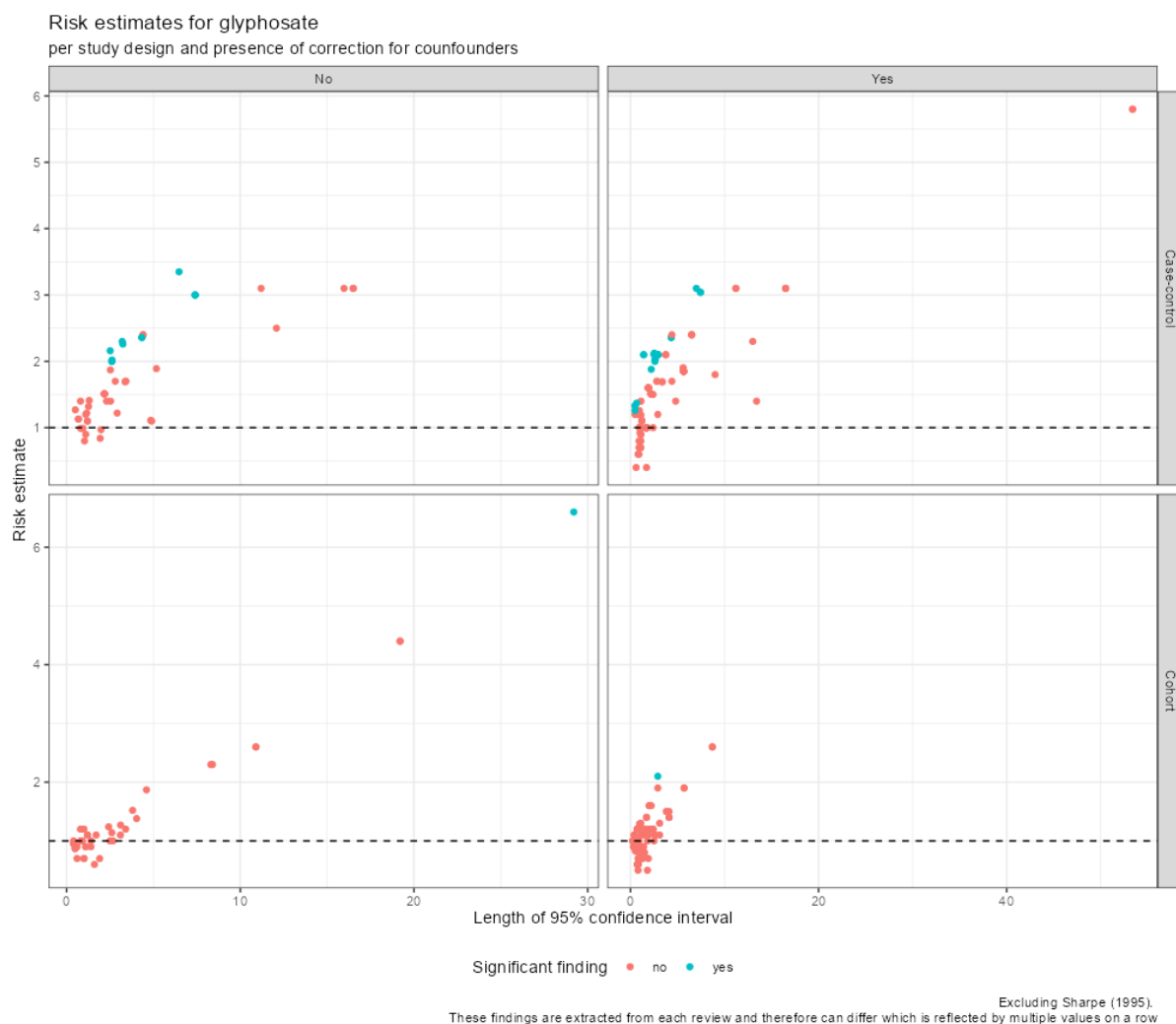
Toch wil ik heel graag zien wat er over de reviews heen zoal gedaan is. Ik heb eerder al beschreven dat ik heb gekozen om geen grote groepen te maken en dat het voor mij belangrijker is om de verschillen tussen studies (en hun schattingen) te tonen dan de overeenkomsten. Als we kijken naar de aan-of afwezigheid van correctiefactoren per *study design* dan zien we de volgende tabel en grafiek (**Figuur 53**):

	study_design	study_adjustments_binary	n
	<chr>	<chr>	<int>
1	Case-control	No	50
2	Case-control	Yes	113
3	Cohort	No	40
4	Cohort	Yes	72

Ik zie niet direct een relatie tussen het wel of niet meenemen van correctiefactoren, maar ik zie wel een sterk verschil per *study design*⁸⁰. Wellicht helpt het als ik de data iets anders weergeef en per kankersoort laat zien of er een verschil is tussen het wel of niet meenemen van correctiefactoren. Het resultaat zien we in **Figuur 54**, maar ook hier zie ik geen grote verschillen. Deze bevinding weerspiegelt mijn eerdere rapporten waarin ik kon putten uit veel meer data. Eigenlijk is dat ook niet gek, want de aan-of afwezigheid van een correctiefactor betekent nog niet dat een bepaalde schatting ook consistent een bepaalde richting op beweegt. Die analyses moet je in de studies zelf doen, maar dat is onmogelijk met data op het niveau van de publicatie. Ook nu blijkt hoe belangrijk het is om te kunnen rekenen met de originele data zelf.

⁸⁰ Nu moet natuurlijk gezegd worden dat ik geen direct vergelijk maak IN een studie, maar alleen tussen studies. Ik weet dus niet hoe een resultaat verandert als er op het niveau van de studie zelf wordt gecorrigeerd. Wat mij rest is alleen maar te zien of het rapporteren van aangepaste berekeningen, gecorrigeerd voor *confounders*, zorgt voor minder statistische significante bevindingen. Deze aanpak kan het beeld vertekenen en men kan niet spreken over causaliteit.

Wat mij wel intrigeert is de grote afwezigheid van correcties voor cohort onderzoek bij beenmergkanker (MM).



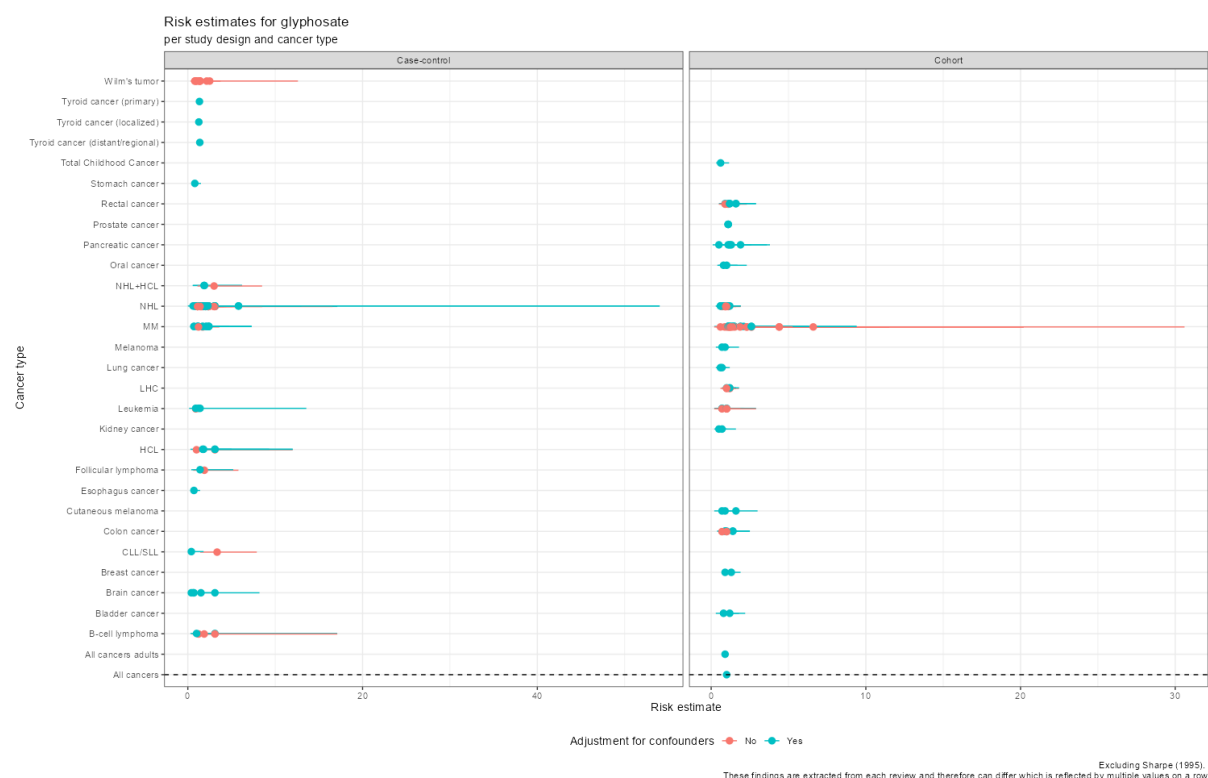
Figuur 53. Risicoschattingen en de lengte van het 95% betrouwbaarheidsinterval per *study design* en de aan-of afwezigheid van correctiefactoren. De kleur geeft statistische significantie aan.

Een snelle splitsing toont de volgende tabel:

study_design	study_adjustments_binary	n
<chr>	<chr>	<int>
1 Case-control	No	2
2 Case-control	Yes	13
3 Cohort	No	18
4 Cohort	Yes	16

Inderdaad lijkt het erop dat er meer correcties zijn gedaan bij *case-control* studies dan bij cohort studies, maar wellicht wordt dit beeld vertekent door één of twee studies. Ik wil daarom alleen naar de cohort studies kijken om vervolgens te zien waar en wanneer er wel of geen correctiefactoren zijn meegenomen. Ik krijg dan de volgende tabel:

study_adjustments_binary	study_author_year	n
<chr>	<chr>	<int>
1 No	De Roos, 2005	10
2 No	Leon, 2019	1
3 No	Sorahan, 2015	7
4 Yes	De Roos, 2005	11
5 Yes	Presutti, 2016	1
6 Yes	Sorahan, 2015	4



Figuur 54. Risicoschatting met 95% betrouwbaarheidsinterval per kankersoort, *study design* en de aan-of-afwezigheid voor correctiefactoren.

Dit is een bijzondere bevinding, omdat voor beide groepen (wel óf geen correctiefactoren) er één enkele studie uitspringt: De Roos(60). Ik ben nu benieuwd welke reviews erbij betrokken zijn en krijg dan de volgende tabel:

# Groups:	study_adjustments_binary	review_author_year	n
	<chr>	<chr>	<int>
1	No	E.T.Chang, 2016	7
2	No	J.V.Varghese, 2021	3
3	Yes	E.T.Chang, 2016	1
4	Yes	J.Acquavella, 2016	8
5	Yes	P.J.Mink, 2012	2

Het verschil lijkt dus door de reviews te komen. Daarom wil ik nu voor alle risicoschattingen zien voor hoeveel exact dezelfde schattingen⁸¹ ik wel of geen informatie heb over de mate van correctie:

review_author_year	risk_estimate	risk_estimate_low	risk_estimate_high	study_adjustments_binary
<chr>	<dbl>	<dbl>	<dbl>	<dbl> <chr>
1 E. T. Chang, 2016	0.6	0.2	1.8	No
2 J. Acquaveilla, 2016	1.1	0.4	3	Yes
3 J. Acquaveilla, 2016	1.1	0.4	3.5	Yes
4 E. T. Chang, 2016	1.1	0.4	3.5	No
5 J. Acquaveilla, 2016	1.1	0.5	2.4	Yes
6 E. T. Chang, 2016	1.2	0.4	3.8	No
7 J. Acquaveilla, 2016	1.2	0.5	2.9	Yes
8 J. Acquaveilla, 2016	1.4	0.4	4.5	Yes
9 J. Acquaveilla, 2016	1.5	0.2	4.3	Yes
10 P. J. Mink, 2012	1.9	0.6	6.3	Yes
11 J. Acquaveilla, 2016	1.9	0.6	6.3	Yes
12 P. J. Mink, 2012	2.1	1.1	4	Yes
13 J. V. Varghese, 2021	2.3	0.6	8.9	No
14 E. T. Chang, 2016	2.3	0.6	9	No
15 J. V. Varghese, 2021	2.6	0.6	11.5	No
16 E. T. Chang, 2016	2.6	0.6	11.5	No
17 J. Acquaveilla, 2016	2.6	0.7	9.4	Yes
18 E. T. Chang, 2016	2.6	0.7	9.4	Yes
19 J. V. Varghese, 2021	4.4	1	20.2	No
20 E. T. Chang, 2016	4.4	1	20.2	No
21 E. T. Chang, 2016	6.6	1.4	30.6	No

De tabel hierboven verraad dat ik soms voor exact dezelfde schattingen wel of geen informatie kreeg vanuit de reviews over de manier waarop er met correctiefactoren is omgegaan. Soms had ik helemaal geen informatie, ook niet vanuit een andere review⁸². Deze bevinding maakt dat het filteren van de data op basis van de gehanteerde correctiefactoren lastig is. Het is daarom wellicht beter om meer vanuit het *study design* te redeneren of de mate van blootstelling. Hoe dan ook wordt het nu tijd om te duiken in mogelijke meta-analyses voor glyfosaat en kanker.

Meta-analyses voor glyfosaat en kanker

Het is nu tijd om zelf met de data te gaan rekenen⁸³ wat betekent dat we studies gaan samenvoegen om tot gewogen risicoschattingen te komen. De meest voor de hand liggende meta-analyse om uit te voeren is het simpelweg samenvoegen van alle studies die naar glyfosaat en kanker hebben gekeken. Dit betekent ook dat we sommige studies dubbel gaan tellen, omdat de database is opgebouwd vanuit de verschillende reviews. In principe weet je al dat dit geen correcte manier van samenvoegen is, al is het maar omdat de studies niet

⁸¹ Bepaald door de puntschatting en het 95% betrouwbaarheidsinterval.

⁸² Nu had ik ervoor kunnen kiezen om, net als bij het invullen van de van missende waarden voor de risicoschattingen, informatie van andere reviews te lenen. Maar dat had niet voor elke schatting gekund. Ook is het invullen van missende waarden per schatting gemakkelijker, omdat ik de mate van blootstelling kan gebruiken om te bepalen of we het over dezelfde schatting hebben. Uiteindelijk zou je kunnen redeneren dat ik met die informatie ook de missende informatie over correctiefactoren had kunnen invullen. Wat overblijft is dus een keuze van mijn kant om dat niet te doen. Met het ter beschikking stellen van de data en de codes kan iedere geïnteresseerde lezer zelf aan de slag om te zien wat het effect zou zijn van het maken van andere keuzes.

⁸³ Een klein voorschot op wat ik wil gaan doen heb ik natuurlijk al laten zien in het trachten te reproduceren van de gepubliceerde schattingen van Zhang, 2019. Ook heb ik al aardig wat beschrijvende statistiek gebruikt.

allemaal dezelfde berekeningen rapporteren. Zo kunnen we onmogelijk⁸⁴ een *odds-ratio* met een *risk-ratio* combineren. Ook is het niet slim om schattingen voor verschillende soorten kanker te combineren omdat niet elke vorm van kanker op eenzelfde manier tot stand komt. En dan hebben we nog het probleem dat de verschillende *study designs* ook van invloed kunnen zijn op de gepresenteerde schatting⁸⁵. Zo begint een *case-control* met de ziekte en gaat dan op zoek naar risicofactoren, terwijl een *cohort* begint met mogelijke risicofactoren om vervolgens te zoeken naar ziektes. Dit zijn twee compleet verschillende manieren van informatie vergaren die een wezenlijke invloed hebben op de schattingen zelf. Ten slotte hebben we nog de mate waarin wel of niet gecorrigeerd is voor mogelijke beïnvloedende factoren. Omdat het niet lijkt dat deze factoren van invloed zijn zal ik hier niet verder op filteren. Wellicht dat de lezer nu denkt dat het weinig zinvol is om met de data te gaan rekenen gegeven het feit dat de reviews verschillen in hun rapportages van de studies, maar het is toch mijn taak om in dit rapport te beschrijven hoe ver ik kan komen. De mate waarin een meta-analyse mogelijk is, is vaak ook de essentie van die meta-analyse⁸⁶.

We zullen dus een aantal keuzes moeten maken om de meta-analyse (statistisch) verantwoord te houden. Om te beginnen zal ik dubbele schattingen verwijderen zodat wat overblijft schattingen zijn met informatie in de meeste kolommen. Op die manier tel ik niet dubbel en zorg ik ervoor dat ik het verst met de data kan duiken (althans dat hoop ik). Verder zal ik alleen studies meenemen die *odds-ratio's* of *risk-ratio's* rapporteren. We hebben gezien dat, hoewel we ook *hazard ratio's* hebben, deze maar een klein deel van het totaal uitmaken. Verder is een *hazard ratio* een hele ander statistiek dan de *odds-ratio* of de *risk-ratio*. Combineren is hier dus eigenlijk onmogelijk. Verder laat ik de studie van Sharpe(85) uit de analyse omdat deze bevinding zo extreem is dat er voor mijn gevoel iets niet klopt. En als het wel 'kopt' dan is de onzekerheid zo groot dat de puntschatting eigenlijk weinig waardevolle informatie oplevert: we weten namelijk helemaal niet zo goed welke waarde het beste past bij de schatting.

Deze keuzes tezamen maken dat ik in totaal 219 (75.2%) van de 291 rijen overhoudt, maar een vlugge blik toont dat ik nog steeds dubbele risicoschattingen heb. Dit komt omdat

⁸⁴ Onmogelijk is een groot woord, maar wat ik bedoel is dat het samenvoegen problemen oplevert die je gewoonweg kunt voorkomen door niet te combineren.

⁸⁵ En niet alleen omdat een ander *study design* vaak ook een andere statistiek betekent. We hebben namelijk al gezien dat er verschillende statistieken zijn per *study design*. Verder is het ook zo dat de ene *case-control* of cohort studies niet de andere is.

⁸⁶ Toch blijft het van belang om te blijven benoemen hoe de uitkomst van een meta-analyse grotendeels bepaald wordt door de keuzes die men maakt in het wel of niet samenvoegen van studies.

sommige reviews een risicoschatting voor een bepaalde studie als *odds-ratio* of *risk-ratio* hebben gepresenteerd, of de mate van blootstelling anders hebben gecategoriseerd⁸⁷. Nu heb ik al eerder gezegd dat ik niet bereid ben om groepen samen te voeren ten behoeve van versimpeling tenzij ik 100% zeker weet wat het getal moet zijn. Ik weet dat nu niet en ik zal later op die implicaties (moeten) terugkomen. Maar nogmaals: deze 219 rijen aan data zijn geen 219 studies noch zijn het 219 unieke risicoschattingen. Het zijn 219 rijen aan informatie gesprokkeld op basis van 20 reviews. Het feit dat ik niet kan bepalen welke waarde ik moet meenemen bekrachtigt de eerder gepresenteerde uitspraak uit Zhang(33). We zullen dus voorzichtig moeten zijn in het waarderen van elke cijfer dat uit de volgende analyses komt rollen. Dit betekent ook dat we toch een poging gaan wagen, want kijken hoe ver we kunnen komen én hoeveel waarde die informatie met zich meebrengt is het gehele doel van deze exercitie.

Om te beginnen voeren we een simpele *random-effects* meta-analyse uit door gebruik te maken van zowel het *metafor* als het *meta* pakket (**Tabel 2**). Elk pakket voegt uiteindelijk 219 rijen samen en toont vergelijkbare resultaten met uitzondering van de I^2 ⁸⁸:

Statistiek	Methode berekening	Metafor pakket	Meta pakket
Summary estimate	REML	1.04 (0.99; 1.09)	1.04 (0.99; 1.09)
Tau ²	REML	0.02	0.02
I^2	REML	22.62%	8.3%
H ²	REML	1.29%	1.04%

Tabel 2. De schattingen rondom de heterogeniteit van de meta-analyse zoals gerapporteerd door twee statistische pakketten voor het statistiek programma R: *metafor* en *meta*.

Deze tabel geeft vier verschillende onderdelen aan(99):

1. **Summary estimate**: dit is de *odds-ratio* plus 95% betrouwbaarheidsinterval van alle studies bij elkaar. Het is een gewogen gemiddelde waarbij het gewicht van elke regel bepaald wordt door de variantie van de schatting.
2. **Tau²**: dit is de geschatte hoeveelheid heterogeniteit tussen de studies. Deze waarde geeft aan hoeveel variatie er zit tussen de bevindingen (in dit geval de *odds-ratio's*).

⁸⁷ Ook dit is een bijzondere bevinding, omdat de schattingen niet zomaar inwisselbaar zijn. In de publicaties zelf lees ik geen reden voor dit verschil, maar het kan heel goed zijn dat ze zelf met de data aan de slag zijn gegaan.

⁸⁸ Nogmaals, dit zijn geen 219 unieke risicoschattingen.

Daar waar I^2 de relatieve heterogeniteit weergeeft, geeft Tau^2 de absolute waarde weer. Hoe dichter de waarde tegen de nul aanzit hoe kleiner die variatie is.

3. I^2 : dit is de totale heterogeniteit gedeeld door de totale variabiliteit. De waarde wordt uitgedrukt in een percentage tussen de 0% en 100%. Hoewel de interpretatie vrij los is, worden er wel adviezen gegeven over wanneer een meta-analyse te veel heterogeniteit laat zien. Zo wordt in het algemeen aangenomen dat een waarde van boven de 75% synoniem staat voor 'te veel' [heterogeniteit](#)⁸⁹.
4. H^2 : dit is de totale variabiliteit tussen studies gedeeld door de steekproefvariabiliteit. Deze waarde helpt om de variatie tussen studies af te zetten tegen de variatie in een studie.

Wat opvalt is dat de *summary estimate* exact hetzelfde is, net als Tau^2 . Maar I^2 verschilt in grootte en H^2 worden anders weergegeven (absoluut in plaats van relatief). Dat beide pakketten I^2 anders berekenen heb ik uitgebreid besproken in mijn [tweede rapport](#). Dat zal ik daarom hier niet herhalen.

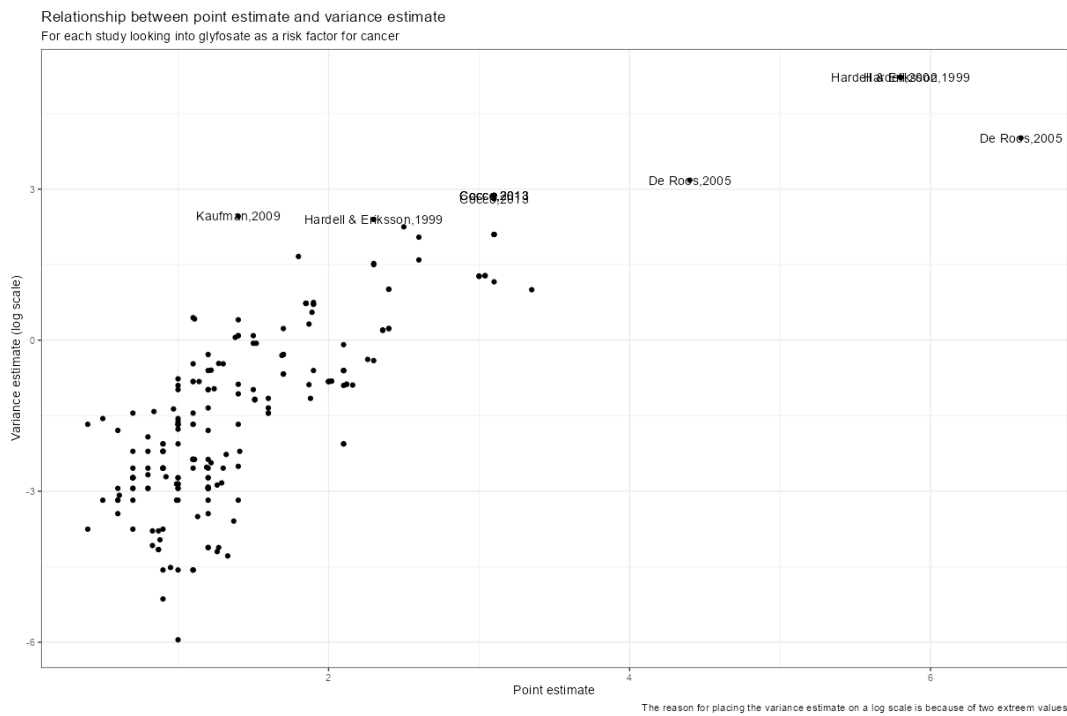
Wat wel zinvol is om te herhalen is het duiden van de relatie tussen de puntschatting en de variatie rondom de puntschatting (**Figuur 55**). Wat we zien is dat twee studies er behoorlijk uitspringen en dan zijn er nog een aantal andere studies met een relatief grootte variatie in de schattingen. Dat is iets om mee te nemen in de beoordeling. Verder zien we, wederom, dat deze dataset niet allemaal unieke schattingen toont⁹⁰. Om dit te duiden kunnen we trachten om de data op te delen. Eerst per *study design* (**Figuur 56**), en dan per kankersoort (**Figuur 57**). We kunnen, wederom, zien dat de effecten voor *case-control* studies groter zijn voor cohort studies. Ook zien we, en dat zagen we in **Figuur 55** al, dat hoe groter de risicoschatting is hoe groter de onzekerheid van die schatting wordt⁹¹. Als we kijken naar **Figuur 57** dan zien we geen eenduidig beeld per kankersoort, maar wel dat het gros van de risicoschattingen voor NHL is. Tezamen laten deze figuren dus eigenlijk niks nieuws zien en ze zijn eigenlijk ook meer bedoeld als inleiding voor de keuzes die ik straks moet maken. Zo is het niet raadzaam om kankersoorten te groeperen noch om statistieken

⁸⁹ "0% to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; 75% to 100%: considerable heterogeneity".

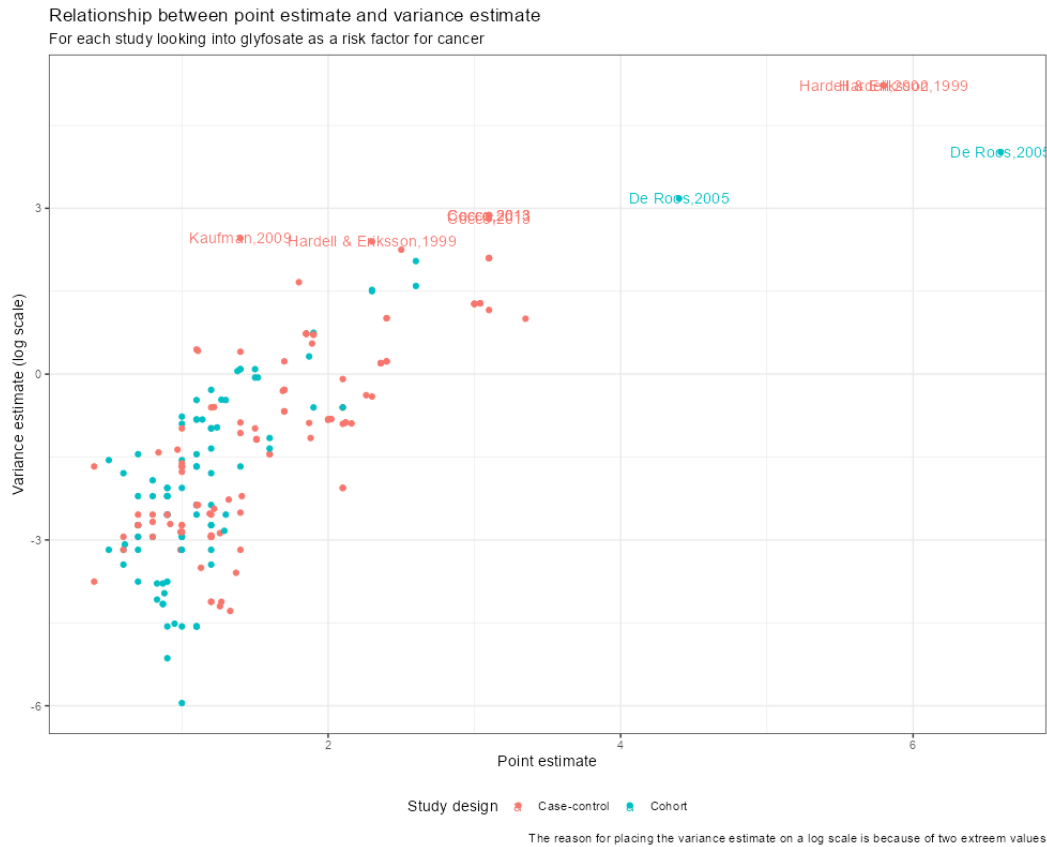
⁹⁰ Tekst overlapt wat aan mij toont dat we te maken hebben met dezelfde puntschatting en schatting van de onzekerheid.

⁹¹ Maar deze relatie lijkt een bepaalde curve te tonen. Geen idee of dit een artefact is of een noemenswaardige relatie.

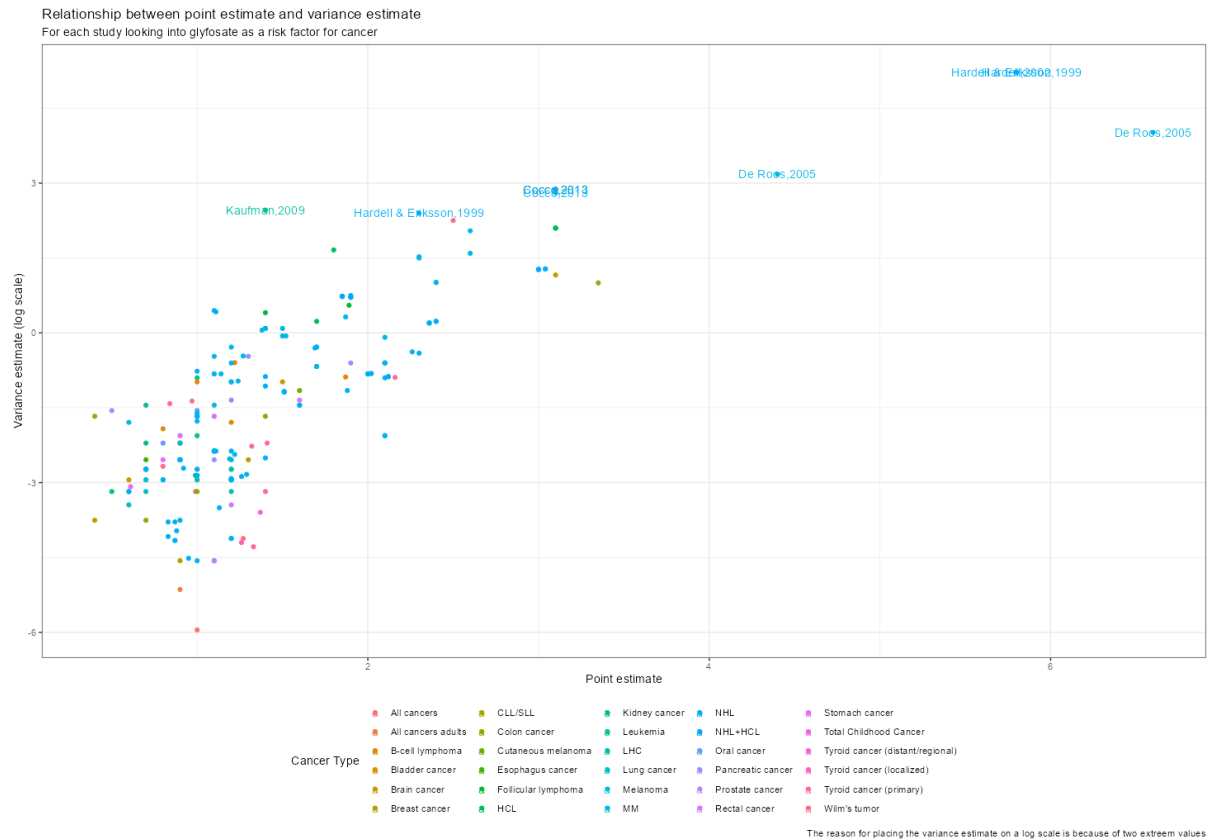
samen te voegen. Daarom is eigenlijk tijd om de analyses per kankersoort en per gebruikte statistiek te presenteren. Ik zal beginnen met de *odds-ratio's*.



Figuur 55. De relatie tussen de puntschatting van het effect van glyfosaat op kanker én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in *odds-ratio's* of *risk ratio's*. Twee studies (60, 67) springen eruit met een groter dan normale variatie.



Figuur 56. De relatie tussen de puntschatting van het effect van glyfosaat op kanker én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in *odds-ratio's* of *risk ratio's*. De kleur geeft het *study design* aan.



Figuur 57. De relatie tussen de puntschatting van een effect van pesticiden én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in *odds-ratio's* of *risk ratio's*. De kleur geeft de kankersoort aan.

Meta-analyses voor glyfosaat – odds-ratio's

Laten we dus eerst maar eens beginnen met een meta-analyse voor *odds-ratio's*. Als ik de rijen selecteer kom ik uit op 84 schattingen (ongeveer 38% van de 219 die we hebben gebruikt in de vorige analyse). Dit zijn dus 84 schattingen over alle kankersoorten heen, maar het lijken geen unieke schattingen (zie **Figuur 59**). Toch heb ik hier om gevraagd, maar als ik naar de cijfers kijk dan zie ik vijf dubbeltellingen⁹². Ik zie geen dubbeltellingen per review⁹³. Toch ga ik door omdat ik wil zien wat de uitkomst ons kan bieden en omdat ik weet dat het gewicht dat elke schatting krijgt best klein is. Het resultaat staat hieronder:

```

Number of studies combined: k = 84

          95%-CI      z  p-value
Random effects model 1.1083 [1.0190; 1.1976] 24.33 < 0.0001
Prediction interval  [0.6671; 1.5495]

Quantifying heterogeneity:
tau^2 = 0.0471 [0.0055; 0.0845]; tau = 0.2171 [0.0740; 0.2907]
I^2 = 32.2% [11.0%; 48.3%]; H = 1.21 [1.06; 1.39]
    
```

⁹² Één voor Cocco; twee voor Eriksson; één voor Hardell, en één voor McDuffie.

⁹³ Het is goed om te blijven onthouden dat sommige studies gebruik maken van eenzelfde cohort en dat wellicht daar het antwoord ligt op de vraag waarom verschillende schattingen en hun 95% betrouwbaarheidsintervallen zo dicht bij elkaar in de buurt liggen.

Test of heterogeneity:

Q	d.f.	p-value
122.43	83	0.0032

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for τ^2
- Q-Profile method for confidence interval of τ^2 and τ
- Hartung-Knapp prediction interval (df = 82) (with ad hoc correction)

Hoewel we met behoorlijk wat voorzichtigheid naar deze studies moeten kijken kunnen we zien dat de *summary estimate* een positief effect laat zien: 1.10 [1.0190;1.1976]. Maar dit effect is erg klein en het 95% voorspellingsinterval is behoorlijk groot (en niet significant). Dit sterkt mij in de gedachte dat we hier te maken hebben met een hoogst onzekere schatting. Om dieper te kijken naar wat de meta-analyse ons kan bieden is het wellicht behulpzaam om de schattingen af te zetten tegen het gewicht dat ze in de meta-analyse hebben gekregen. Dit helpt ons ook om de impact van de vijf dubbeltellingen te plaatsen. Het resultaat is zichtbaar in **Figuur 60** en **Figuur 61**. We kunnen nu duidelijk zien dat de 95% betrouwbaarheidsintervallen van veel schattingen behoorlijk groot zijn, wat verraadt dat het hier om zeldzame gebeurtenissen gaat. Ook zien we dat het gewicht in de meta-analyse direct gelieerd is aan de lengte van het onzekerheidsinterval.

We zouden deze exercitie kunnen herhalen voor de *risk-ratio's* en wie daar interesse in heeft kan zich behelpen met onderstaand resultaat⁹⁴ en **Figuur 58**:

```
Number of studies combined: k = 219
          95%-CI      z p-value
Random effects model 1.0406 [0.9938; 1.0874] 43.54      0
Prediction interval  [0.7602; 1.3210]

Quantifying heterogeneity:
tau^2 = 0.0197 [0.0000; 0.0208]; tau = 0.1403 [0.0000; 0.1442]
I^2 = 8.3% [0.0%; 23.3%]; H = 1.04 [1.00; 1.14]

Test of heterogeneity:
      Q d.f. p-value
237.65 218 0.1721

Results for subgroups (random effects model):
      k          95%-CI      tau^2      tau      Q      I^2
risk_metric = OR  84 1.1083 [1.0190; 1.1976] 0.0471 0.2170 122.43 32.2%
risk_metric = RR 135 0.9871 [0.9460; 1.0283] 0      0      107.19 0.0%

Test for subgroup differences (random effects model):
      Q d.f. p-value
Between groups 5.84 1 0.0157
```

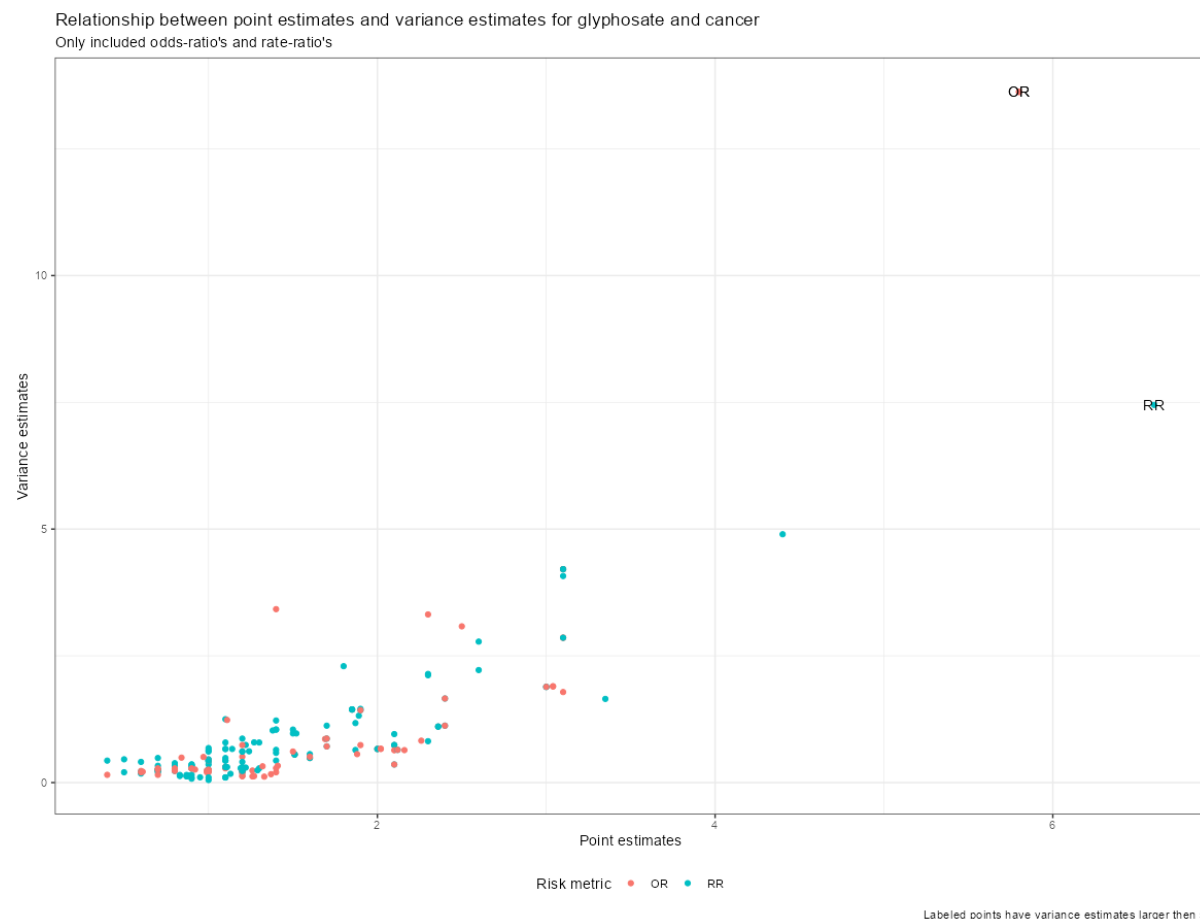
⁹⁴ De oplettende lezer zal zien dat het model niet significant is, maar wel als significant wordt aangemerkt door het model. Dat komt omdat de waarde waartegen wordt afgezet niet de 1 is, maar de 0. Ik heb namelijk niet gespecificeerd dat het om *odds-ratio's* of *risk-ratio's* gaat. Dit maakt dat de schattingen van het summary effect niet helemaal correct is en eigenlijk zouden we de exercitie moeten splitsen per statistiek.

Prediction interval for subgroups:

95%-PI
 risk_metric = OR [0.6671; 1.5495]
 risk_metric = RR [0.9456; 1.0287]

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for τ^2
- Q-Profile method for confidence interval of τ^2 and τ
- Prediction interval based on t-distribution (df = 217)



Figuur 58. Schattingen en onzekerheden voor elke *odds-ratio* en *risk-ratio* die in de reviews is opgenomen.

Maar wellicht is het beter om wat dieper stil te staan bij de heterogeniteit die ook uit de meta-analyse rolt. Vaak wordt er namelijk alleen maar gekeken naar wat het gewogen gemiddelde uit een meta-analyse ons te bieden heeft, maar niet naar wat de verschillen tussen schattingen (vanuit studies, reviews) ons kunnen vertellen.

Extra analyses voor de heterogeniteit

Om eerlijk te zijn is er door de jaren heen een complete bibliotheek aan mogelijkheden ter beschikking gekomen om te ontdekken of het resultaat van een meta-analyse betrouwbaar genoeg is of niet. Zo weten we uit de laatste analyse dat er toch wat heterogeniteit tussen de studies te vinden is, ondanks dat we vijf dubbeltellingen over vier studies (59, 64, 66, 76) heen hebben. Hieronder staan de bevindingen van dezelfde analyse uitgevoerd door de twee statistische pakketten die we tot nu toe steeds inzetten:

```
Number of studies combined: k = 84

Random effects model 1.1083 [1.0190; 1.1976] 24.33 < 0.0001
Prediction interval [0.6671; 1.5495]

Quantifying heterogeneity:
tau^2 = 0.0471 [0.0055; 0.0845]; tau = 0.2171 [0.0740; 0.2907]
I^2 = 32.2% [11.0%; 48.3%]; H = 1.21 [1.06; 1.39]

Test of heterogeneity:
Q d.f. p-value
122.43 83 0.0032

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 82)
(with ad hoc correction)
```

```
Random-Effects Model (k = 84; tau^2 estimator: REML)

loglik deviance AIC BIC AICC
-75.4660 150.9320 154.9320 159.7697 155.0820

tau^2 (estimated amount of total heterogeneity): 0.0471 (SE = 0.0207)
tau (square root of estimated tau^2 value): 0.2171
I^2 (total heterogeneity / total variability): 36.79%
H^2 (total variability / sampling variability): 1.58

Test for Heterogeneity:
Q(df = 83) = 122.4310, p-val = 0.0032

Model Results:

estimate se zval pval ci.lb ci.ub ***
1.1083 0.0456 24.3314 <.0001 1.0190 1.1976 ***

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Het verschil tussen beide resultaten zit hem in het gebruik van een net iets ander model met net iets andere correctiefactoren (zoals de Hartung-Knapp correctie voor kleine studies(100, 101)). Ook zit er een verschil in het pakket, maar onder de streep zien we grotendeels hetzelfde. Wat we ook zien is dat de gepresenteerde waarden voor de heterogeniteit soms

wel (en soms niet) met een 95% betrouwbaarheidsinterval worden getoond. Wel worden ze in beide pakketten statistisch getoetst en beide pakketten geven eenzelfde statistisch significant resultaat. Wat dit resultaat exact betekent is niet direct helder, omdat het niks zegt over de invloed van die heterogeniteit.

De heterogeniteit rondom een schatting kent dus zelf ook onzekerheid en dit zagen we al voor de I^2 waarde. We kunnen die onzekerheid ook tonen voor de andere schattingen door middel van een zogenaamde permutatietest(102). Deze levert een 95% betrouwbaarheidsinterval op basis van [profile likelihood estimates](#). Het resultaat toont aardig brede betrouwbaarheidsintervallen⁹⁵:

```
Random-Effects Model (k = 84; tau^2 estimator: REML)

  logLik deviance      AIC      BIC      AICC
-75.4660 150.9320 154.9320 159.7697 155.0820

tau^2 (estimated amount of total heterogeneity): 0.0471 (SE = 0.0207)
tau (square root of estimated tau^2 value):      0.2171
I^2 (total heterogeneity / total variability):   36.79%
H^2 (total variability / sampling variability):  1.58

Test for Heterogeneity:
Q(df = 83) = 122.4310, p-val = 0.0032

Model Results:

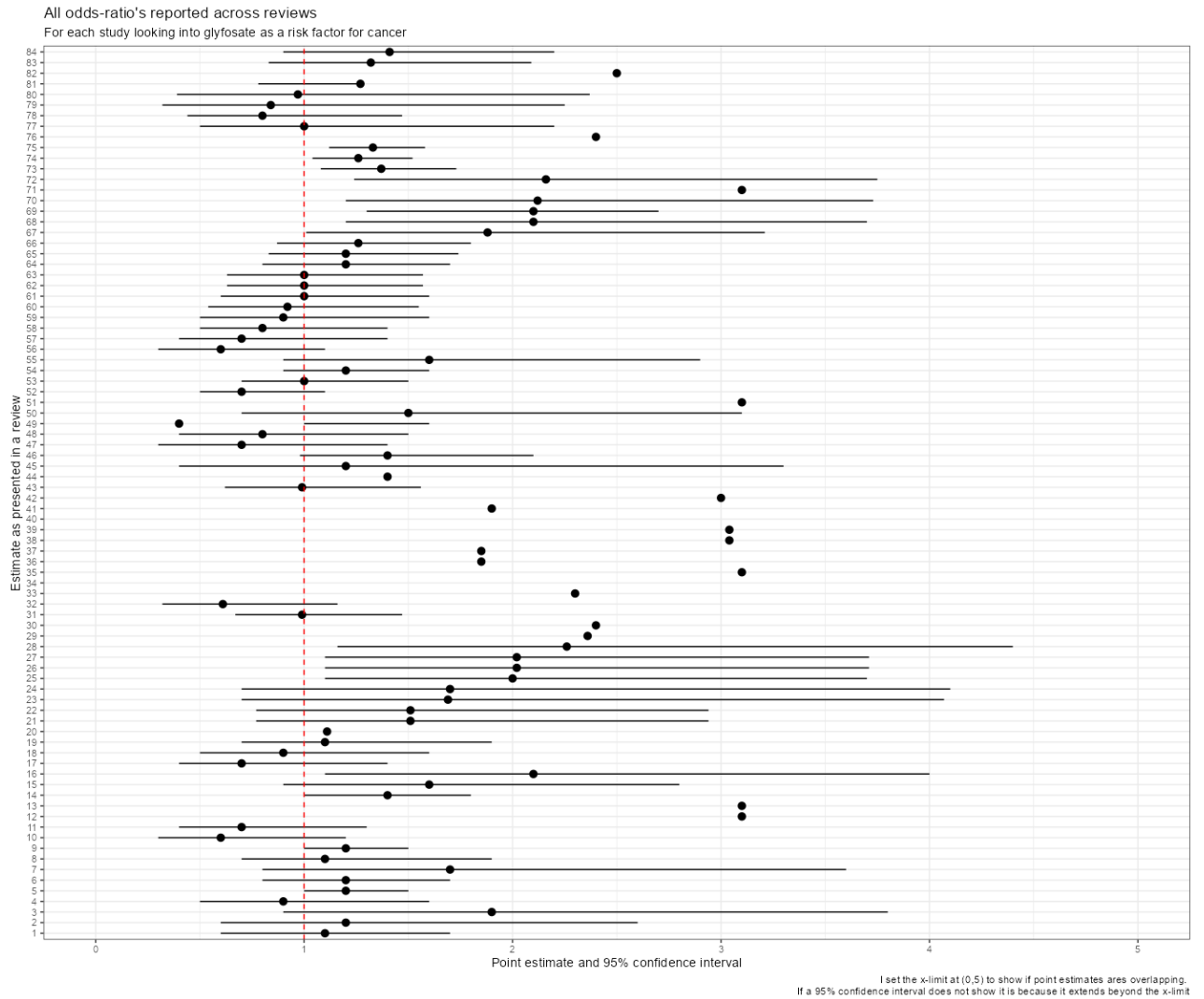
estimate      se      zval      pval      ci.lb      ci.ub      ***
  1.1083    0.0456   24.3314   <.0001   1.0190    1.1976

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

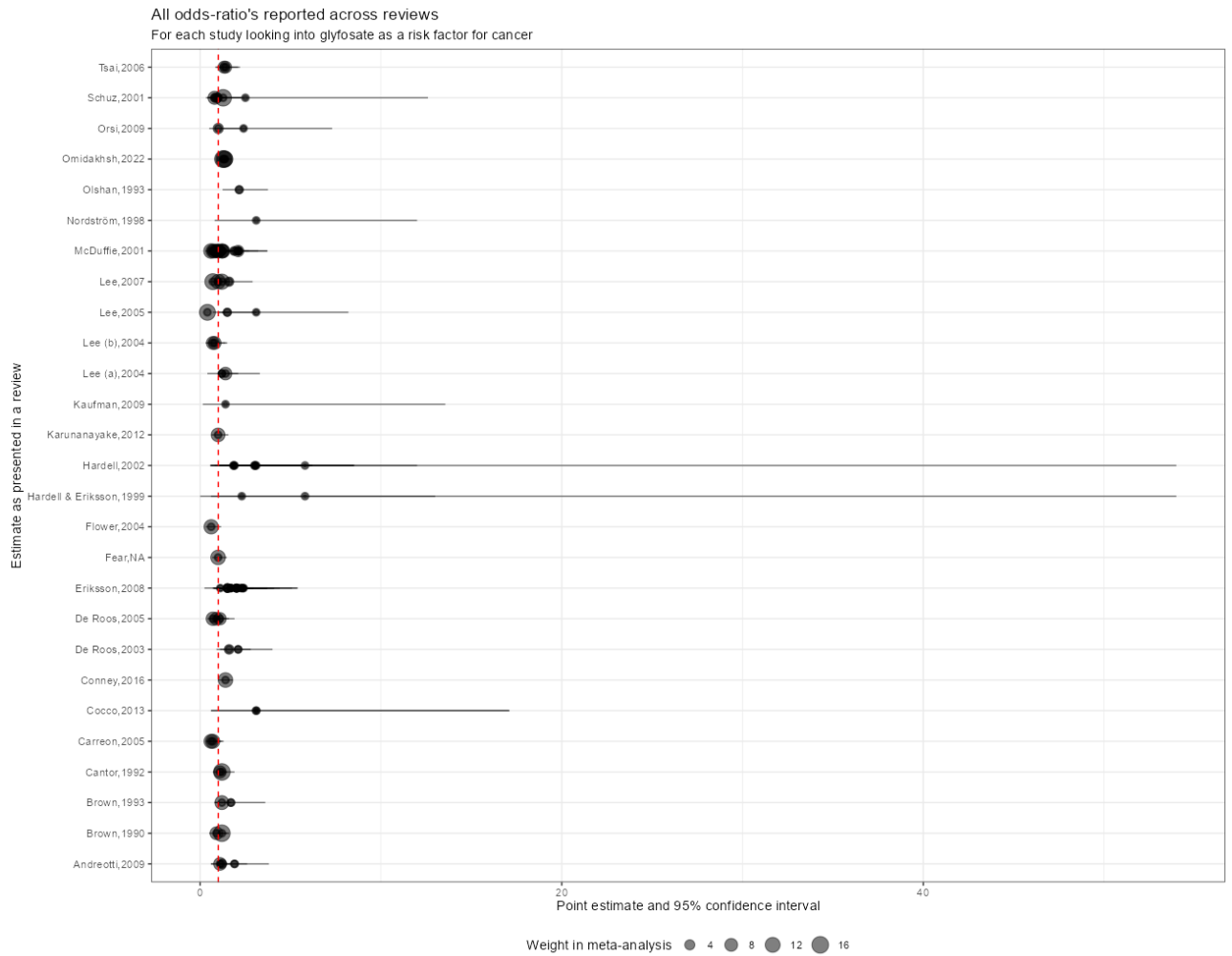
tau^2      estimate      ci.lb      ci.ub
tau         0.2171    0.1332    0.3211
I^2(%)     36.7853   17.9717   56.0201
H^2         1.5819    1.2191    2.2738
```

Op basis van deze bevindingen is het zinvol om te ontdekken wat de mogelijke oorzaak is van die heterogeniteit(27). Zo kan het heel goed zijn dat bepaalde studies significant meer bijdragen aan die heterogeniteit. Een dergelijke exercitie is zinvol omdat het ons kan helpen bij het uitbreiden van de meta-analyse in zinvolle subgroep analyses of meta-regressies. Een grafiek die ons hierbij kan helpen is de zogenaamde [Baujat grafiek](#)(103) waarvan het resultaat zichtbaar is in **Figuur 62**. Deze grafiek laat zien hoeveel elke studie bijdraagt aan de eergenoemde Q-waarde (die weer gebruikt wordt om I^2 te berekenen). We zien duidelijk één schatting die van invloed is op het resultaat.

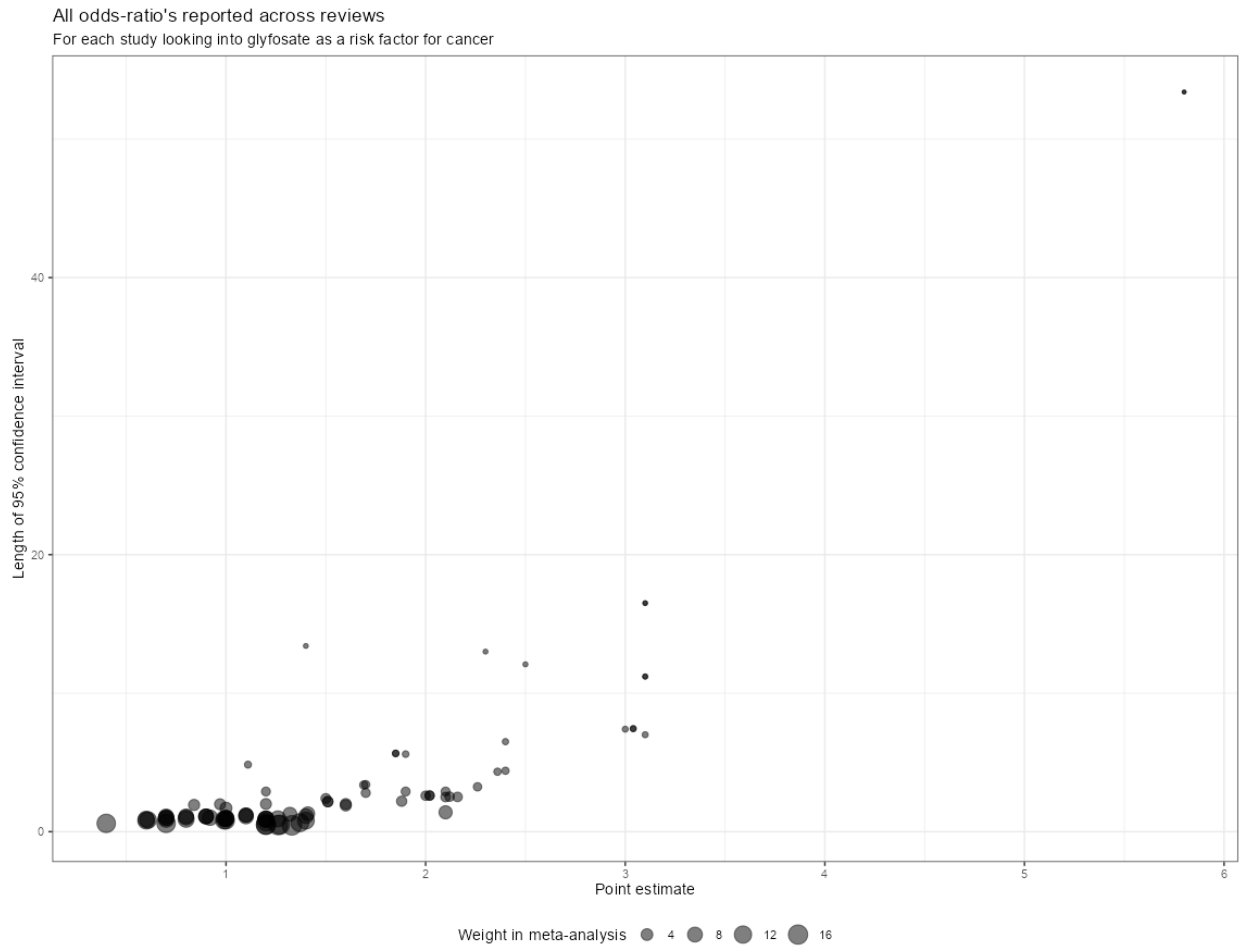
⁹⁵ De oplettende lezer zal hebben gezien dat de I^2 waarden nu niet meer rond de 32% zijn, maar wederom rond de 36%. Dit heeft te maken met het pakket waarmee ik de berekening uitvoer: [metafor](#) in plaats van [meta](#). De lezer die dus dieper in dit rapport wil duiken doet er goed om de codes tot zich te nemen. Dan kun je direct zien welk pakket voor welke analyse is gebruikt.



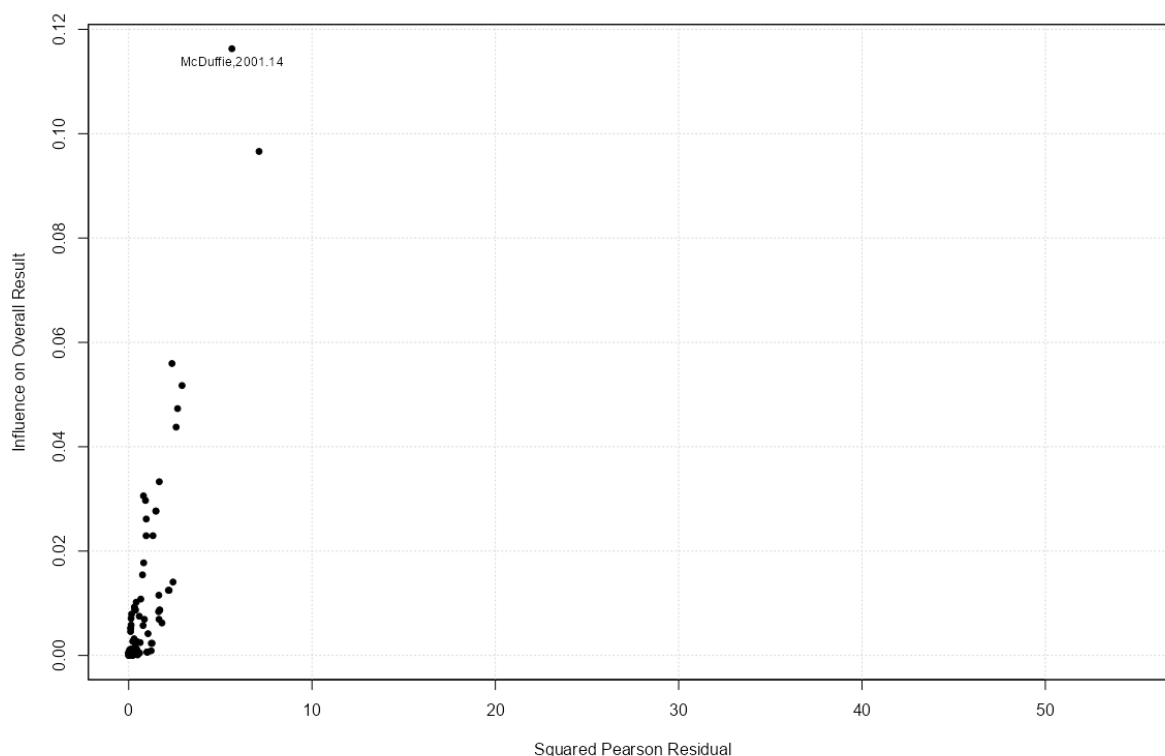
Figuur 59. Odds-ratio's met 95% betrouwbaarheidsintervallen zoals verzameld over de reviews heen. De rode stippellijn is de lijn van geen-effect.



Figuur 60. Odds-ratio's met 95% betrouwbaarheidsintervallen. De grootte van elke schatting is afhankelijk van het gewicht in de meta-analyse. Zichtbaar is dat niet elke dubbel telling eenzelfde gewicht heeft gekregen.

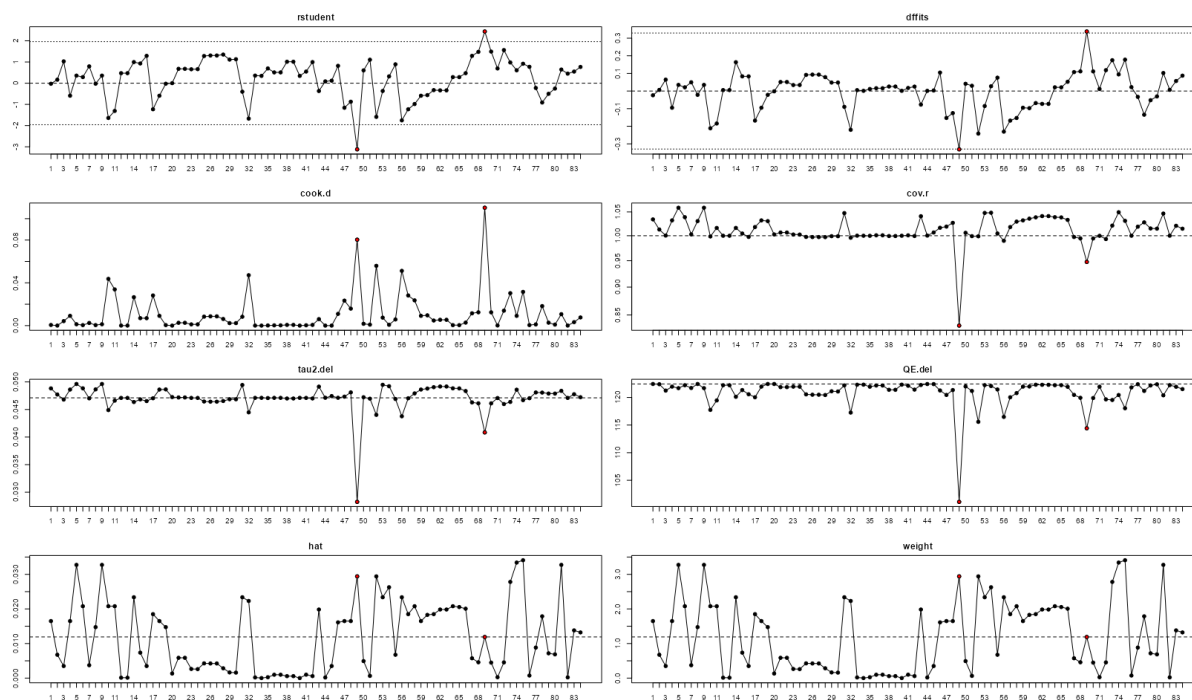


Figuur 61. Puntchattingen van alle *odds-ratio's* afgezet tegen de lengte van de 95% betrouwbaarheidsintervallen. De grootte van elke schatting is afhankelijk van het gewicht in de meta-analyse. Duidelijk is dat het gewicht direct gebonden is aan de lengte van het 95% betrouwbaarheidsinterval.



Figuur 62. Baujat grafiek voor de meta-analyse van alle studies in de database die hebben gekeken naar de relatie tussen glyfosaat en kanker.

Een verdere statistische uitbreiding van dit concept is het iteratief weglaten (**Figuur 63**) van studies. Elke keer als een studie wordt weggelaten worden de statistieken zoals we deze kennen uit onder andere **Tabel 2** opnieuw berekend. Hiermee kunnen we zien hoe stabiel onze bevindingen zijn. Hoewel we in deze exercitie verschillende schattingen dubbeltellen, lijkt het alsof er maar twee schattingen een grote impact hebben: de 2^{de} schatting van Lee(72) en de 14^{de} schatting van McDuffie(76). De invloed van de schatting van McDuffie(76) zagen we in **Figuur 62** ook al terugkomen.



Figuur 63. Influence statistics voor de meta-analyse van alle studies met odds-ratio's die hebben gekeken naar de relatie tussen glyfosaat en kanker.

Een laatste exercitie die ik nog kunnen uitvoeren is het daadwerkelijk verwijderen van beide schattingen. Dit zijn dan de resultaten:

```

Number of studies combined: k = 82

Random effects model 1.0913 [1.0041; 1.1784] 95%-CI z p-value
Prediction interval [0.6791; 1.5035] 24.54 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.0409 [0.0025; 0.0721]; tau = 0.2023 [0.0504; 0.2686]
I^2 = 28.9% [6.2%; 46.2%]; H = 1.19 [1.03; 1.36]

Test of heterogeneity:
Q d.f. p-value
113.95 81 0.0093

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 80)
(with ad hoc correction)
    
```

De *summary estimate* blijft significant⁹⁶, het voorspellingsinterval blijft niet significant en de heterogeniteit daalt iets maar blijft noemenswaardig (en met een groot 95% betrouwbaarheidsinterval).

⁹⁶ Maar is dat dus niet, want hier wordt de verkeerde statistiek gebruikt door het programma.

Publication bias

Als laatste wil ik stil staan bij een ander belangrijk onderdeel. In dit onderdeel onderzoeken we of er sprake is van zogenaamde *publication bias*: het niet publiceren van een bepaald resultaat (positief of negatief) vanwege de aard van de bevinding (104, 105).

Een resultaat wat duidt op de **afwezigheid** van publicatiebias is een resultaat waarin het gros van de studies een spreiding toont in het gevonden effect én de variatie van dat effect. Met andere woorden: je verwacht dat studies een reeks aan bevindingen laat zien. Ook bevindingen die gek lijken, zoals een beschermende rol van pesticiden zoals glyfosaat. Wat je niet wil zien is dat een meta-analyse alleen maar kleine positieve studies meeneemt.

De bekendste instrumenten voor het testen van de mogelijkheid op *publication bias* zijn de *Egger Test*, *Funnel plot*(104) en de *Trim-and-Fill Test*(106). De resultaten van de *Egger Test* en de *Funnel plot* laten zien dat publicatie bias een mogelijkheid is (**Figuur 64**) De *Egger Test* is statistisch significant en de *Funnel plot* vertoont een concentratie rechtsboven:

```
Eggers' test of the intercept
=====
intercept      95% CI      t      p
      0.738 0.33 - 1.14  3.561 <0.001
Eggers' test indicates the presence of funnel plot asymmetry
```

De *Trim-and-Fill* procedure is een simulatie procedure die op basis van de data met publicatie bias een nieuwe dataset maakt waarin de publicatiebias als het ware wordt opgeheven. Om de publicatie bias 'op te heven' worden er 29 studies' toegevoegd' die elk een spiegel vormen van een andere studie. Op die manier kom ik op 113 studies uit (**Figuur 65**)⁹⁷. Wanneer ik die exercitie uitvoer en de meta-analyse opnieuw uitvoer krijg ik een niet significante *summary estimate*⁹⁸:

```
Estimated number of missing studies on the left side: 29 (SE = 5.7657)
Random-Effects Model (k = 113; tau^2 estimator: REML)

logLik  deviance  AIC      BIC      AICC
-134.8793  269.7586  273.7586  279.1956  273.8687
```

⁹⁷ Dit is een verre van perfecte procedure en wie in de codes duikt kan zien dat, afhankelijk van het type statistiekpakket én soort model, het aantal toegevoegde studies tussen de 26 en 31 studies ligt. Wel is het zo dat in elk van die analyse de *summary estimate* niet meer significant is.

⁹⁸ Dit is uiteraard een gesimuleerde bevinding en dient eerder ter verduidelijk van de mogelijke impact van de *publication bias* dan dat het een weerspiegeling is van het 'echte', maar niet gevonden effect van glyfosaat op kanker als we alleen naar de *odds-ratio's* kijken.

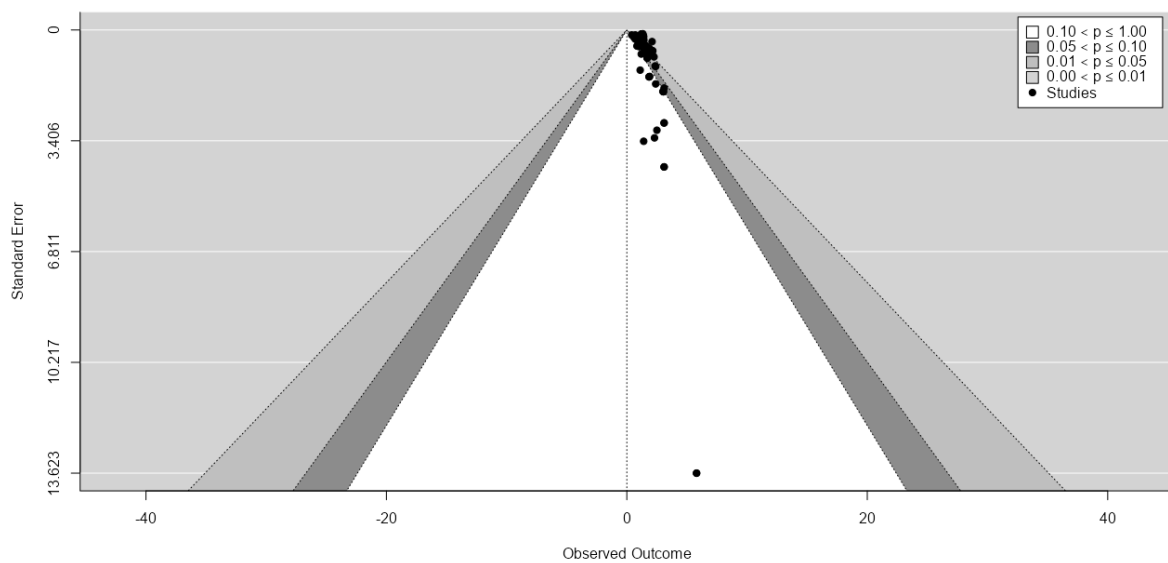
```
tau^2 (estimated amount of total heterogeneity): 0.0556 (SE = 0.0225)
tau (square root of estimated tau^2 value): 0.2358
I^2 (total heterogeneity / total variability): 34.51%
H^2 (total variability / sampling variability): 1.53
```

```
Test for Heterogeneity:
Q(df = 112) = 168.2635, p-val = 0.0005
```

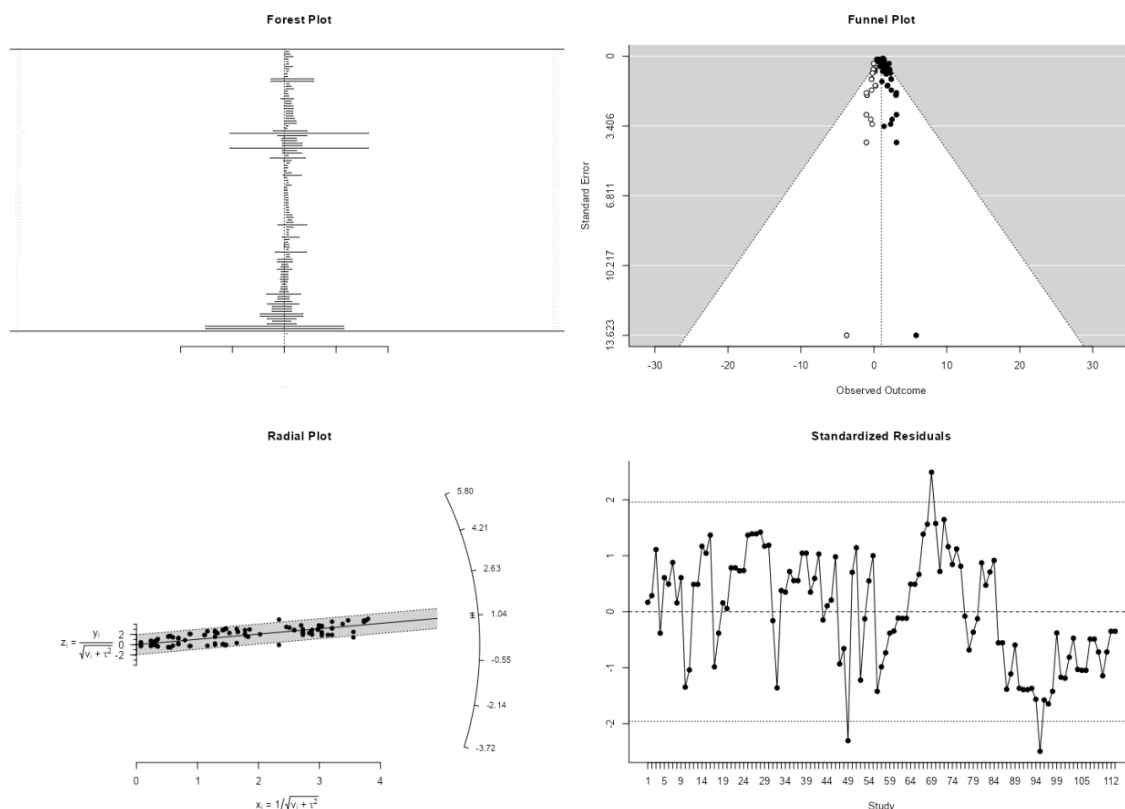
Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
1.0392	0.0459	22.6519	<.0001	0.9493	1.1291	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Figuur 64. *Funnel plot* van de geobserveerde studies. De plot laat zien dat de meeste bevindingen aan de rechterzijde en bovenaan de funnel zitten. Dit wijst op *publication bias*.



Figuur 65. Diagnostische plots rondom de aanwezigheid van *publication bias*. De figuur rechtsboven toont de *funnel plot* voor *publication bias* met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de *publication bias* teniet te doen.

Prediction interval

Voordat ik verderga wil ik stil staan bij een belangrijk verschil tussen het 95% betrouwbaarheidsinterval en het 95% voorspellingsinterval ([prediction interval](#))(107-109). Het 95% betrouwbaarheidsinterval beschrijft de onzekerheid rondom de puntschatting. Dit is de onzekerheid op groepsniveau. Het 95% voorspellingsinterval beschrijft de onzekerheid op het niveau van de persoon. Dat betekent dat het 95% voorspellingsinterval iets zegt over de kans dat iemand die is blootgesteld aan glyfosaat ook kanker krijgt (**Tabel 3**). In dit geval is de onzekerheid zo groot dat er niet gesproken kan worden van een associatie. De vraag is nu of die onzekerheid komt door de onzekerheid tussen de studies en/of de onzekerheid in een studie. Daarom wordt het tijd om wat dieper in de subgroep analyses te duiken waarbij een meta-analyse per kankersoort de meest logische beginstap is⁹⁹.

⁹⁹ En eigenlijk een noodzakelijke stap.

Statistiek	Methode berekening	Metafor pakket	Meta pakket
Summary estimate	REML	1.11 (1.02; 1.19)	1.11 [1.02;1.19]
Prediction interval	REML		1.11 (0.67; 1.55)

Tabel 3. *Summary estimate* en *prediction interval* voor het risico op kanker door glyfosaat berekend vanuit beide pakketten voor gerapporteerde *odds-ratio's*.

Meta-analyses per kankersoort

We hebben al eerder gezien dat we een hoop kankersoorten hebben geïnccludeerd, maar alleen van een aantal voldoende studies hebben (minstens vijf¹⁰⁰) om een enigszins betrouwbare schatting te maken(110). Waarbij betrouwbaar nu alleen een functie is van aantallen en we het nog niet eens gehad hebben over wat er aan heterogeniteit te vinden is in elke kankersoort. Die uitstap gaan we nu maken en de eerste splitsing van de data toont dat we eigenlijk maar voor vier kankersoorten een meta-analyse kunnen uitvoeren (**Figuur 66**):

```
disease      n
<chr>      <int>
1 HCL        5
2 MM         8
3 NHL       14
4 wilm's tumor 5
```

Een verdere splitsing voor het soort statistiek dat is gerapporteerd toont dat het gros *odds-ratio's* zijn, maar ook dat er voor een ziekte zoals NHL ongeveer evenveel *odds-ratio's* als *risk-ratio's* zijn:

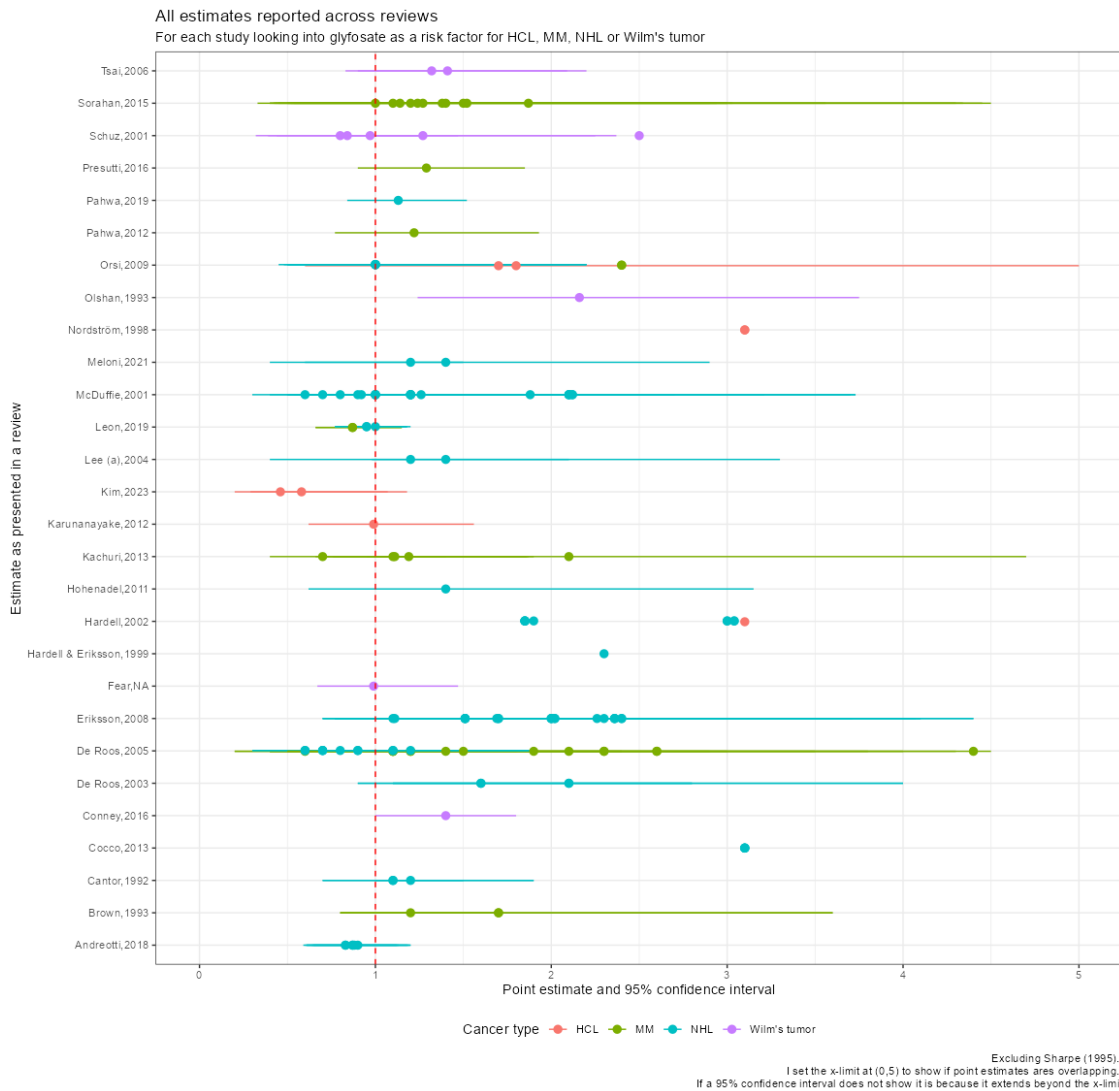
```
# A tibble: 11 x 3
# Groups:   disease [4]
disease      risk_metric      n
<chr>      <chr>      <int>
1 HCL        HR          2
2 HCL        OR          3
3 HCL        RR          3
4 MM         HR          1
5 MM         OR          3
6 MM         RR         37
7 NHL        HR          1
8 NHL        OR         44
9 NHL        RR         49
10 NHL       NA          2
11 wilm's tumor OR         10
```

¹⁰⁰ Ik noem hier nu een getal waarvan ik het gevoel heb dat dit passend genoeg zou moeten zijn om er een variantie over te berekenen, maar eigenlijk kun je niet a priori zeggen dat een bepaald getal voldoende is. Het aantal studies dat nodig is om een enigszins betrouwbare meta-analyse uit te voeren is van meer dan alleen het aantal studies afhankelijk: er is dus niet ergens een lijst voorhanden die je kunt doorlopen om een beslissing te maken. Wederom nodig ik de lezer uit om zelf met de data en de codes aan de slag te gaan.

Deze tabel wordt echter een stuk kleiner als ik eis dat ik ook voor de mate van blootstelling informatie heb:

```
# A tibble: 5 x 3
# Groups:   disease [3]
  disease risk_metric n
<chr>    <chr>    <int>
1 MM      OR         1
2 MM      RR        24
3 NHL     OR        29
4 NHL     RR        28
5 Wilm's tumor OR         8
```

Zelf lijkt het mij beter om bij deze laatste tabel te blijven, omdat we informatie rondom blootstelling later toch nog zullen moeten meenemen. Dan kunnen we dat maar beter vanaf het begin doen en het gemis aan informatie voor deze belangrijke variabele opnemen als een belangrijke bevinding.



Figuur 66. Risicoschattingen met 95% betrouwbaarheidsinterval voor elke studie opgenomen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM), hairy cell leukemie (HCL) en nierkanker (*Wilm's tumor*).

Ook de splitsing per statistiek is een belangrijke stap om te nemen en zelfs noodzakelijk, maar voor de eerste meta-analyse wil ik dat nog niet doen. Het resultaat zien we hieronder (en afgebeeld in **Figuur 67**):

```

Number of studies combined: k = 90

Random effects model 1.0784 [0.9905; 1.1662] 24.06 < 0.0001
Prediction interval [0.7102; 1.4465]

Quantifying heterogeneity:
tau^2 = 0.0323 [0.0000; 0.0815]; tau = 0.1797 [0.0000; 0.2854]
I^2 = 16.7% [0.0%; 36.6%]; H = 1.10 [1.00; 1.26]

Test of heterogeneity:
Q d.f. p-value
106.84 89 0.0957

Results for subgroups (random effects model):
k 95%-CI tau^2 tau Q I^2
disease = NHL 57 1.0617 [0.9599; 1.1635] 0.0372 0.1928 85.23 34.3%
disease = MM 25 1.1014 [0.8536; 1.3493] 0.0029 0.0541 10.55 0.0%
disease = wilm's tumor 8 1.2159 [1.0249; 1.4069] 0 0 6.30 0.0%

Test for subgroup differences (random effects model):
Q d.f. p-value
Between groups 1.95 2 0.3768

Prediction interval for subgroups:
95%-PI
disease = NHL [0.6596; 1.4637]
disease = MM [0.8169; 1.3860]
disease = wilm's tumor [0.9774; 1.4543]

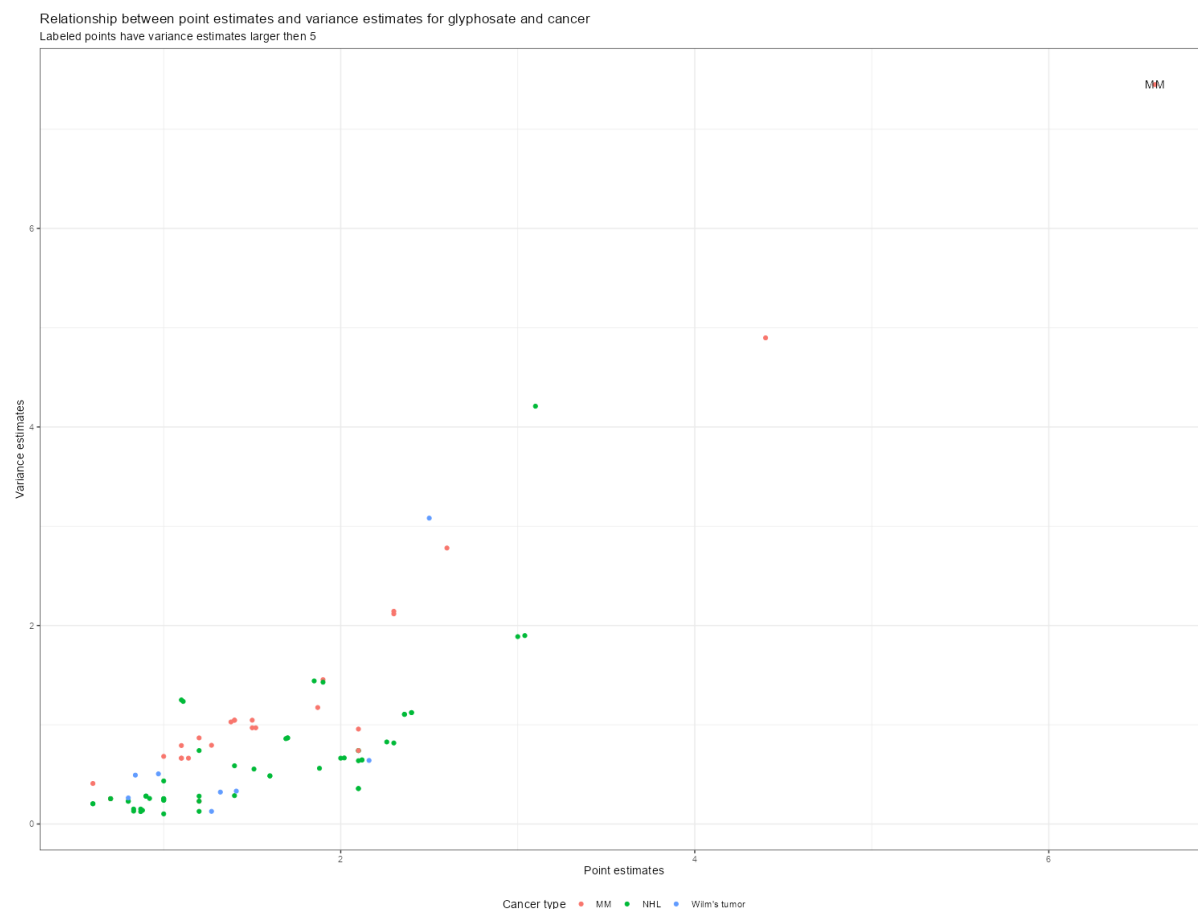
Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 88)
  (with ad hoc correction)
    
```

Bovenstaand resultaat laat zien dat als we alle studies meenemen en op één hoop gooien het resultaat niet significant is. Als we een splitsing maken per ziekte is deze ook niet significant, behalve voor nierkanker. Maar van alle schattingen zijn de 95% voorspellingsintervallen niet significant. De heterogeniteit tussen studies is alleen substantieel voor NHL wat erop kan wijzen dat we te verschillende maten van blootstelling hebben meegenomen (of door het verschil in statistiek).

Maar nogmaals, deze bevindingen zijn eigenlijk te onbruikbaar om te waarderen omdat we té veel op één hoop hebben gegooid¹⁰¹. De reden om deze exercitie toch te doen is omdat het een eerste stap is in een verdere lijn van stappen om zo te komen bij schattingen waarvan de waardebepaling meer zin heeft. De tabel hieronder en **Figuur 68**

¹⁰¹ Ondanks dat alleen NHL een mix van *odds-ratio*'s en *risk-ratio*'s laat zien, kan het heel goed zijn dat de andere kankersoorten heterogeen zijn vanwege andere zaken zoals de mate van blootstelling. Dat vereist nog wat extra onderzoek, tenzij een verdere splitsing de analyse té onbetrouwbaar maakt.

toont de resultaten van een *trimm-and-fill* procedure die voor nu alleen als illustratie bruikbaar is.



Figuur 67. Schattingen en varianties per schattingen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en nierkanker (*Wilms tumor*).

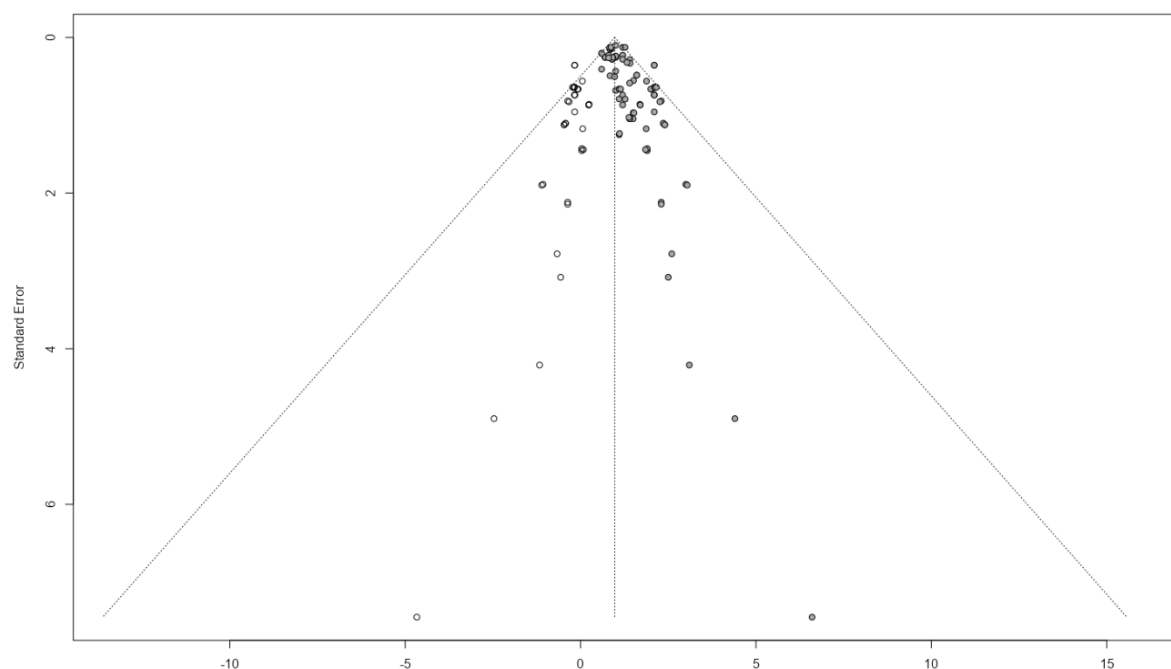
```

Number of studies combined: k = 125 (with 35 added studies)
                                95%-CI      z  p-value
Random effects model 0.9747 [0.8797; 1.0697] 20.11 < 0.0001
Prediction interval  [0.4826; 1.4668]

Quantifying heterogeneity:
tau^2 = 0.0595 [0.0236; 0.2451]; tau = 0.2438 [0.1536; 0.4951]
I^2 = 31.3% [14.0%; 45.1%]; H = 1.21 [1.08; 1.35]

Test of heterogeneity:
  Q d.f. p-value
180.44 124 0.0007

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Prediction interval based on t-distribution (df = 123)
- Trimm-and-fill method to adjust for funnel plot asymmetry
    
```



Figuur 68. De *funnel plot* voor *publication bias* met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de *publication bias* teniet te doen.

Het wordt nu tijd om nog wat dieper in de materie te duiken waarbij een focus op NHL een goede keus is. We hebben voor NHL de meeste schattingen opgenomen in de database en zagen al dat we haast net zoveel *odds-ratio's* als *risk-ratio's* hadden. Voor MM hadden we één *odds-ratio* op 24 *risk-ratio's* en voor Wilm's tumor (nierkanker) hadden we allemaal *odds-ratio's*.

Meta-analyses voor NHL en per statistiek

Een eerstvolgende logische stap is dus het opsplitsen van de data voor NHL per statistiek. Ook kunnen we een stap maken door te kijken naar een specifiek soort *study design*, juist omdat we weten dat er een relatie is tussen het ontwerp van de studie en de gebruikte statistiek¹⁰². De tabel hieronder laat zien wat er met de data gebeurt als we voor NHL verdere indelingen maken. Zo zien we, wederom, dat we net zoveel *odds-ratio's* als *risk-ratio's* hebben. Ook zien we dat we zowel *case-control* als cohort studies hebben en we zien

¹⁰² Met de gehanteerde analyse methode als mediërende variabele.

welke studies we in totaal voor NHL hebben opgenomen. Uiteindelijk gaan we werken met 57 schattingen:

```

disease risk_metric      n
<chr>    <chr>          <int>
1 NHL     OR              29
2 NHL     RR              28

study_design risk_metric  n
<chr>        <chr>      <int>
1 Case-control OR        27
2 Case-control RR        18
3 Cohort      OR         2
4 Cohort      RR        10

study_author_year risk_metric  n
<chr>              <chr>    <int>
1 Andreotti,2018   RR         5
2 Cantor,1992      OR         1
3 Cocco,2013       RR         1
4 De Roos,2003     OR         2
5 De Roos,2003     RR         2
6 De Roos,2005     OR         2
7 De Roos,2005     RR         4
8 Eriksson,2008    OR         8
9 Eriksson,2008    RR         6
10 Hardell,2002    OR         2
11 Hardell,2002    RR         2
12 Lee (a),2004    OR         2
13 Leon,2019       RR         1
14 McDuffie,2001   OR        12
15 McDuffie,2001   RR         4
16 Meloni,2021     RR         2
17 Orsi,2009       RR         1
    
```

In een meta-analyse kan ik de schattingen voor NHL splitsen per type statistiek en het resultaat toont zich hieronder. Wat we zien is dat geen van de schattingen statistisch significant is¹⁰³, dat de 95% voorspellingsintervallen groot zijn en dat de heterogeniteit noemenswaardig blijft voor NHL ongedacht de splitsing per type statistiek (zie ook **Figuur 69**):

```

Number of studies combined: k = 57

          95%-CI      z  p-value
Random effects model 1.0617 [0.9599; 1.1635] 20.44 < 0.0001
Prediction interval  [0.6596; 1.4637]

Quantifying heterogeneity:
tau^2 = 0.0372 [0.0081; 0.1963]; tau = 0.1928 [0.0898; 0.4431]
I^2 = 34.3% [8.9%; 52.6%]; H = 1.23 [1.05; 1.45]

Test of heterogeneity:
  Q d.f. p-value
85.23 56 0.0071
    
```

¹⁰³ Wederom zien we dat het programma dit niet automatisch herkent en dus de verkeerde p-waarde presenteert. Ik kom daar later op terug.

```

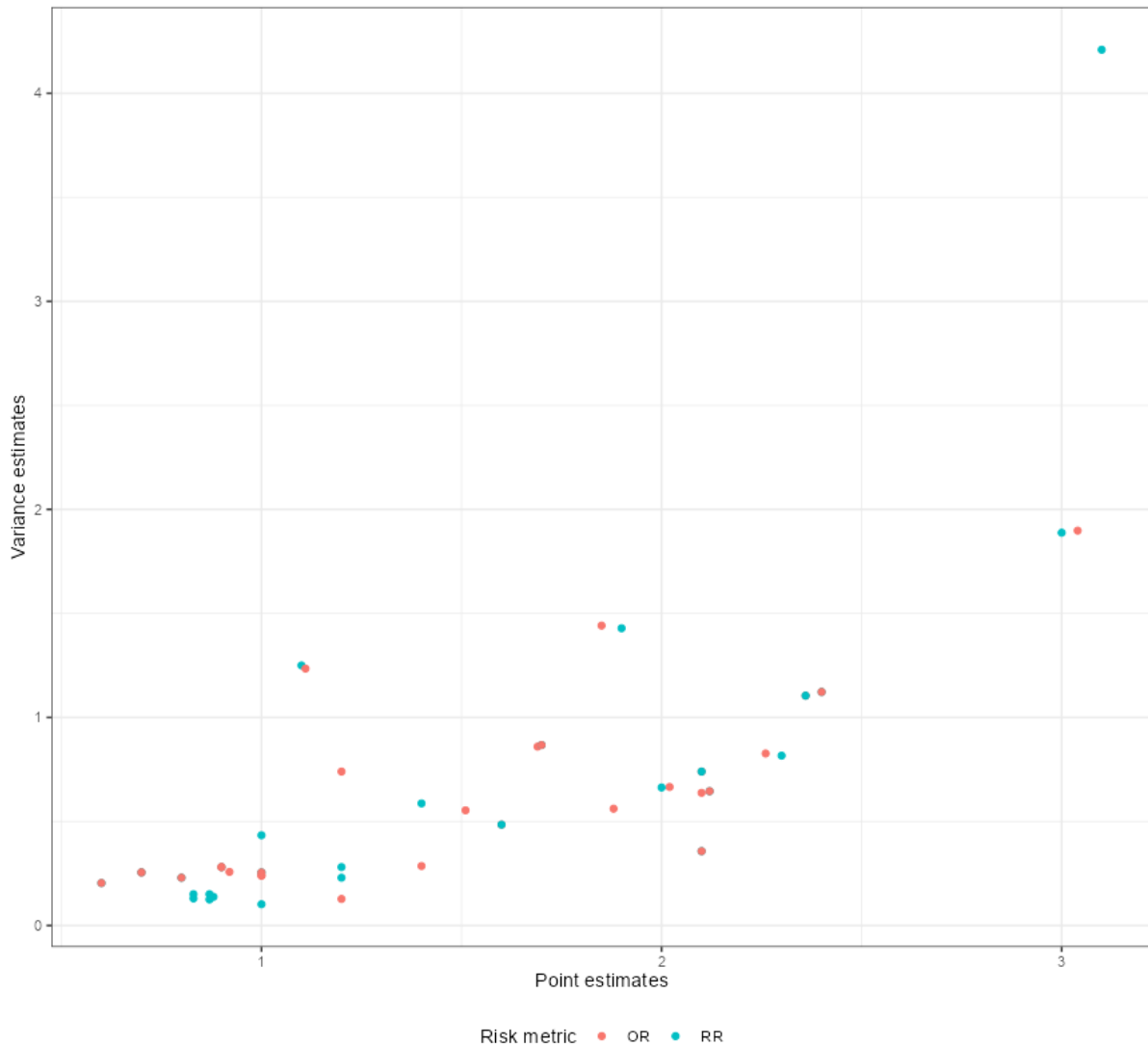
Results for subgroups (random effects model):
          k          95%-CI  tau^2    tau    Q    I^2
risk_metric = RR  28  0.9505 [0.8638; 1.0373]  0    0  38.71  30.2%
risk_metric = OR  29  1.1666 [0.9859; 1.3474]  0.0776  0.2786  43.26  35.3%

Test for subgroup differences (random effects model):
          Q d.f. p-value
Between groups 4.46    1  0.0346

Prediction interval for subgroups:
          95%-PI
risk_metric = RR [0.8417; 1.0594]
risk_metric = OR [0.5645; 1.7687]

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 55)
(with ad hoc correction)
    
```

Relationship between point estimates and variance estimates for glyphosate and non-hodgkin lymphoma
 Labeled points have variance estimates larger then 5



Figuur 69. Risicoschattingen voor glyphosaat en non-hodgkin lymfoom (NHL) afgezet tegen de variantieschattingen. De kleur geeft de gebruikte statistiek aan.

Meta-analyses voor NHL en per study design

Een andere subgroep analyse die we kunnen uitvoeren is er een waarbij we expliciet een splitsing maken tussen *case-control* en cohort studies. Nu weten we dat als we dit doen dat we toch statistieken combineren die eigenlijk niet goed te combineren zijn, maar voor nu gaat het om wat splitsen van de data met de resultaten doet. Het resultaat zien we hieronder:

study_design <chr>	risk_metric <chr>	n <int>
1 Case-control	OR	27
2 Case-control	RR	18
3 Cohort	OR	2
4 Cohort	RR	10

```

Number of studies combined: k = 57

Random effects model 1.0617 [0.9599; 1.1635] 20.44 < 0.0001
Prediction interval [0.6596; 1.4637]

Quantifying heterogeneity:
tau^2 = 0.0372 [0.0081; 0.1963]; tau = 0.1928 [0.0898; 0.4431]
I^2 = 34.3% [8.9%; 52.6%]; H = 1.23 [1.05; 1.45]

Test of heterogeneity:
Q d.f. p-value
85.23 56 0.0071

Results for subgroups (random effects model):
k 95%-CI tau^2 tau Q I^2
study_design = Cohort 12 0.8626 [0.7712; 0.9541] 0 0 4.52 0.0%
study_design = Case-control 45 1.3215 [1.1516; 1.4915] 0.1011 0.3179 60.52 27.3%

Test for subgroup differences (random effects model):
Q d.f. p-value
Between groups 21.72 1 < 0.0001

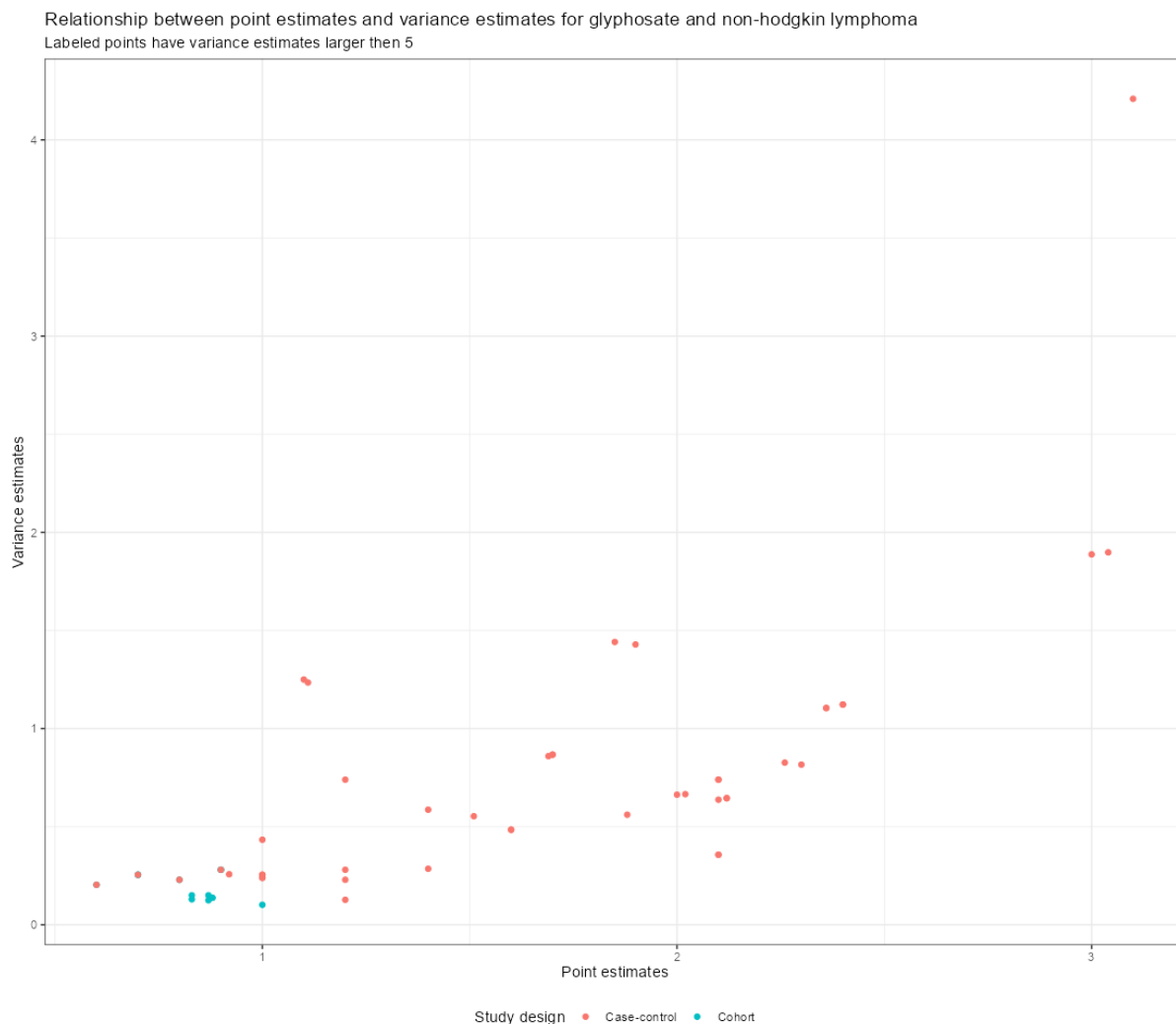
Prediction interval for subgroups:
95%-PI
study_design = Cohort [0.7586; 0.9666]
study_design = Case-control [0.6570; 1.9861]

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 55)
(with ad hoc correction)
    
```

De meta-analyse laat geen significante effecten zien¹⁰⁴, behalve voor de *case-control* studies. Dit hebben we al eerder gezien in dit rapport en zagen we ook in het [rapport](#) over risicofactoren voor de ziekte van Parkinson. Omdat *case-control* studies starten met de ziekte, en niet met de risicofactor, is het makkelijker om een statistisch significant verschil te vinden. We kunnen verder zien dat er, in tegenstelling tot de cohort studies, toch nog wat

¹⁰⁴ Wederom het probleem met de p-waarde.

heterogeniteit is. Dit zien we beter in **Figuur 70** waar de schattingen voor de cohort studies op elkaar clusteren terwijl die van de *case-control* studies verspreid liggen. Een mogelijke reden hiervoor is dat er maar een paar cohort studies zijn en verschillende studies aan de haal zijn gegaan met dezelfde data.



Figuur 70. Punt-schattingen en onzekerheidsschattingen voor de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL). De kleur geeft het type *study design* aan.

Het wordt nu langzaam tijd om strakker te selecteren en zo, minstens voor NHL, te achterhalen wat alle studies zeggen over een relatie met glyfosaat.

Meta-analyses voor NHL odds-ratio's en exposure

Een volgende subgroep analyse is eigenlijk geen subgroep analyse zoals we die voorheen hebben gezien, maar een analyse waarin ik een selecte groep tot mij neem. Ik kier er bewust

voor om alleen te kijken naar *odds-ratio's* en daarmee kijk ik dus ook alleen maar naar *case-control* studies. Met die selectie houd ik nog maar zes studies over (57, 61, 64, 66, 73, 76):

```
# Groups:   study_design [2]
  study_design risk_metric      n
  <chr>      <chr>          <int>
1 Case-control OR             11
2 Case-control RR             10
3 Cohort      RR              1

# A tibble: 6 x 3
# Groups:   study_author_year [6]
  study_author_year risk_metric      n
  <chr>             <chr>          <int>
1 Cantor,1992      OR              1
2 De Roos,2003    OR              2
3 Eriksson,2008   OR              2
4 Hardell,2002    OR              2
5 Lee (a),2004    OR              2
6 McDuffie,2001   OR              2
```

Het resultaat van deze keuze zien we hieronder. Ik zie geen dubbelen, en ik zie een statistisch significant effect voor het 95% betrouwbaarheidsinterval¹⁰⁵, maar niet voor het 95% voorspellingsinterval (**Figuur 71**). Ook zie ik weinig heterogeniteit¹⁰⁶:

```
Review:      Glyphosate and NHL for the exposed group

              OR           95%-CI %w(random)
Cantor,1992  1.2000 [1.0000; 1.5000] 29.6
De Roos,2003 2.1000 [1.1000; 4.0000] 6.8
De Roos,2003 1.6000 [0.9000; 2.8000] 8.5
Eriksson,2008 1.5100 [0.7700; 2.9400] 6.4
Eriksson,2008 2.0200 [1.1000; 3.7100] 7.6
Hardell,2002 1.8500 [0.5500; 6.2000] 2.2
Hardell,2002 3.0400 [1.0800; 8.5200] 2.9
Lee (a),2004 1.2000 [0.4000; 3.3000] 2.8
Lee (a),2004 1.4000 [0.9800; 2.1000] 15.4
McDuffie,2001 0.9200 [0.5400; 1.5500] 9.5
McDuffie,2001 1.8800 [1.0100; 3.2100] 8.2

Number of studies combined: k = 11

              OR           95%-CI      z  p-value
Random effects model 1.4492 [1.2073; 1.7396] 3.98 < 0.0001
Prediction interval  [0.9972; 2.1060]

Quantifying heterogeneity:
tau^2 = 0.0186 [0.0000; 0.1916]; tau = 0.1364 [0.0000; 0.4377]
I^2 = 10.7% [0.0%; 50.7%]; H = 1.06 [1.00; 1.42]

Test of heterogeneity:
  Q d.f. p-value
11.20 10 0.3423

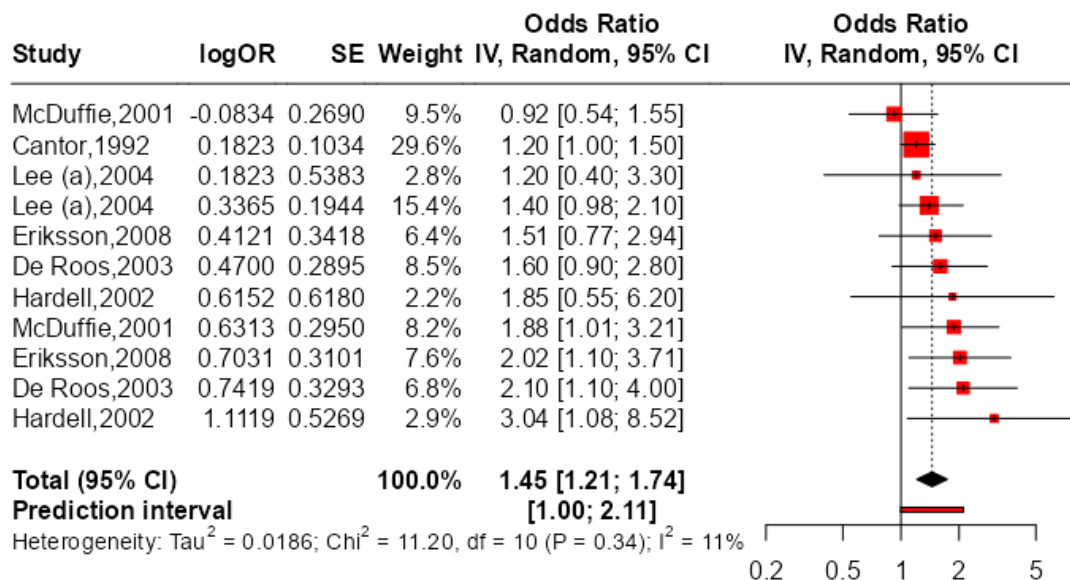
Details on meta-analytical method:
```

¹⁰⁵ Wat deze keer wel klopt.

¹⁰⁶ Voor de oplettende lezer: dit is de eerste meta-analyse waarin ik in de analyse *a priori* specificeer dat het om *odds-ratio's* gaat. Zou ik dit niet doen, dan zou het programma automatisch de 0 als de waarde van geen-effect nemen (bij een *odds-ratio* is dat de 1) en de gewichten anders verdelen. Het zou tot een andere schatting leiden. Dus, wat we tot nu toe hebben gedaan is het combineren van studies die zich eigenlijk helemaal niet goed laten combineren. Daarom is dit pas de eerste meta-analyse waar we ook echt iets over kunnen zeggen over de gepresenteerde waarden.

```

- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 9)
  (with ad hoc correction)
    
```



Figuur 71. Forest plot die de odds-ratio's laat zien voor de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL) in de groep die is blootgesteld aan glyfosaat (in welke hoedanigheid dan ook).

Maar de vraag is of we deze studies wel helemaal bij elkaar mogen zetten. Daarvoor helpt het wellicht als we ook de *exposure category* tot ons nemen en dan zien we de volgende tabel:

```

# A tibble: 11 x 6
  study_author_year risk_metric risk_estimate risk_estimate_low risk_estimate_high exposure_category
<chr>              <chr>          <dbl>          <dbl>          <dbl>          <dbl> <chr>
1 Cantor,1992      OR              1.2            1              1.5  Farmer/nonfarmer
2 De Roos,2003    OR              2.1            1.1            4      =1 day/lifetime
3 De Roos,2003    OR              1.6            0.9            2.8  =1 day/lifetime
4 Eriksson,2008   OR              1.51           0.77           2.94 =1 day (multivariate)
5 Eriksson,2008   OR              2.02           1.1            3.71 =1 day (univariate)
6 Harde11,2002    OR              1.85           0.55           6.2  Exposed (multivariate)
7 Harde11,2002    OR              3.04           1.08           8.52 Exposed (univariate)
8 Lee (a),2004    OR              1.2            0.4            3.3  Asthmatics
9 Lee (a),2004    OR              1.4            0.98           2.1  Nonasthmatics
10 McDuffie,2001  OR              0.92           0.54           1.55 Glyphosate only
11 McDuffie,2001  OR              1.88           1.01           3.21 High exposure
    
```

Ik heb niet direct het idee dat ik onmogelijke schattingen aan het combineren ben, maar ik wat ik wel zie is dat ik *univariate* en *multivariate* schattingen uit dezelfde studie combineer. Het is beter om alleen de *multivariate* schatting te nemen, omdat deze voor meerdere factoren statistisch gecorrigeerd heeft. Dat wil niet zeggen dat deze analyse per definitie beter is, maar het voelt beter om deze te kiezen dan om beide schattingen voor dezelfde studie mee te nemen. Ook wil ik voor Lee, 2004 de astmatische groep eruit halen, omdat dit

een specifieke groep is. Dat betekent dat een nieuwe analyse drie rijen aan informatie minder heeft. Het resultaat zien we hieronder wat betekent dat we een kleinere *odds-ratio* zien met mindere heterogeniteit én daardoor ook een statistisch significant 95% voorspellingsinterval (marginaal significant).

```

Review:      Glyphosate and NHL for the exposed group

              OR              95%-CI  %w(random)
Cantor,1992  1.2000 [1.0000; 1.5000]  42.2
De Roos,2003 2.1000 [1.1000; 4.0000]   6.6
De Roos,2003 1.6000 [0.9000; 2.8000]   8.4
Eriksson,2008 1.5100 [0.7700; 2.9400]   6.2
Harde11,2002 1.8500 [0.5500; 6.2000]   2.0
Lee (a),2004 1.4000 [0.9800; 2.1000]  16.9
McDuffie,2001 0.9200 [0.5400; 1.5500]   9.6
McDuffie,2001 1.8800 [1.0100; 3.2100]   8.1

Number of studies combined: k = 8

              OR              95%-CI  z  p-value
Random effects model 1.3542 [1.1406; 1.6077] 3.46 0.0005
Prediction interval  [1.0022; 1.8299]

Quantifying heterogeneity:
tau^2 = 0.0075 [0.0000; 0.2106]; tau = 0.0864 [0.0000; 0.4589]
I^2 = 0.9% [0.0%; 67.9%]; H = 1.00 [1.00; 1.76]

Test of heterogeneity:
  Q d.f. p-value
  7.06  7  0.4226

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 6)
  (with ad hoc correction)
    
```

We kunnen nu een aantal extra analyse doen rondom heterogeniteit en de betrouwbaarheid van de schattingen. Om te beginnen werk ik met een permutatietest, maar omdat sommige testen niet goed werken met het ene pakket en wel met het andere¹⁰⁷, krijg ik een uitkomst op de log-odds schaal en niet op de schaal van de *odds-ratio*. Dit is makkelijk verholpen door de exponent te nemen. We zien dan exact dezelfde resultaten, behalve voor de I^2 , wat betekent dat de *summary estimate* significant blijft. De onzekerheid rondom de I^2 is behoorlijk groot:

```

Random-Effects Model (k = 8; tau^2 estimator: REML)

  logLik deviance      AIC      BIC      AICC
  -1.1199  2.2399    6.2399    6.1317    9.2399

tau^2 (estimated amount of total heterogeneity): 0.0075 (SE = 0.0301)
tau (square root of estimated tau^2 value):      0.0865
I^2 (total heterogeneity / total variability):   11.43%
H^2 (total variability / sampling variability):   1.13

Test for Heterogeneity:
Q(df = 7) = 7.0606, p-val = 0.4226
    
```

¹⁰⁷ Dit probleem heb ik mijn eerdere rapporten al beschreven.

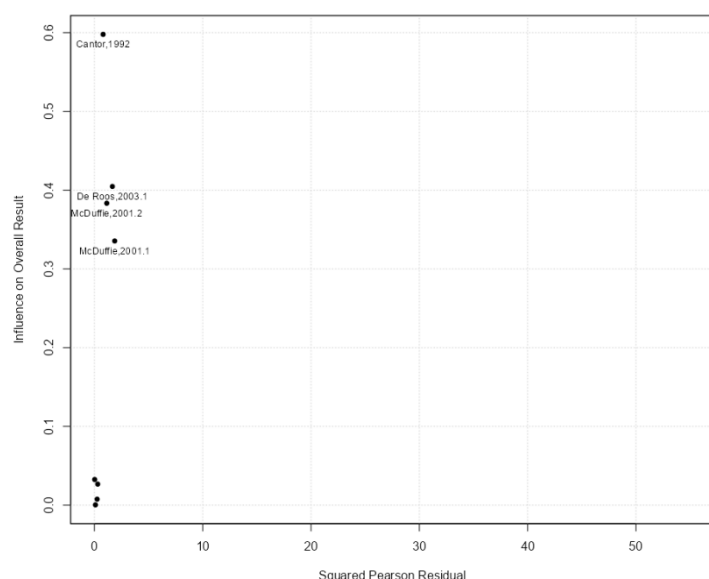
```

Model Results:
estimate      se      zval      pval      ci.lb      ci.ub      ***
0.3032      0.0876      3.4626      0.0005      0.1316      0.4748

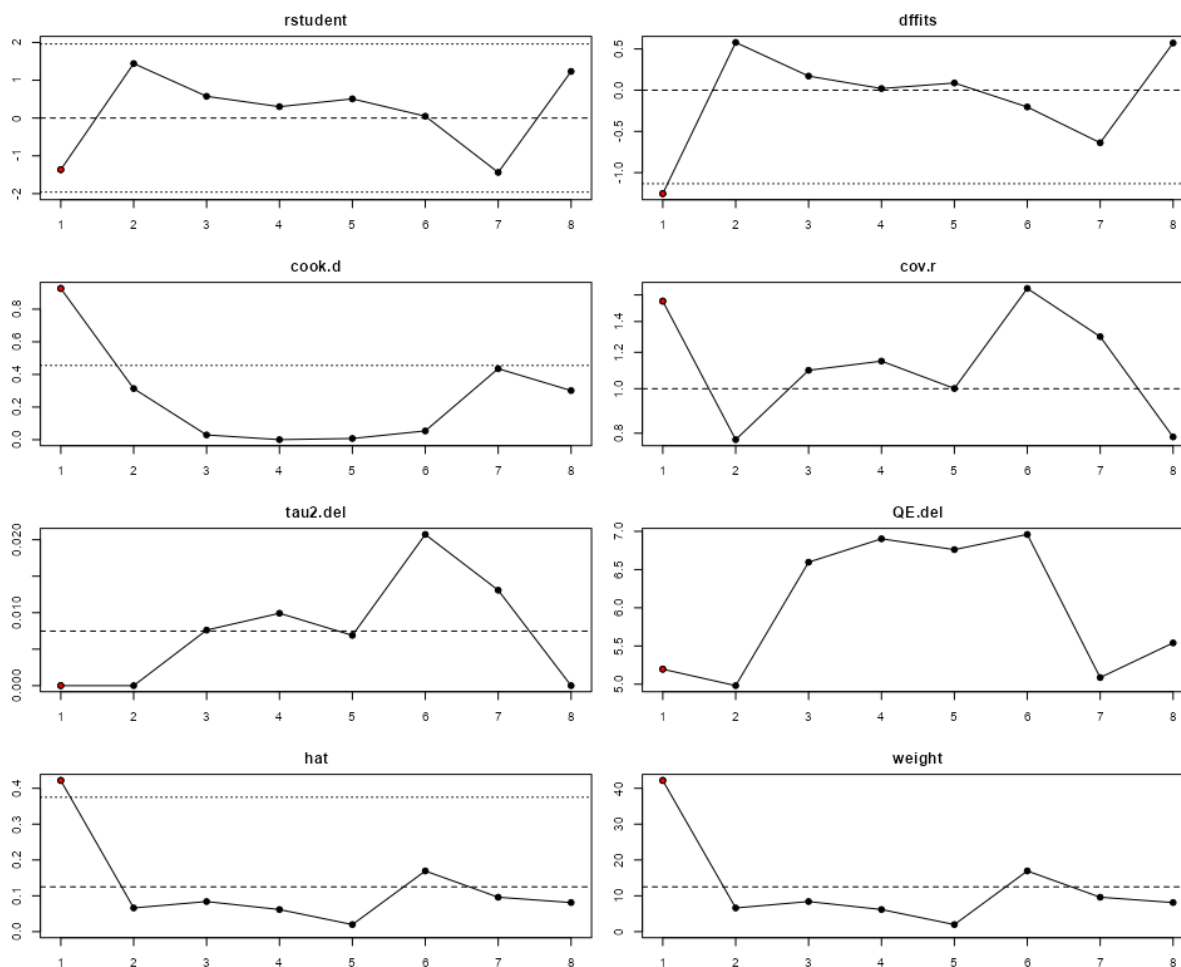
Model Results based on permutation testing:
estimate      se      zval      pval1      ci.lb      ci.ub      *
0.3032      0.0876      3.4626      0.0234      0.1316      0.4748

      estimate      ci.lb      ci.ub
tau^2      0.0075      0.0000      0.1938
tau      0.0865      0.0000      0.4402
I^2(%)      11.4326      0.0000      76.9902
H^2      1.1291      1.0000      4.3460
    
```

Figuur 72 en **Figuur 73** laten zien dat er minstens één schatting is die van invloed is op het geheel. Dit kon je eigenlijk ook al zien in **Figuur 71** omdat deze schatting het grootste gewicht heeft in de meta-analyse. Zonder die schatting zou het resultaat anders zijn. We kunnen nu nog verder kijken of er sprake is van een zogenaamde *publication bias*: de neiging om alleen datgene te publiceren wat één kant van het verhaal laat zien. Daarvoor passen we de Egger test toe en de *trimm-and-fill* procedure. En hoewel de Egger test aangeeft dat er geen indicatie is voor *publication bias*, wil ik toch graag zien wat de *trimm-and-fill* procedure op tafel neerlegt. Vanuit die procedure, wat uiteraard een simulatie blijft, krijgen we een schatting van 0.22 [0.082; 0.358] is. Oftewel, een *odds-ratio* van 1.246 [1.085; 1.430]. Oftewel, nog steeds statistisch significant. Ik verwacht dat het 95% voorspellingsinterval ook significant zal blijven.



Figuur 72. Baujat plot van de meta-analyse. De studies springen allemaal een beetje uit het veld, wat ook het gevolg is van het samenvoegen van weinig studies die ook nog eens verschillend zijn.



Figuur 73. Influence plot die laat zien welke van de 11 schattingen de grootste invloed heeft op het eindresultaat. Dat blijkt de schatting uit Cantor (57) te zijn wat we ook al zagen in **Figuur 72**.

```

Eggers' test of the intercept
=====
intercept      95% CI      t      p
      1.162 -0.08 - 2.4  1.839 0.12
Eggers' test does not indicate the presence of funnel plot asymmetry.
Estimated number of missing studies on the left side: 3 (SE = 1.8667)
Random-Effects Model (k = 11; tau^2 estimator: REML)

logLik deviance      AIC      BIC      AICc
-3.7614  7.5228  11.5228  12.1280  13.2371

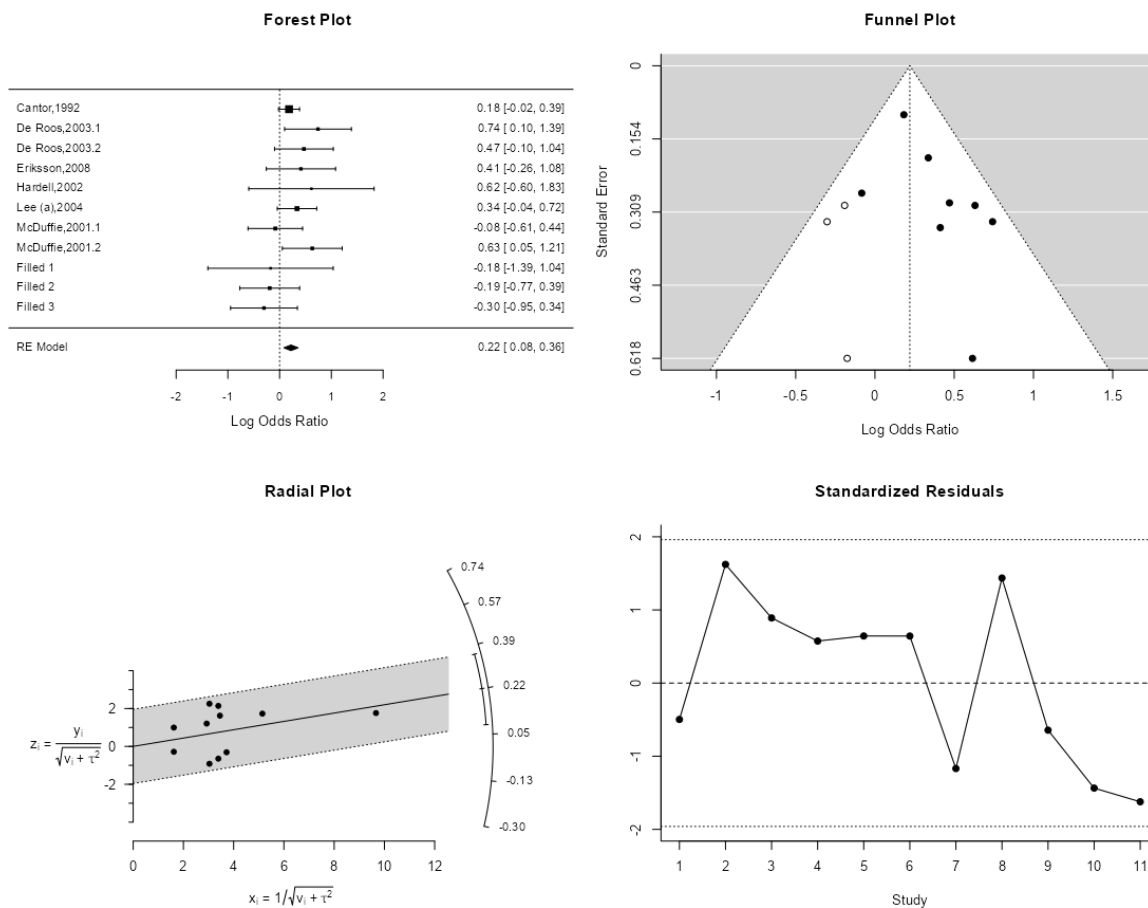
tau^2 (estimated amount of total heterogeneity): 0.0000 (SE = 0.0211)
tau (square root of estimated tau^2 value):      0.0028
I^2 (total heterogeneity / total variability):    0.01%
H^2 (total variability / sampling variability):    1.00

Test for Heterogeneity:
Q(df = 10) = 12.5556, p-val = 0.2496

Model Results:

estimate      se      zval      pval      ci.lb      ci.ub
0.2200  0.0706  3.1171  0.0018  0.0817  0.3583  **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```



Figuur 74. Diagnostische plots rondom de aanwezigheid van *publication bias*. De figuur rechtsboven toont de *funnel plot* voor *publication bias* met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de *publication bias* teniet te doen.

Meta-analyses voor NHL odds-ratio's en days of exposure

Nadat we een meta-analyse hebben gedaan voor het wel of niet blootgesteld zijn aan glyfosaat wil ik nog een meta-analyse doen waarbij ik kijk naar de mate van blootstelling. Zoals al aangegeven in de methode sectie (en ook in de resultaten sectie) is blootstelling een hele belangrijke parameter en we hebben eigenlijk té vaak al alles op één hoop gegooid. De vorige meta-analyse gaf al een sterker signaal voor het risico van glyfosaat voor NHL omdat we exclusief keken naar *odds-ratio's* en naar blootstelling (*exposure metric*). Ik wil nu kijken naar de mate van blootstelling (*exposure category*) om te zien om we wederom een signaal zien. Als we dan kijken naar de hoeveelheid beschikbare data dan zien we 18 schattingen, verdeeld over drie studies(60, 64, 76) waarvan één cohort studie(60):

```

study_design risk_metric n
<chr> <chr> <int>
1 Case-control OR 16
2 Cohort OR 2
    
```

```

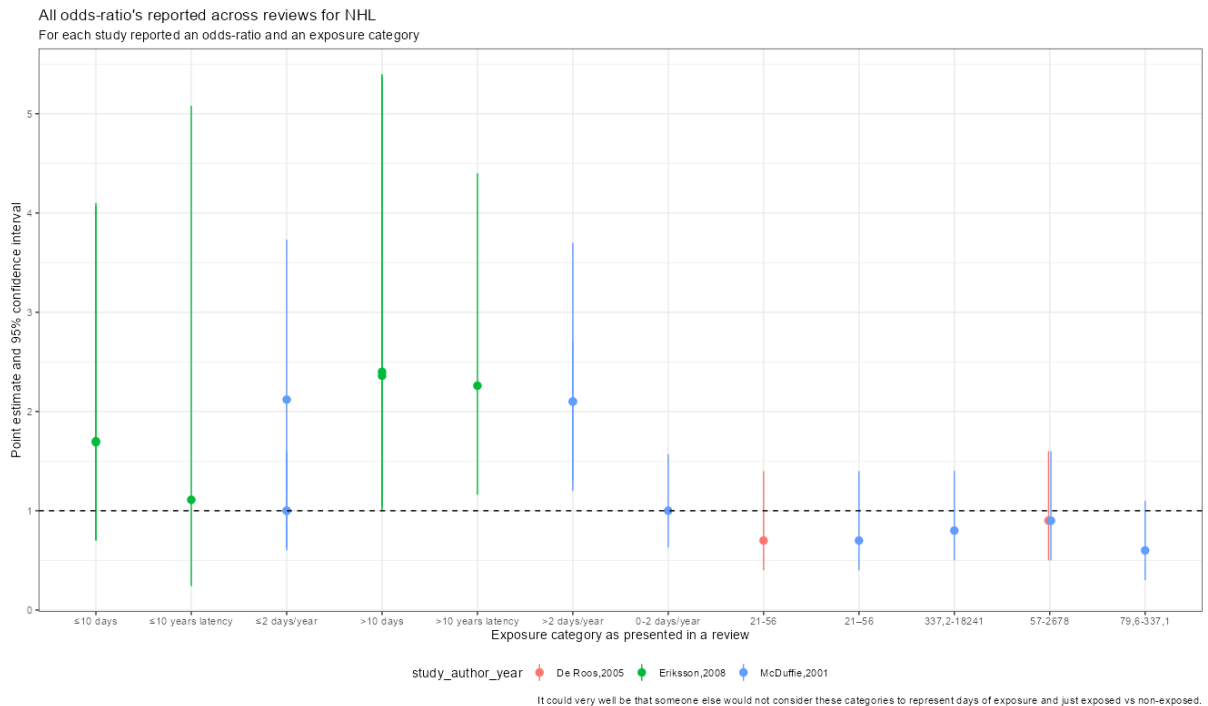
study_author_year risk_metric n
<chr> <chr> <int>
1 De Roos,2005 OR 2
2 Eriksson,2008 OR 6
3 McDuffie,2001 OR 10

# A tibble: 18 x 6
  study_author_year risk_metric risk_estimate risk_estimate_low risk_estimate_high exposure_category
  <chr> <chr> <dbl> <dbl> <dbl> <chr>
1 De Roos, 2005 OR 0.7 0.4 1.4 21-56
2 De Roos, 2005 OR 0.9 0.5 1.6 57-2678
3 Eriksson, 2008 OR 2.36 1.04 5.37 >10 days
4 Eriksson, 2008 OR 2.4 1 5.4 >10 days
5 Eriksson, 2008 OR 2.26 1.16 4.4 >10 years latency
6 Eriksson, 2008 OR 1.69 0.7 4.07 =10 days
7 Eriksson, 2008 OR 1.7 0.7 4.1 =10 days
8 Eriksson, 2008 OR 1.11 0.24 5.08 =10 years latency
9 McDuffie, 2001 OR 2.1 1.3 2.7 >2 days/year
10 McDuffie, 2001 OR 2.1 1.2 3.7 >2 days/year
11 McDuffie, 2001 OR 1 0.63 1.57 =2 days/year
12 McDuffie, 2001 OR 2.12 1.2 3.73 =2 days/year
13 McDuffie, 2001 OR 1 0.6 1.6 =2 days/year
14 McDuffie, 2001 OR 1 0.63 1.57 0-2 days/year
15 McDuffie, 2001 OR 0.7 0.4 1.4 21-56
16 McDuffie, 2001 OR 0.8 0.5 1.4 337, 2-18241
17 McDuffie, 2001 OR 0.9 0.5 1.6 57-2678
18 McDuffie, 2001 OR 0.6 0.3 1.1 79,6-337, 1

```

In **Figuur 75** zien we deze 18 schattingen afgebeeld op basis van de mate van blootstelling uitgedrukt in dagen, voor zover dat kan. Ook zien we voor elke schatting het 95% betrouwbaarheidsinterval en of dat interval de lijn van geen-verschil raakt. Bijna alle schattingen zijn niet significant. Omdat we hier eigenlijk te maken hebben met een dose-response relatie (108) is het uitvoeren van een meta-analyse niet zonder haar beperkingen. Zo wordt blootstelling vaak uitgedrukt in cumulatieve dagen (blootstelling).

¹⁰⁸ Althans voor de numerieke waarden die daadwerkelijk worden uitgedrukt in cumulatieve dagen blootstelling.



Figuur 75. Odds-ratio's voor non-hodgkin lymfoom (NHL) en de verschillende maten van blootstelling uitgedrukt in (cumulatieve) dagen.

De resultaten van de review zien we hieronder. Wat opvalt is dat we geen dubbele schattingen hebben. Wat ook opvalt is dat we maar over drie studies een uitspraak kunnen doen en dat we een uitspraak doen op basis van gelijkenissen in de gebruikte statistiek, ziekte en blootstelling waarbij de overeenkomst in blootstelling eigenlijk maar marginaal is (zie ook **Figuur 75**). Met andere woorden: we zijn nu echt studies aan het samenvoegen waarvan elke schatting hetzelfde is. Ook heb ik het vermoeden dat twee studies (60, 76) hetzelfde meten gezien hun overlap op sommige waardes in **Figuur 75**.

```

Review:      Glyphosate and NHL for the exposed group
      OR      95%-CI %w (random)
De Roos,2005 0.7000 [0.4000; 1.4000] 5.6
De Roos,2005 0.9000 [0.5000; 1.6000] 6.0
Eriksson,2008 2.3600 [1.0400; 5.3700] 4.3
Eriksson,2008 2.4000 [1.0000; 5.4000] 4.2
Eriksson,2008 2.2600 [1.1600; 4.4000] 5.3
Eriksson,2008 1.6900 [0.7000; 4.0700] 4.0
Eriksson,2008 1.7000 [0.7000; 4.1000] 4.0
Eriksson,2008 1.1100 [0.2400; 5.0800] 1.8
McDuffie,2001 2.1000 [1.3000; 2.7000] 7.9
McDuffie,2001 2.1000 [1.2000; 3.7000] 6.2
McDuffie,2001 1.0000 [0.6300; 1.5700] 7.1
McDuffie,2001 2.1200 [1.2000; 3.7300] 6.1
McDuffie,2001 1.0000 [0.6000; 1.6000] 6.8
McDuffie,2001 1.0000 [0.6300; 1.5700] 7.1
McDuffie,2001 0.7000 [0.4000; 1.4000] 5.6
McDuffie,2001 0.8000 [0.5000; 1.4000] 6.6
McDuffie,2001 0.9000 [0.5000; 1.6000] 6.0
McDuffie,2001 0.6000 [0.3000; 1.1000] 5.5
    
```

```

Number of studies combined: k = 18

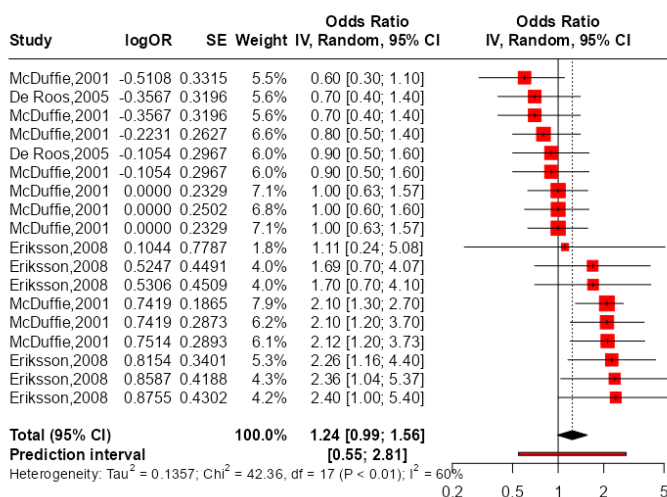
OR          95%-CI      z  p-value
Random effects model 1.2407 [0.9886; 1.5571] 1.86 0.0627
Prediction interval [0.5473; 2.8129]

Quantifying heterogeneity:
tau^2 = 0.1357 [0.0317; 0.4141]; tau = 0.3683 [0.1782; 0.6435]
I^2 = 59.9% [32.7%; 76.1%]; H = 1.58 [1.22; 2.04]

Test of heterogeneity:
Q d.f. p-value
42.36 17 0.0006

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp prediction interval (df = 16)
  (with ad hoc correction)
    
```

De vraag is nu wat we met het resultaat kunnen. Deze is niet significant en ook het 95% voorspellingsinterval is groot. Dat mag eigenlijk niet verbazen, omdat we schattingen met elkaar combineren die niet goed combineerbaar zijn. De I^2 statistiek toont dit ook en is met bijna 60% substantieel te noemen. **Figuur 76** toont de dose-response relatie waarbij een bepaalde mate van blootstelling zelfs beschermend lijkt te zijn. Er is dus een ruime spreiding in de relatie tussen glyfosaat en de kans op NHL. Ook heeft elke schatting een behoorlijk 95% betrouwbaarheidsinterval. Eigenlijk kunnen we met deze meta-analyse helemaal niks.



Figuur 76. Forest plot van een meta-analyse voor glyfosaat en non-hodgkin lymfoom (NHL) voor studies die de mate van blootstelling voor glyfosaat hebben gerapporteerd.

Het wordt nu tijd om de studies bij mensen af te sluiten en te kijken wat we kunnen vinden in de dierstudies.

Glyfosaat en kanker bij muizen en ratten

Het wordt nu tijd om aan de slag te gaan met de onderzoeksgegevens van de dierproeven. Zoals al beschreven in de methode sectie (en deels in de resultaten sectie) had ik niet verwacht dat ik geen enkele meta-analyse zou vinden voor pesticiden en kanker bij dieren. De enige meta-analyse¹⁰⁹ is de studie van Portier(1) en die studie heb ik uitgebreid beschreven in mijn [eerste rapport](#) zoals al aangegeven in de inleiding. Het feit dat ik dus alleen een meta-analyse kon vinden voor glyfosaat en kanker bij dieren, maakt ook dat ik me alleen ben gaan richten op glyfosaat en kanker bij mensen. Deze redenering heb ik al beschreven, maar ik vind het belangrijk vinden dat de lezer dit ook blijft onthouden. Het dossier pesticiden en kanker is dus helemaal niet zo eenvoudig en zomaar verwijzen naar losse (dier)studies is onwenselijk en ik denk zelfs schadelijk¹¹⁰.

Mijn zoektocht naar aanvullende dierstudies voor glyfosaat hebben letterlijk één nieuwe publicatie opgeleverd en dat is de veelbesproken Panzacchi studie (3) van het Ramazzini instituut. Ik kan dus nu twee dingen doen:

1. Ik beschrijf de studie van Panzacchi(3) in de wetenschap dat de meta-analyse van Portier(1) helemaal niet sterk was.
2. Ik voeg de studie van Panzacchi(3) toe aan de meta-analyse van Portier(1) om te zien of het toevoegen van die studie een verschil maakt in het eindresultaat. Met andere woorden: ik doe een update.

Misschien is het wel het beste als ik tracht om beiden te doen. Niet omdat het wetenschappelijk te verdedigen is, maar puur vanwege de media-aandacht voor de Panzacchi studie (3) en het feit dat de Portier meta-analyse ook letterlijk in de Panzacchi studie (3) wordt beschreven.

Laten we eerst maar eens kijken waar we mee te maken hebben en zien wat mogelijk vervolgstappen zijn. Zo weten we dat voor mensen alleen de gegevens rondom non-hodgkin lymfoom (NHL) robuust genoeg zijn om een uitspraak over te doen én dat het gevonden

¹⁰⁹ En die studie werd door ChatGPT niet eens direct opgemerkt als meta-analyse. Dit is waarom ik zelf op zoek ben gegaan naar alle studies die refereren naar de studie van Portier. Ook ben ik hierdoor de analyse van ChatGPT vaker én dieper gaan controleren dan ik origineel van plan was. Het blijkt dat je heel goed moet beschrijven aan AI wat je wil én er dan alsnog maar vanuit moet gaan dat er op je afkomt niet vaker niet dan wel klopt.

¹¹⁰ Tenzij je bij voorbaat overtuigd bent dat glyfosaat kanker veroorzaakt. Dan zul je er geen moeite mee hebben.

effect statistisch significant is, maar zeer klein. Het 95% voorspellingsinterval toon geen significant effect en laat veel onzekerheid zien.

Laten we daarom eerst maar eens beginnen met te beschrijven wat we hebben. Om inzicht te krijgen heb ik de studie tot me genomen én het supplement, en aan ChatGPT gevraagd om alle data te extraheren die ik daarna handmatig heb gecontroleerd. Wat direct opviel, en dat zal ik later uitgebreid beschrijven, is dat er voor deze studie een diverse hoeveelheid aan kankersoorten is onderzocht. Althans, er is een diverse hoeveelheid gevonden. Om dit enigszins te stoomlijnen met de rest van de dataset die ik op basis van de review van Portier heb ontwikkeld voor [mijn eerste rapport](#), heb ik de benamingen voor sommige kankersoorten aangepast. Dat betekent dat als in de meta-analyse van Portier(1) een bepaalde kankersoort op een bepaalde manier werd benoemd, en in de Panzacchi studie (3) men het over dezelfde kankersoort had deze net iets anders noemde, ik de beschrijving uit Portier heb overgenomen. Omdat de Panzacchi studie (3) veel meer kankersoorten heeft opgenomen is dat zeker niet voor alle beschrijvingen gelukt¹¹¹. We zullen straks zien hoe we hier mee om kunnen gaan.

Een ander belangrijk punt om te noemen is dat ik mij expliciet tot de stof glyfosaat heb beperkt en niet heb gekeken naar de producten die op de markt zijn, zoals RoundUp of RangerPro. Dit waren namelijk de andere twee groepen die zijn onderzocht, maar hun samenstelling bestaat uit meer dan glyfosaat. Ook werd er in de meta-analyse van Portier alleen maar gekeken naar glyfosaat. Ik ben dus, samengevat, alleen geïnteresseerd in de relatie tussen de stof glyfosaat en kanker¹¹². Laten we beginnen met daadwerkelijk te beschrijven wat de Panzacchi studie (3) te bieden heeft¹¹³.

De Panzacchi studie

Deze dierstudie uit 2025 heeft de relatie tussen glyfosaat en kanker onderzocht bij Sprague-Dawley ratten in de dosering van 0 (controle), 0.5, 5 en 50 mg/kg per lichaamsgewicht per

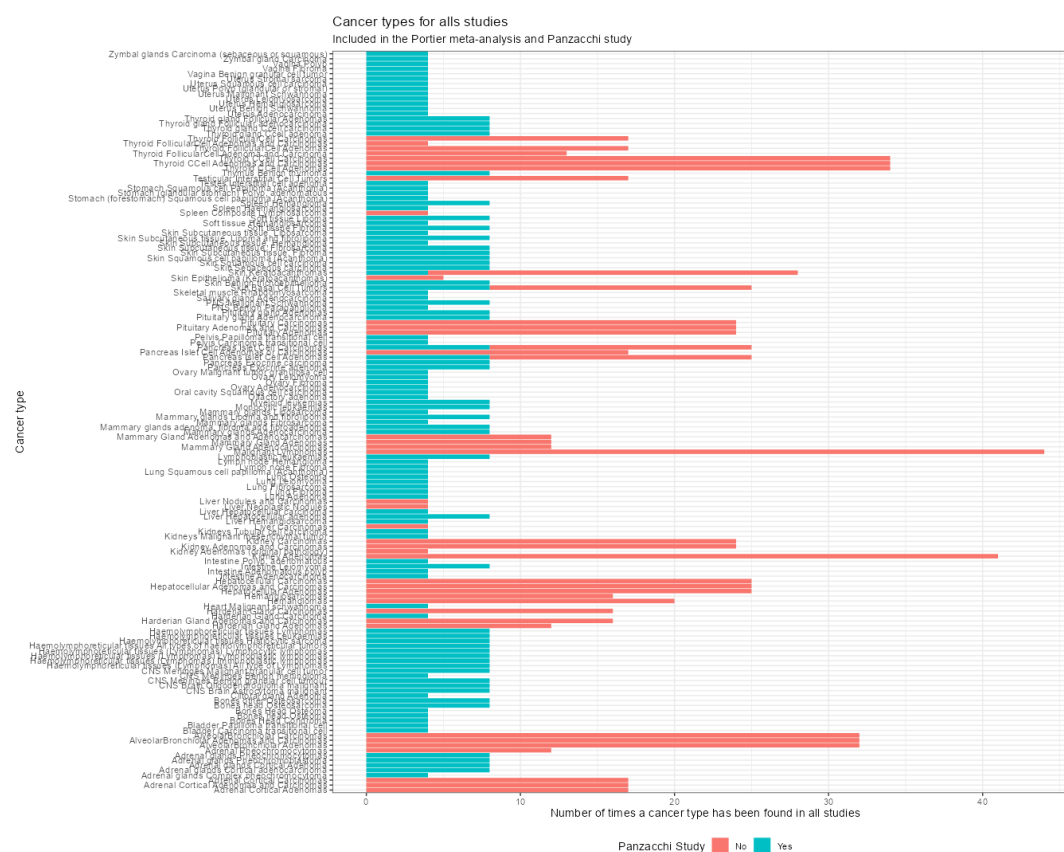
¹¹¹ Eigenlijk voor veel meer studies niet dan wel.

¹¹² Dat mensen in aanraking komen met meer dan puur glyfosaat, en dus wel met producten als RoundUp of RangerPro, maakt een direct vergelijk sowieso al lastig. Maar zou ik de data voor beide producten uit de Panzacchi studie nemen dan zou ik ze niet kunnen vergelijken met de data uit de Portier studie die naar puur glyfosaat keek. Zou ik de data voor de producten wel nemen en vergelijken met de data in studies onder mensen waarvan ik weet dat die expliciet naar dit product hebben gekeken, dan is het nog maar de vraag of er gebruik is gemaakt van dezelfde formule. En als de formule toch exact is, dan heb ik één dierstudie tegenover meerdere menselijke studies. Elke keuze laat dus wat te wensen over. Ik laat het aan de lezer om met de data en de codes zelf aan de slag te gaan, mocht die noodzaak zich aandienen.

¹¹³ Ik ga hier verder niet herhalen wat ik in mijn eerste rapport al beschreven heb. Dit rapport is al groot genoeg zoals het is, en voor veel van de technische aspecten van de analyse is het [eerste rapport](#) een betere plek om de context van mijn gemaakte keuzes hier te begrijpen. Daar waar ik van mening ben dat extra uitleg, of herhaling, wel zinvol is zal ik dat uiteraard ook bieden.

dag. Dit is per geslacht onderzocht waarbij in elke groep 51 dieren zijn geïncludeerd. Elke diergroep is 104 weken (2 jaar) gevolgd. Dit betekent dat er voor glyfosaat in totaal 408 dieren zijn geïncludeerd¹¹⁴. Deze studie lijkt in dat opzicht heel veel op de andere dierproeven die ik in de meta-analyse van Portier gezien heb¹¹⁵. Het meest belangrijke verschil tussen beide studies ligt waarschijnlijk in de aantallen kankersoorten die zijn ontdekt. Zo zagen we in de meta-analyse van Portier al dat de studies verschilden in welke kankersoorten werden geobserveerd en dat de manier waarop Portier hiermee is omgegaan heeft geleid tot methodologische problemen, zoals het niet meenemen van de afwezigheid van een bepaalde soort kanker.

Een snelle analyse laat zien dat in de Panzacchi studie(3) maar liefst 109 verschillende tumorsoorten zijn meegenomen¹¹⁶. Dat is vele malen meer dan de 44 tumoren die in totaal in de Portier meta-analyse zijn meegenomen. Dit kunnen we laten zien via **Figuur 77**.



Figuur 77. Alle soorten kanker die zijn opgenomen in de review van Portier en in de Panzacchi studie. Duidelijk is het verschil in de hoeveelheid soorten kanker.

¹¹⁴ Op basis van vier groepen (controle, 0,5, 5 en 50 mg/kg), twee verschillende geslachten en 51 dieren per groep*geslacht = 4*2*51 = 408 dieren.

¹¹⁵ Het blijkt dat de studiegrooite en de lengte duur dé standaard is voor dierproeven in dit veld.

¹¹⁶ Verschillend voor zover ik het kan beoordelen. Ik ben geen oncoloog, of bioloog, en baseer mijn berekening op datgene wat er in de studie vermeldt staat.

Als ik in de meta-analyse van Portier duik dan lees ik het volgende wat wellicht het verschil kan verklaren:

Reanalyses are conducted on all primary tumors where there are at least 3 tumors in all of the animals in a sex/ species/strain combination (regardless of dosing). In addition, any tumor where a positive finding ($p \leq 0.05$, one-sided C-A trend test) is seen in at least one study is also evaluated, regardless of number of animals with the tumor, in all studies of the same sex/species/strain. When adenomas and carcinomas are seen in the same tissue, a combined analysis of adenomas and carcinomas is also conducted.

Dit zou betekenen dat Portier al een selectie heeft gemaakt om te bepalen welke tumorsoorten er wel of niet worden meegenomen in de analyse. De Panzacchi studie lijkt deze *a priori* keuze niet te hebben gemaakt¹¹⁷ en als dat zo is dan betekent dit dat we de getallen ook iets anders moeten gaan interpreteren. Het betekent ook dat de Panzacchi studie toevoegen aan de Portier meta-analyse haar uitdagingen kent. En het zou natuurlijk verklaren waarom ik in de Panzacchi studie meer dan twee keer zoveel kankersoorten zie dan in de Portier analyse. Maar voordat ik met deze data en getallen verder ga rekenen, wil ik zeker weten dat ik hier geen fout maak. Daarom neem ik voor nu alle data tot me voor de groep met de grootste blootstelling (50 mg/kg bw¹¹⁸/dag) om te zien welke soorten kanker er allemaal zijn ontdekt¹¹⁹. Als eerste kijk ik naar de mannetjes en zie het volgende:

```
# A tibble: 25 x 3
  Tumor                                N Cases
  <chr>                                <dbl> <dbl>
1 Adrenal glands Pheochromocytomas      51     4
2 Bones Head Condroma                   51     1
3 Bones Head Osteoma                     51     1
4 CNS Meninges Malignant granular cell tumor 51     2
5 Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas 51     1
6 Haemolymphoreticular tissues All types of haemolymphoreticular tumors 51     8
7 Haemolymphoreticular tissues Histiocytic sarcoma 51     1
8 Haemolymphoreticular tissues Leukaemias 51     1
9 Haemolymphoreticular tissues Lymphomas 51     6
10 Liver Hepatocellular carcinoma        51     2
11 Lung Osteoma                           51     1
12 Lung Squamous cell papilloma (Acanthoma) 51     1
13 Lymphoblastic leukaemias               51     1
14 Mammary glands adenoma, fibroma and fibroadenoma 51     1
15 Mammary glands Lipoma and fibrolipoma 51     1
16 Pancreas Islet Cell Adenomas          51     1
17 Pituitary gland Adenomas               51    16
18 Pituitary gland Adenocarcinoma        51     2
19 Skin Keratoacanthomas                  51     1
20 Skin Squamous cell carcinoma           51     1
21 Skin Squamous cell papilloma (Acanthoma) 51     2
22 Skin Subcutaneous tissue, Fibroma      51     1
23 Stomach (forestomach) Squamous cell papilloma (Acanthoma) 51     2
24 Testes Interstitial cell adenoma       51     1
25 Thyroid gland Follicular adenocarcinoma 51     1
```

¹¹⁷ Ik ga er gemakshalve maar vanuit dat de Portier keuze *a priori* gemaakt is.

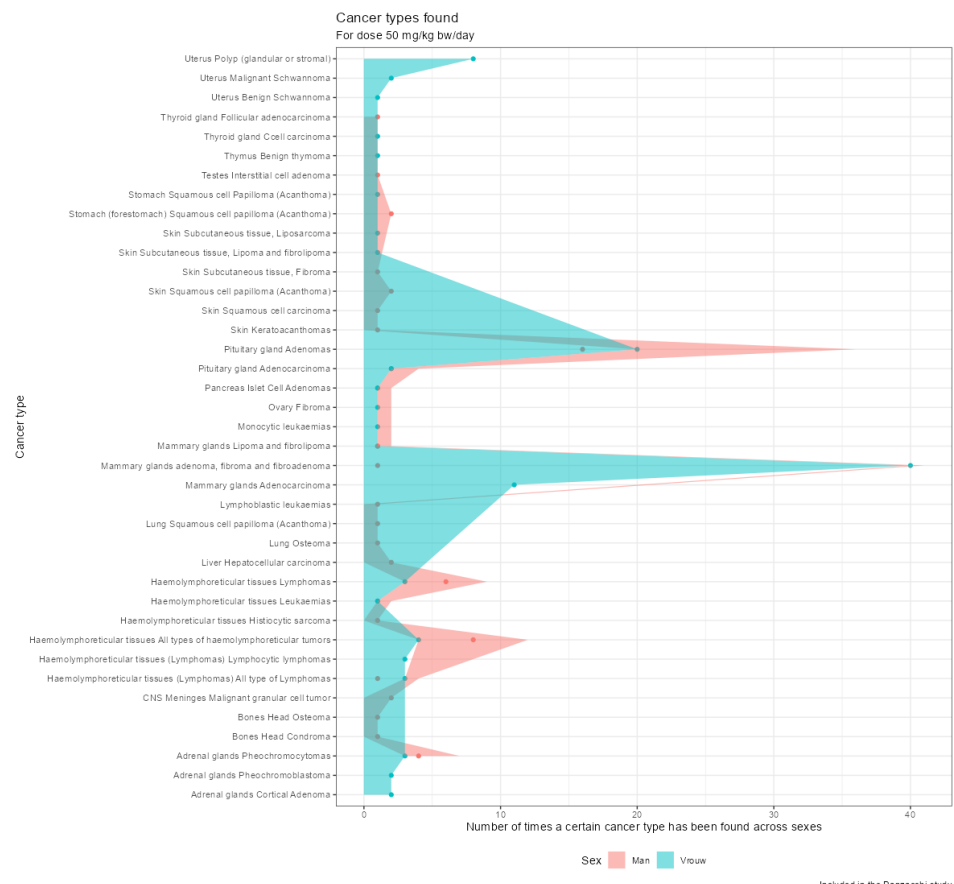
¹¹⁸ Bodyweight (bw)

¹¹⁹ Ik kies deze groep *a priori* uit omdat ik verwacht dat in deze groep de grootste aantallen kanker te vinden zijn. Ook verwacht ik dat elke kankersoort zich laat zien in deze groep.

We zien 25 soorten kanker met als uitschieter het aantal tumoren in de hypofyse. Als we kijken bij de vrouwtjes zien we het volgende:

```
# A tibble: 24 x 3
  Tumor                                N Cases
  <chr>                                <dbl> <dbl>
1 Adrenal glands Cortical Adenoma      51     2
2 Adrenal glands Pheochromoblastoma    51     2
3 Adrenal glands Pheochromocytomas     51     3
4 Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas 51     3
5 Haemolymphoreticular tissues (Lymphomas) Lymphocytic lymphomas 51     3
6 Haemolymphoreticular tissues All types of haemolymphoreticular tumors 51     4
7 Haemolymphoreticular tissues Leukaemias 51     1
8 Haemolymphoreticular tissues Lymphomas 51     3
9 Mammary glands Adenocarcinoma        51    11
10 Mammary glands adenoma, fibroma and fibroadenoma 51    40
11 Mammary glands Lipoma and fibrolipoma 51     1
12 Monocytic leukaemias                 51     1
13 Ovary Fibroma                        51     1
14 Pancreas Islet Cell Adenomas         51     1
15 Pituitary gland Adenocarcinoma       51     2
16 Pituitary gland Adenomas             51    20
17 Skin Subcutaneous tissue, Lipoma and fibrolipoma 51     1
18 Skin Subcutaneous tissue, Liposarcoma 51     1
19 Stomach Squamous cell Papilloma (Acanthoma) 51     1
20 Thymus Benign thymoma                51     1
21 Thyroid gland Ccell carcinoma        51     1
22 Uterus Benign Schwannoma              51     1
23 Uterus Malignant Schwannoma          51     2
24 Uterus Polyp (glandular or stromal)  51     8
```

Dat is een hele andere verdeling. Als we beide lijsten grafisch combineren krijgen we **Figuur 78** wat laat zien dat vrouwtjes en mannetjes andere tumoren ontwikkelen.



Figuur 78. Aantal kankersoorten zoals gezien in de Panzacchi studie bij mannetjes en vrouwtjes voor de 50 mg/kg dosering.

Uiteindelijk lijkt dus alsof de grote hoeveelheden tumoren die we zien geen fout is in de data. Wel lijkt het alsof ik, met een specifieke selectie voor de groep met de hoogste glyfosaat concentratie, zeker niet alle kankersoorten tegenkom. Het kan dus haast niet anders dan er ook kankersoorten zijn gezien in de lagere doseringen, maar daar wil ik straks wat specifiek naar kijken. Eerst wil ik nog wat tijd spenderen aan de verschillen en overeenkomsten tussen de Panzacchie studie en de Portier meta-analyse.

Het verschil tussen beide studies maakt namelijk dat ik niet zomaar alle 'nieuwe' kankersoorten uit de Panzacchi studie (3) kan samenvoegen met de kankersoorten uit de Portier analyse (of met andere kankersoorten uit de Panzacchi studie zelf). Zo heb ik al eerder genoemd dat je per kankersoort moet beoordelen of er een verschil is, en dat is exact wat Portier ook gedaan heeft. Wat **Figuur 77** verder laat zien is dat van sommige kankersoorten uit Panzacchi (3) de beschrijving overeenkomt met die van Portier, maar voor het gros is dat niet het geval. Dit maakt dat we van een grote groep de gegevens moeilijk kunnen combineren, tenzij we de proberen om de benamingen over te nemen van één van de studies. Dit is iets waar ik in een later stadium nog een keer naar wil (en moet) kijken.

Maar voordat ik begin met combineren wil ik eerst nog een aantal andere zaken bekijken. Zo zou het zinvol zijn om te achterhalen of de Portier meta-analyse, gecombineerd met de Panzacchi studie(3), überhaupt dezelfde tumorsoorten heeft ontdekt als die we in de menselijke studies hebben gezien. Als dit niet het geval zou zijn, kunnen we net zo goed stoppen: er is dan geen manier om de data van dierstudies af te zetten tegen menselijke studies.

In de studies onder mensen hebben gezien dat daar eigenlijk alleen maar genoeg gegevens zijn voor non-hodgkin lymfoom (NHL), beenmergkanker (MM), hairy cell leukemie (HCL), en nierkanker bij kinderen (Wilm's tumor). Omdat wij geen dieren zijn hoeft je ook niet te verwachten dat dezelfde kanker bij muizen eenzelfde naam kent als die bij mensen. Zo hoeven we niet te verwachten dat men bij ratten NHL kent wat maakt dat we deze in de lijst inderdaad niet tegenkomen. Zo heet NHL bij ratten en muizen T-cel lymfoom. Maar als ik zoek naar alle lymfomen die zijn meegenomen, dan zie ik de volgende lijst:

Tumor	n
1 Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas	1
2 Haemolymphoreticular tissues (Lymphomas) Immunoblastic lymphomas	1
3 Haemolymphoreticular tissues (Lymphomas) Lymphoblastic lymphomas	1
4 Haemolymphoreticular tissues (Lymphomas) Lymphocytic lymphomas	1
5 Haemolymphoreticular tissues All types of haemolymphoreticular tumors	1
6 Haemolymphoreticular tissues Histiocytic sarcoma	1
7 Haemolymphoreticular tissues Leukaemias	1
8 Haemolymphoreticular tissues Lymphomas	1

9	Lymph node Fibroma	1
10	Lymph node Hemangioma	1
11	Lymphoblastic leukaemias	1
12	Malignant Lymphomas	1
13	Spleen Composite Lymphosarcoma	1

Dit maakt een vergelijking helemaal niet zo makkelijk, maar we zouden onze hoop kunnen vestigen op *malignant lymphomas*: kwaadwaardige tumoren in de lymfeklieren. Zoek ik verder specifiek naar alle kankersoorten die verwijzen naar 'cell' (zoals *hairy cell leukemia*), dan krijg ik een langere lijst, maar zie geen HCL:

```
# A tibble: 36 x 2
  Tumor                                n
  <chr>                                <int>
1 Bladder Carcinoma transitional cell 1
2 Bladder Papilloma transitional cell 1
3 CNS Meninges Benign granular cell tumour 1
4 CNS Meninges Malignant granular cell tumor 1
5 Hepatocellular Adenomas 1
6 Hepatocellular Adenomas and Carcinomas 1
7 Hepatocellular Carcinomas 1
8 Kidneys Tubular cell carcinoma 1
9 Liver Hepatocellular adenoma 1
10 Liver Hepatocellular carcinoma 1
11 Lung Squamous cell papilloma (Acanthoma) 1
12 Oral cavity Squamous cell carcinoma 1
13 Ovary Malignant tumor granulosa cell 1
14 Pancreas Islet Cell Adenomas 1
15 Pancreas Islet Cell Adenomas or Carcinomas 1
16 Pancreas Islet Cell Carcinomas 1
17 Pelvis Carcinoma transitional cell 1
18 Pelvis Papilloma transitional cell 1
19 Skin Basal Cell Tumors 1
20 Skin Squamous cell carcinoma 1
21 Skin Squamous cell papilloma (Acanthoma) 1
22 Stomach (forestomach) Squamous cell papilloma (Acanthoma) 1
23 Stomach Squamous cell Papilloma (Acanthoma) 1
24 Testes Interstitial cell adenoma 1
25 Testicular Interstitial Cell Tumors 1
26 Thyroid CCell Adenomas 1
27 Thyroid CCell Adenomas and Carcinomas 1
28 Thyroid CCell Carcinomas 1
29 Thyroid FollicularCell Adenoma and Carcinoma 1
30 Thyroid FollicularCell Adenomas 1
31 Thyroid FollicularCell Adenomas and Carcinomas 1
32 Thyroid FollicularCell Carcinomas 1
33 Thyroid gland Ccell adenoma 1
34 Thyroid gland Ccell carcinoma 1
35 Uterus Squamous cell carcinoma 1
36 Vagina Benign granular cell tumor 1
```

Ga ik zoeken naar leukemie, beenmergkanker en nierkanker¹²⁰, dan krijg ik de volgende resultaten, respectievelijk:

```
# A tibble: 4 x 2
  Tumor                                n
  <chr>                                <int>
1 Haemolymphoreticular tissues Leukaemias 1
2 Lymphoblastic leukaemias 1
3 Monocytic leukaemias 1
4 Myeloid leukemias 1

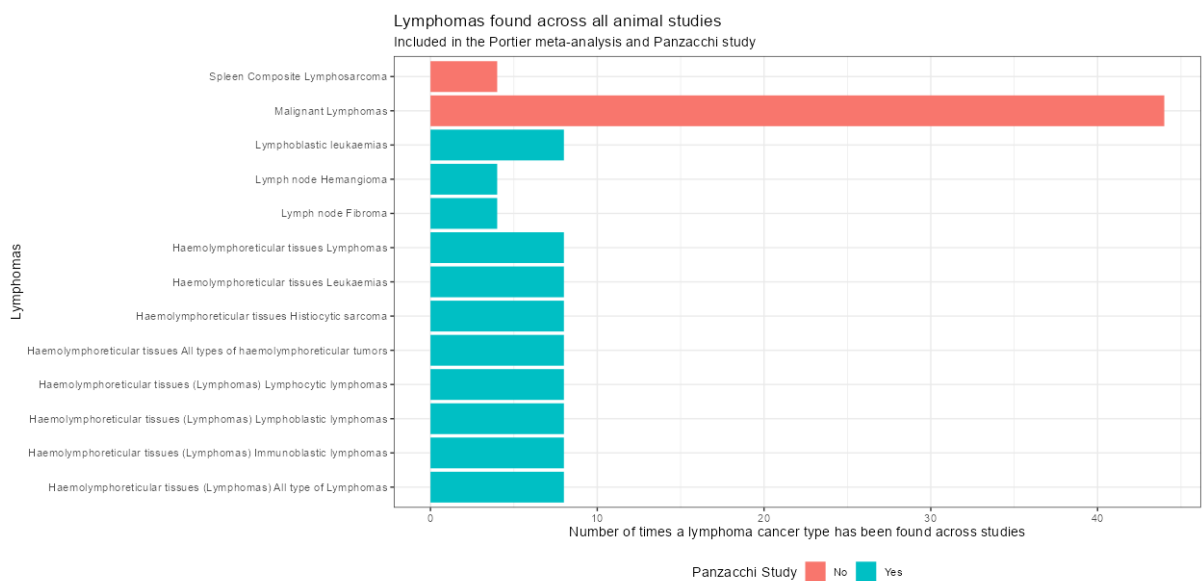
# A tibble: 0 x 2
# 2 variables: Tumor <chr>, n <int>

# A tibble: 6 x 2
  Tumor                                n
  <chr>                                <int>
```

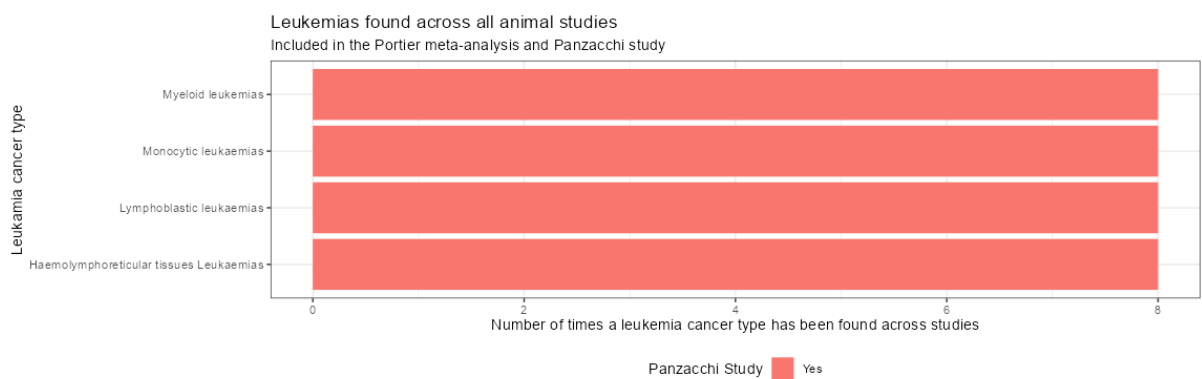
¹²⁰ Hoewel de Latijnse benaming voor nierkanker *renal* is, kreeg ik niks passends bij die term. Daarom heb ik gezocht op *kidney*. Voor beenmergkanker (*multiple myeloma*) kon ik niks vinden. Als ik op *hairy* zoek vind ik ook niks.

1	Kidney Adenomas	1
2	Kidney Adenomas (original pathology)	1
3	Kidney Adenomas and Carcinomas	1
4	Kidney Carcinomas	1
5	Kidneys Malignant mesenchymal tumor	1
6	Kidneys Tubular cell carcinoma	1

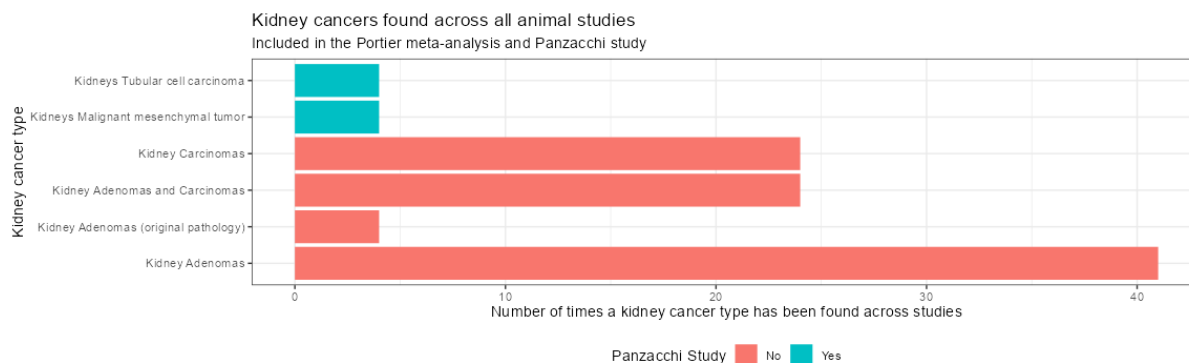
Met deze bevindingen kunnen we nu gaan kijken waar (in Portier(1) of Panzacchi (3)) we resultaten voor lymfoom (**Figuur 79**), leukemie (**Figuur 80**) en nierkanker (**Figuur 81**) vinden. Het blijkt dat het gros van de soorten lymfoomkanker zich in de Panzacchi studie bevinden, leukemie in Portier niet voorkomt, en dat ook voor nierkanker er meer soorten worden gezien in de Panzacchi studie.



Figuur 79. Soorten lymfoom kanker en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.



Figuur 80. Soorten kanker die te maken hebben met leukemie en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.



Figuur 81. Soorten nierkanker en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.

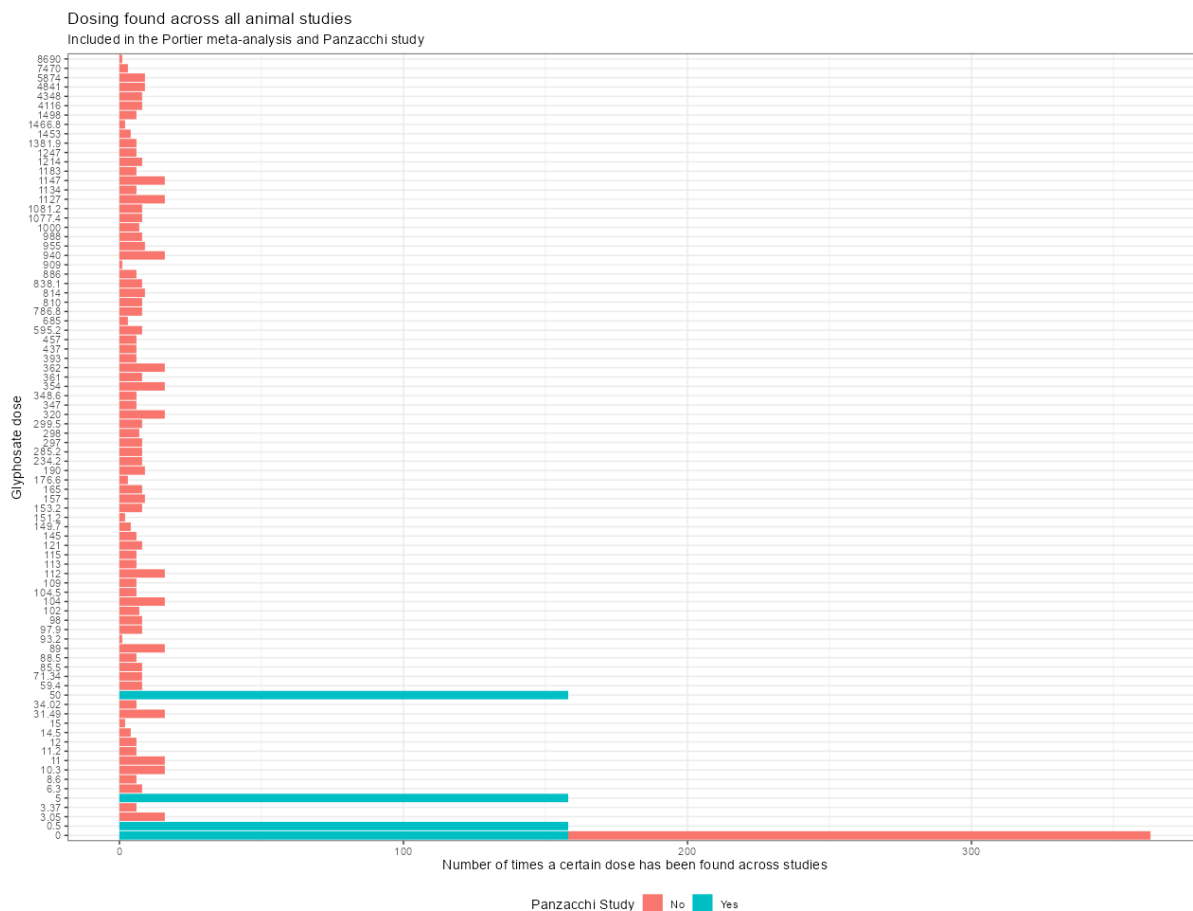
Het lijkt er dus toch echt op dat de Panzacchi studie veel meer verschillende kankersoorten rapporteert, maar onduidelijk is nog in hoeverre er ook echt meer kanker is gezien. Daar komen we straks op terug.

Verder ben ik geïnteresseerd in de dosering van de Panzacchi studie en waar die valt tussen de rest van de doseringen zoals opgenomen in de meta-analyse van Portier (**Figuur 82**). Ook wil ik zien hoeveel data we nu hebben per diersoort (**Figuur 83**).

Steeds duidelijker wordt hoeveel data de Panzacchi studie toevoegt¹²¹. Ik wil dus nu toch dieper gaan duiken in de aantallen om te ontdekken hoeveel kanker er in totaal is geïdentificeerd bij alle dieren die zijn blootgesteld aan glyfosaat én hoeveel kanker we zien in de controle groep. Ik wil me dus voor nu blijven richten op de Panzacchi studie omdat deze een stuk afwijkt ten opzichte van de andere studies die zijn opgenomen in de Portier meta-analyse. Later zal ik weer terugkomen bij de mogelijke overlap met de Portier analyse.

Ik heb al eerder beschreven dat het onderzoek naar glyfosaat in principe 408 dieren telt. En omdat één dier meerdere soorten kanker kan ontwikkelen, of wellicht meerdere tumoren per kankersoort heeft, kan het heel goed zijn dat de totale hoeveelheid kanker (veel) meer is dan 408. Maar het kan ook goed zijn dat het (veel) minder is.

¹²¹ Of er ook echt meer informatie wordt toegevoegd moet nog blijken, maar in aantallen is het een fikse toevoeging.



```
# A tibble: 4 x 2
  Dosering sum
  <dbl> <dbl>
1 0 179
2 0.5 181
3 5 168
4 50 174

# A tibble: 8 x 3
# Groups:   Dosering [4]
  Dosering Geslacht sum
  <dbl> <chr> <dbl>
1 0 Man 58
2 0 Vrouw 121
3 0.5 Man 64
4 0.5 Vrouw 117
5 5 Man 56
6 5 Vrouw 112
7 50 Man 60
8 50 Vrouw 114
```

Er zijn in totaal 702 gevallen van kanker ontdekt waarvan ongeveer net zoveel in de controle groep als in de glyfosaat groepen (behalve de 5 mg/kg bw/dag die minder tumoren laat zien). Een kleine rekensom zegt ons dat de hoeveelheid tumoren 1.7 keer zo groot is als het aantal proefdieren (onder de aanname dat elk dier maar één tumor kan krijgen). Ook zien we per dosering 1.75, 1.77, 1.65 en 1.71 keer zoveel kanker als dat we ratten hebben¹²².

Ik ben verder benieuwd naar het soort kanker dat we zien bij elke dosering. Hoewel de aantallen, in totaal, niet lijken te verschillen (met een uitschieter naar beneden bij de 5mg/kg dosering) zou een afwijken patroon een interessant gegeven zijn. Is het bijvoorbeeld zo dat we sommige kankers helemaal niet zien in de controlegroep, maar wel bij glyfosaat (waarbij er wellicht sprake is van een *dose-response* relatie), dan hebben we iets op ons op te gaan richten, maar wel met gepaste voorzichtigheid¹²³.

We weten al dat we 109 unieke tumoren zien in de Panzacchi studie. Een snelle analyse toont ons dat we 74 soorten tumoren vinden bij de mannetjes en 84 soorten tumoren bij de vrouwtjes¹²⁴. Ik ben nu benieuwd voor hoeveel van deze tumoren er geen casus werd gevonden in de controlegroep¹²⁵. Een snelle analyse laat zien dat van de 74 tumoren bij mannetjes en 84 tumoren bij vrouwtjes er in 56 en 55 kankergevallen,

¹²² Uiteraard is het een beetje gek om op deze manier de hoeveelheid kanker af te zetten tegen de hoeveelheid muizen. Zeker als je bedenkt dat de controle groep in feite meer kanker laat zien dan de 5mg/kg en 50 mg/kg dosering. Ook dit is een constatering die mij in het [eerste rapport](#) al deed fronsen.

¹²³ Met een zo grote hoeveelheid kanker en zoveel (mogelijke) analyses is het vinden van vals-positieven sowieso een groot probleem. Wederom iets wat ik mijn eerste rapport uitvoerig heb beschreven.

¹²⁴ Dat zijn 158 soorten tumoren bij elkaar geteld, wat betekent dat er overlap moet zijn. Hoeveel, en vooral voor welke soort tumor, is niet direct duidelijk.

¹²⁵ In principe verwacht ik dat er voor elke genoemde kankersoort een casus wordt gevonden anders zou dit betekenen dat er ook kankersoorten worden genoemd die niet zijn gevonden in de studie. Dat zou op zich niet verkeerd zijn omdat het zou betekenen dat de lijst van kankersoorten *a priori* is opgesteld.

respectievelijk, geen tumoren werden gevonden in de controle groep. Maar dit betekent nog niet dat er wel (of geen) tumoren werden gevonden in de doseringsgroepen. Want als ik kijk voor hoeveel tumoren er in geen enkele dosering iets werd gevonden, dan zien we dat voor 24 en 30 tumoren, respectievelijk, geen enkele bevinding werd gedaan (zie ook **Figuur 88**). We kunnen nu langzaam een tabel gaan opbouwen (**Tabel 4**):

Vraag om te beantwoorden	Mannetjes	Vrouwtjes
Aantal unieke tumoren beschreven in de Panzacchi studie?	74	84
Aantal unieke tumoren met geen casus in de controlegroep?	56	55
Aantal unieke tumoren met geen casus in alle groepen?	24	30

Tabel 4. Gevallen per kankersoort en geslacht gezien in de Panzacchi studie.

De bevindingen hierboven zijn al bijzonder te noemen, want nu blijkt dat de grote hoeveelheid kankersoorten die is meegenomen zeker niet allemaal zijn gesignaleerd in de studie. De eerstvolgende vraag die dan bij mij naar boven schiet is de vraag wat er verandert in de hoeveelheden als we de controlegroep vergelijken met de glyfosaatgroepen. Ik verwacht daarbij dat, vergeleken met de controlegroep, elke glyfosaatgroep een grotere hoeveelheid tumoren laat zien¹²⁶. Dit kunnen we grafisch het best laten zien. Zo toont **Figuur 84** dat het zeker niet zo is dat het aantal tumoren in elke glyfosaatgroep groter is dan de controle groep. Dit kunnen we ook zien in **Figuur 85** en nog veel beter in **Figuur 86**. Waar **Figuur 85** voor elke tumorgroep het pad laat zien van het verschil ten opzichte van de controle groep, geeft **Figuur 86** met de dikte van de lijnen ook aan hoe vaak dit gebeurt (de incidentie). We kunnen nu duidelijk zien dat voor het gros van de tumoren er geen verschil zit tussen de glyfosaatgroepen en de controlegroep (wat waarschijnlijk betekent dat er helemaal geen tumoren zijn gezien). Ook zien we dat het net zo goed kan zijn dat we meer tumoren vinden in de glyfosaatgroepen als dat het kan zijn dat we minder tumoren vinden.

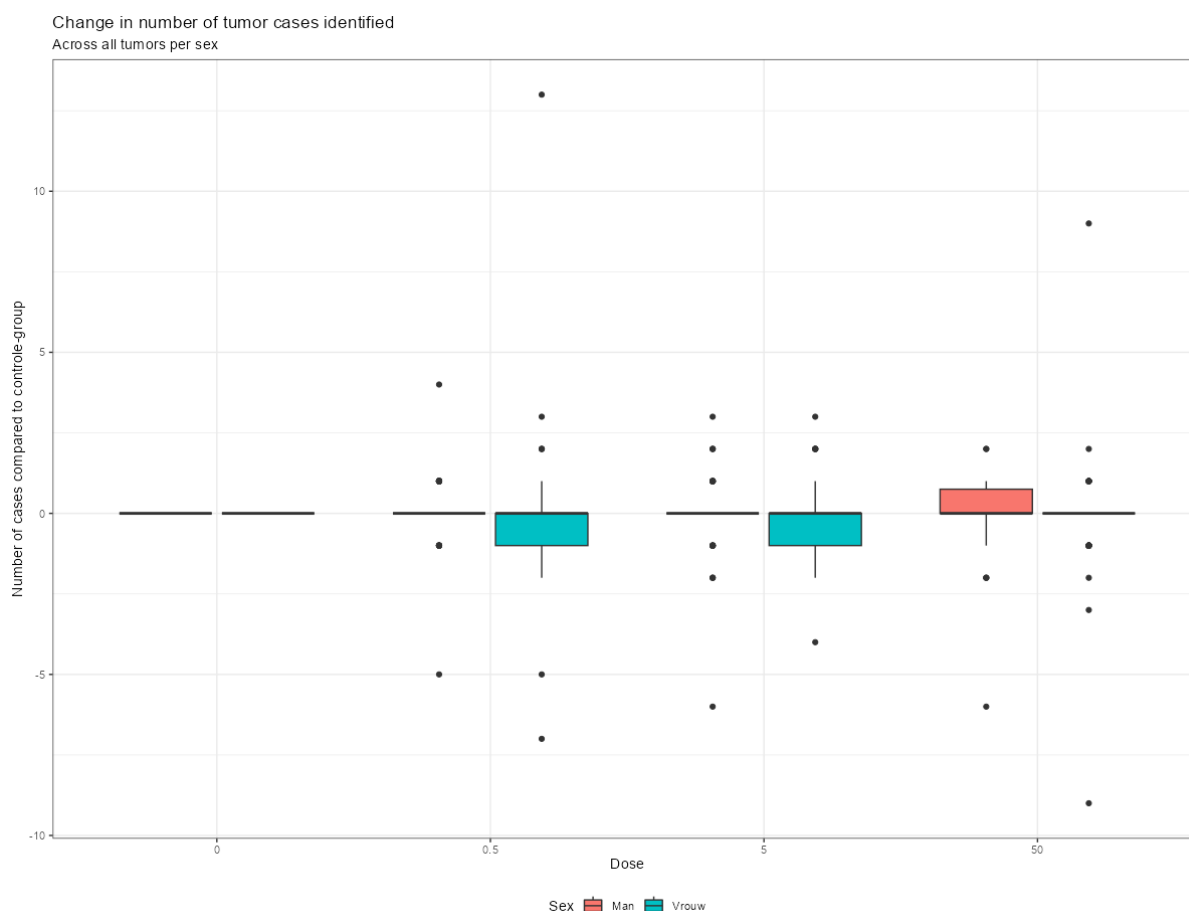
Wat wel duidelijk is, is dat er meer verschillende paden zijn bij vrouwtjes dan bij mannetjes, met een flinke uitschieter van 13 extra tumoren in de glyfosaatgroep. Dit blijkt in de studie te worden gebeuren voor *Mammary glands adenoma, fibroma and fibroadenoma*. Als ik deze tumorsoort selecteer voor alle doseringen en geslacht, dan zie ik het volgende:

```
# A tibble: 8 x 6
# Groups:   Geslacht, Tumor [2]
  Geslacht Tumor          Dosering    N Cases change
  <chr>    <chr>          <dbl> <dbl> <dbl> <dbl>
1 Vrouw  Mammary glands adenoma, fibroma and fibroadenoma 0 51 31 0
2 Vrouw  Mammary glands adenoma, fibroma and fibroadenoma 0.5 51 44 13
3 Vrouw  Mammary glands adenoma, fibroma and fibroadenoma 5 51 32 1
```

¹²⁶ Maar weet dus nu dat dit niet kan, omdat voor sommige tumorsoorten er geen enkele casus werd gezien in geen enkele groep.

4	Vrouw	Mammary glands adenoma, fibroma and fibroadenoma	50	51	40	9
5	Man	Mammary glands adenoma, fibroma and fibroadenoma	0	51	0	0
6	Man	Mammary glands adenoma, fibroma and fibroadenoma	0.5	51	0	0
7	Man	Mammary glands adenoma, fibroma and fibroadenoma	5	51	0	0
8	Man	Mammary glands adenoma, fibroma and fibroadenoma	50	51	1	1

Het lijkt alsof de relatie anders is bij vrouwtjes dan bij mannetjes, maar ik zie in beiden geen duidelijk *dose-response*. En hoewel ik bij mannetjes weinig zie gebeuren, is er bij vrouwtjes wel degelijk sprake van een toename van kanker in de glyfosaatgroepen ten opzichte van de controlegroep.

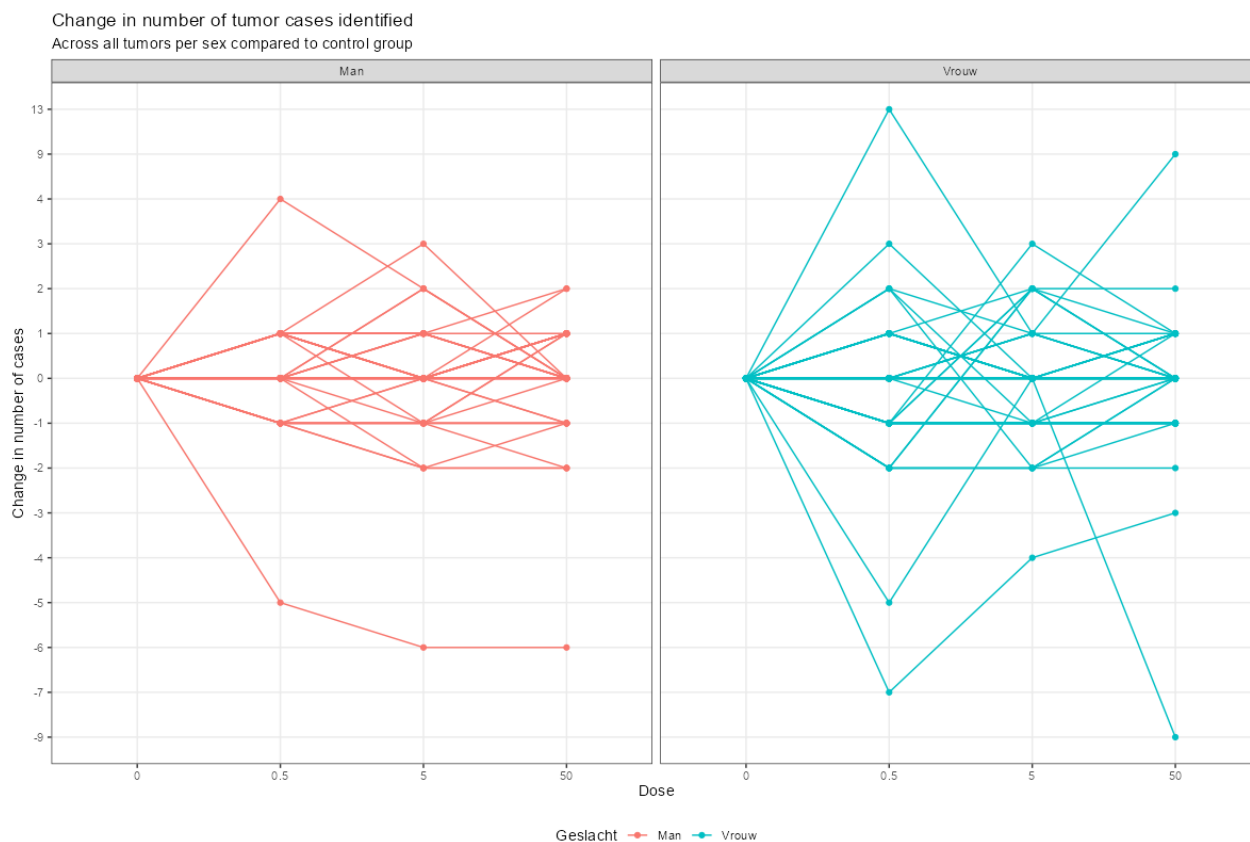


Figuur 84. Verdeling van de veranderingen in het aantal tumoren, per tumorsoort en geslacht, wanneer vergeleken met de controlegroep.

Ik ben nu wel benieuwd hoe vaak we een stijging zien ten opzichte van de controle groep en hoe vaak we een daling zien. Ik kijk dus nu even niet per tumor, maar gewoon naar het totaal. We zien dan dat van de 632 mutaties, er 537 net zo veel of minder kanker laten zien dan de controle groep. In 95 gevallen is het meer kanker. Als we deze getallen splitsen per geslacht dan krijgen we de volgende verdeling:

Geslacht	Equal or less cancer	More cancer
----------	----------------------	-------------

Man	246	50
Vrouw	291	45

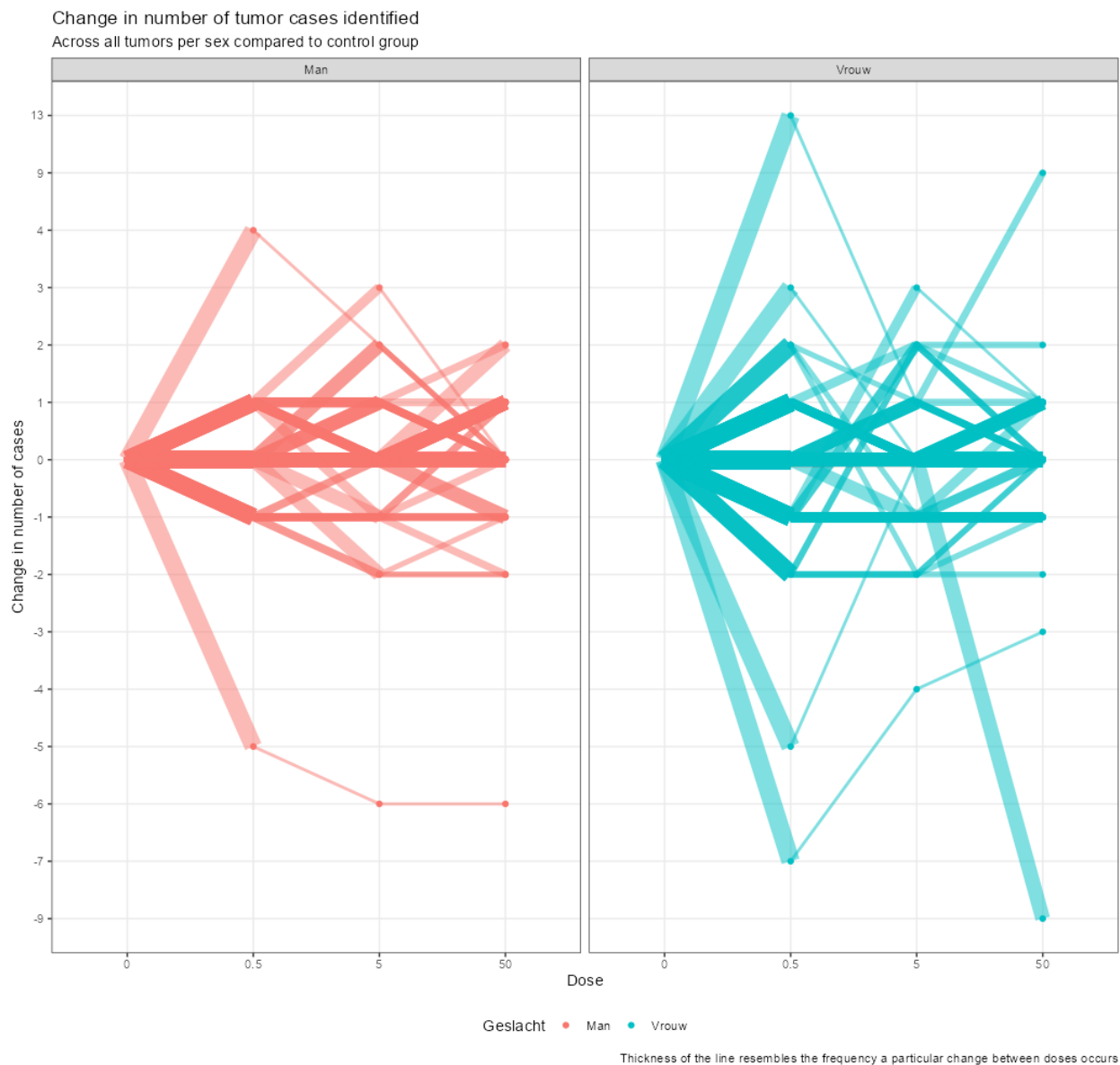


Figuur 85. De verandering in het aantal tumoren, per tumor en geslacht, vergeleken met de controlegroep.

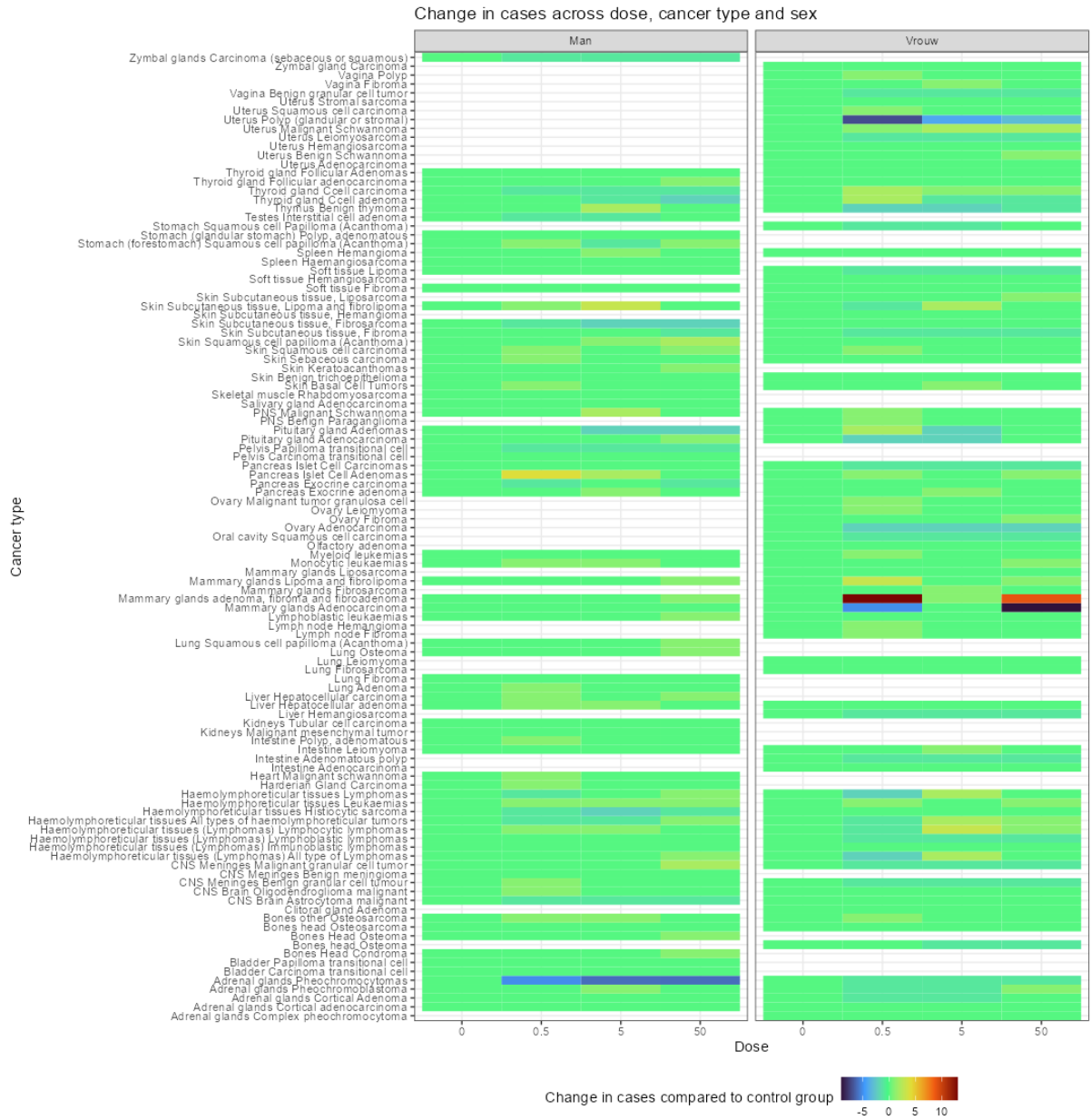
Ik had eigenlijk verwacht meer kanker te zien bij de vrouwtjes zeker na het zien van **Figuur 86**. Ook als ik kijk naar de ruwe data dan lijkt het alsof er meer kanker wordt vastgesteld bij vrouwtjes dan bij mannetjes. Het probleem met de vorige analyse is natuurlijk dat ik elke groep vergelijk met de controle groep en het helemaal niet zo hoeft te zijn dat een hogere dosis zorgt voor meer kanker. Dat zagen we in **Figuur 84** al. Misschien is het wel raadzaam om te laten zien hoe, per geslacht, de aantallen voor elke groep veranderen per tumor. Het resultaat zien we in **Figuur 87** en wat wederom duidelijk wordt is voor hoeveel verschillende soorten kanker er wel niet is gekeken én hoezeer het aantal tumorgevallen dat eigenlijk zou moeten stijgen door glyfosaat wel meevalt.

Ik ben nu eigenlijk ook wel weer opnieuw benieuwd of er tumoren zijn waar bij geen enkele groep iets is gevonden en welke dat dan zijn. Door middel van **Figuur 88** zien we dat er wel degelijk heel wat tumoren zijn meegenomen waarbij in geen enkele groep kanker is vastgesteld. Dit is in principe een goede methodologische keuze omdat het aangeeft voor

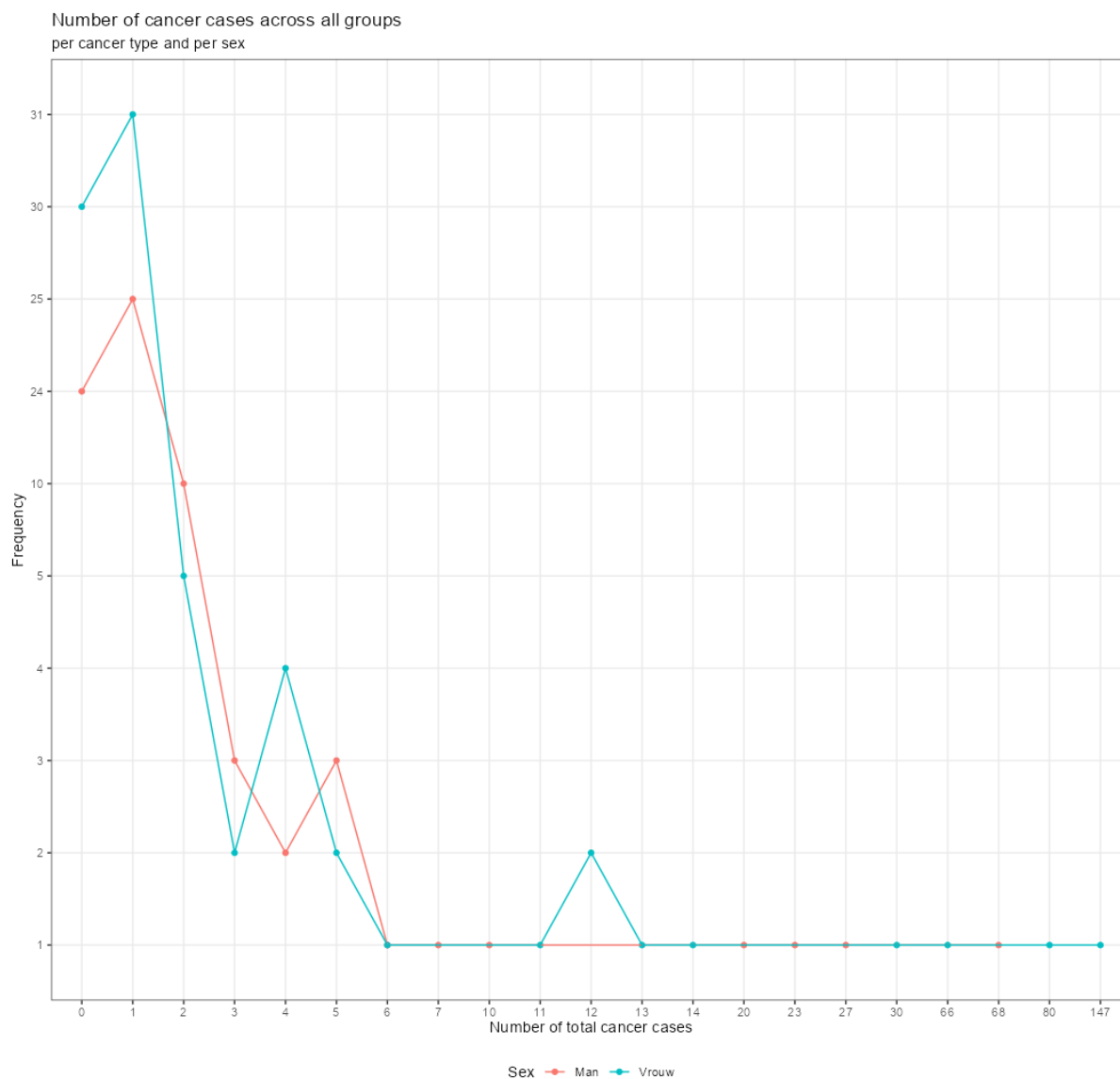
welke mogelijke kankersoorten er wel en geen kanker is gevonden. Daarmee doet de Panzacchi studie iets waar ik in mijn rapport over de Portier analyse kritiek op had: laten zien waar en wanneer er geen tumoren zijn.



Figuur 86. De verandering in het aantal tumoren, per tumor en geslacht, vergeleken met de controlegroep. De dikte van de lijn geeft de frequentie aan.



Figuur 87. Verandering in aantal tumoren, per tumorsoort en geslacht, wanneer vergeleken met de controlegroep.



Figuur 88. Totaal aantal tumoren gezien per kankersoort en geslacht. Wanneer in totaal nul tumoren zijn gezien dan betekent dit dat de kankersoort wel wordt genoemd, maar niet is geobserveerd.

We weten dus nu dat de grote hoeveelheid tumoren in de Panzacchi studie niet betekent dat al deze tumoren ook gezien zijn. Het volgende wat ik wil gaan doen is kijken of ik de resultaten uit de studie kan nabootsen. Maar dat ga ik niet voor elke tumor doen, want in de Panzacchi studie is geen correctie uitgevoerd voor het uitvoeren van meerdere statistische testen. Wie hier aan twijfelt doet er goed aan om in het supplement te kijken: daar staan alle significante resultaten beschreven¹²⁷. We zien dan iets heel bijzonders, want van de 74 tumoren beschreven bij de mannetjes zien we 10 statistisch significante effecten, en van de 84 tumoren beschreven bij de vrouwtjes zien we er maar drie (**Tabel 5**). Hier is overduidelijk

¹²⁷ De niet-significante resultaten zien we gek genoeg niet, wat eigenlijk betekent dat er selectief wordt gerapporteerd, maar dat terzijde.

niet gecorrigeerd voor het uitvoeren van meerdere testen, want het gros van de p-waardes is allemaal hetzelfde. Zouden we een hele simpele berekening doen voor het mogelijk aantal vals-positieven, waarbij we alleen kijken naar de tumoren in **Tabel 5**, dan is die kans ongeveer 54%¹²⁸.

Tumor	Mannetjes	Vrouwtjes
Skin Squamous cell papilloma (Acanthoma)	0.0392	-
Skin Keratoacanthoma	0.0419	-
Skin Subcutaneous tissue, Liposarcoma	-	0.0419
Mammary glands adenoma, fibroma and fibroadenoma	0.0419	-
Mammary glands lipoma and fibrolipoma	0.0419	-
Lung squamous cell pappiloma (ancanthoma)	0.0419	-
Lung Osteoma	0.0419	-
Ovary Fibroma	-	0.0419
Thyroid gland follicular adenocarcinoma	0.0419	-
Uterus Benign Schwannoma	-	0.0419
CNS malignant granular cell tumor	0.0071	-
Bnes head condroma	0.0419	-
Bones head osteoma	0.0419	-
Lymphoblastic leukaemias	0.0419	-
Monocytic leukaemias	-	0.0419

Tabel 5. Alle statistisch significante verbanden per unieke kankersoort zoals gerapporteerd in Panzacchi, voor mannetjes en vrouwtjes apart.

Maar in het artikel van Panzacchi staat nog iets anders beschreven wat verraadt waarom we meerdere p-waardes zien, namelijk dit:

Survival differences among groups were considered in assessing the incidence of neoplastic lesions and the Polyk test was used [36–39]. This test is a survival-adjusted quantal-response procedure that modifies the Cochran-Armitage linear trend test to account for survival differences by assigning a reduced risk of neoplasm, proportional to the power of the fraction of time on study, to only those lesion-free animals that did not reach terminal euthanasia. More specifically, this method modifies the denominator in the quantal estimate of lesion incidence to approximate more closely the total number of animal years at risk. Each animal is assigned a risk weight: this value is 1 if the animal had a lesion at the site under consideration or if it survived until terminal euthanasia; if the animal died before terminal sacrifice and did not have a lesion at the site under consideration, its risk weight is the fraction of the entire study time that it survived, raised to the k-th power. A value of k = 3 was used in the analysis, as recommended by Bailer and Portier [36, 37]. Reported p values for these analyses are one-sided.

¹²⁸ De formule voor het uitrekenen van de *family-wise error rate* (zie mijn [eerste rapport](#) voor een uitvoerige beschrijving) is $1 - (1 - \alpha)^{\text{aantal vergelijkingen}}$. In dit geval is dat dus $1 - (1 - 0.5)^{15} = 53.67\%$

Wat hierboven staat klinkt indrukwekkend, maar betekent in principe alleen maar dat de toets om te bepalen of er statistische significantie is, is aangepast(111). Meer specifiek betekent dit dat dieren die eerder sterven zonder tumor(en) in de analyse minder gewicht krijgen dan dieren die wel tumoren krijgen, maar overleven. Op deze manier kan men met minder informatie een berekening uitvoeren, maar ik vraag me oprecht af wat de waarde is door met minder informatie te rekenen. Sterker nog, ik krijg het gevoel dat men alleen maar probeert om sneller statistische significantie te bereiken zodat men geen bevinding 'over het hoofd ziet'. Gecombineerd met de eenzijdige Cochran-Armitage toets (CA-trend test die ik ook uitvoerig heb besproken in mijn [eerste rapport](#)) lijkt deze studie eerder op een screeningsstudie dan op doorslaggevende informatie. Dit verklaart wellicht waarom er geen correctie voor statistische significantie is gedaan. Het maakt ook dat we er wellicht beter aan doen om vanaf nu beschrijvend aan het werk te gaan, omdat we weten dat de p-waarden niet de juiste correctie hebben noch opereren op basis van volledige informatie. Voordat ik de focus ga leggen op specifieke tumoren, wil ik **Tabel 5** uitbreiden met de bevindingen op basis van de *poly-k-adjusted test*. We krijgen dan **Tabel 6** wat een lagere p-waarde toont voor de aangepaste statistiek (Poly-k) ten opzichte van de eenzijdige CA-trend test.

Tumor	Mannetjes		Vrouwtjes	
	CA-trend	Poly-k	CA-trend	Poly-k
Skin Squamous cell papilloma (Acanthoma)	0.0392	0.0274	-	
Skin Keratoacanthoma	0.0419	0.0311	-	
Skin Subcutaneous tissue, Liposarcoma	-		0.0419	
Mammary glands adenoma, fibroma and fibroadenoma	0.0419		-	
Mammary glands lipoma and fibrolipoma	0.0419		-	
Lung squamous cell papilloma (ancanthoma)	0.0419		-	
Lung Osteoma	0.0419		-	
Ovary Fibroma	-		0.0419	0.0411
Thyroid gland follicular adenocarcinoma	0.0419	0.0311	-	
Uterus Benign Schwannoma	-		0.0419	0.0411
CNS malignant granular cell tumor	0.0071	0.0040	-	
Bones head condroma	0.0419	0.0311	-	
Bones head osteoma	0.0419	0.0335	-	
Lymphoblastic leukaemias	0.0419	0.0335	-	
Monocytic leukaemias	-		0.0419	0.0411

Tabel 6. Statistisch significante resultaten zoals beschreven in de Panzacchi studie – nu per analysesoort.

We kunnen nu twee dingen gaan doen. Ten eerste kunnen we de bevindingen uit de Parzacchi studie combineren met de bevindingen van de Portier analyse om te zien of de

bevindingen uit die analyse, en die uit mijn eerste rapport, veranderen. We zagen in **Figuur 77** al dat het combineren van beide bronnen wat voeten in de aarde heeft, omdat de benamingen verschillen. Maar wellicht dat we niet zover hoeven te gaan, want we kunnen ook zien of de bevindingen uit de dierstudies vergelijkbaar zijn met de bevindingen uit de studies onder mensen. Dat is eigenlijk ook het doel van dit rapport. Maar als we dan kijken naar **Tabel 5** en **Tabel 6** dan zien we niet direct de kankersoorten terug die we beschreven in het eerste deel van dit rapport. Daar was de focus vooral op lymfoom, nierkanker, hersenkanker en beenmergkanker. En hoewel beenmergkanker en leukemie gelijkenissen hebben zijn het kankersoorten die zich heel anders ontwikkelen. De enige tumoren uit **Tabel 5** en **Tabel 6** die ogenschijnlijk een mogelijke connectie kunnen hebben met de studies onder mensen zijn *CNS malignant granular cell tumor*, *Lymphoblastic leukaemias* en *Monocytic leukaemias*, maar een bioloog/oncoloog zou moeten bevestigen of dit echt zo is¹²⁹.

Ik wil nu teruggaan naar alle studies uit de Portier analyse om te zien of deze soorten ook daar gezien zijn. Het kan heel goed zijn dat ik wellicht de benamingen moet aanpassen, maar dat is dan maar zo. En ondanks al mijn kritiek op hoe die analyse is gedaan vanuit een methodologisch en statistisch oogpunt, kan het wel zinvol zijn om minstens te beschrijven wat er in die studies gevonden is. Uiteindelijk draait het om het vast stellen van de relatie tussen dierstudies en studies bij mensen.

Eerst begin ik met de lymfomen (elke vorm van lymfoomkanker) waarbij ik ook expliciet zal zoeken op elke kankersoort met het woord *cell* erin¹³⁰. Dat geeft een lijst van ongeveer 49 kankersoorten:

Tumor	n
<chr>	<int>
1 Bladder Carcinoma transitional cell	1
2 Bladder Papilloma transitional cell	1
3 CNS Meninges Benign granular cell tumour	1
4 CNS Meninges Malignant granular cell tumor	1
5 Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas	1
6 Haemolymphoreticular tissues (Lymphomas) Immunoblastic lymphomas	1
7 Haemolymphoreticular tissues (Lymphomas) Lymphoblastic lymphomas	1
8 Haemolymphoreticular tissues (Lymphomas) Lymphocytic lymphomas	1
9 Haemolymphoreticular tissues All types of haemolymphoreticular tumors	1
10 Haemolymphoreticular tissues Histiocytic sarcoma	1
11 Haemolymphoreticular tissues Leukaemias	1
12 Haemolymphoreticular tissues Lymphomas	1
13 Hepatocellular Adenomas	1
14 Hepatocellular Adenomas and Carcinomas	1
15 Hepatocellular Carcinomas	1
16 Kidneys Tubular cell carcinoma	1
17 Liver Hepatocellular adenoma	1
18 Liver Hepatocellular carcinoma	1

¹²⁹ En nee, dit soort vragen laat ik niet aan AI over.

¹³⁰ Vanwege T-cell of B-cell lymfoom.

19	Lung Squamous cell papilloma (Acanthoma)	1
20	Lymph node Fibroma	1
21	Lymph node Hemangioma	1
22	Lymphoblastic leukaemias	1
23	Malignant Lymphomas	1
24	Oral cavity Squamous cell carcinoma	1
25	Ovary Malignant tumor granulosa cell	1
26	Pancreas Islet Cell Adenomas	1
27	Pancreas Islet Cell Adenomas or Carcinomas	1
28	Pancreas Islet Cell Carcinomas	1
29	Pelvis Carcinoma transitional cell	1
30	Pelvis Papilloma transitional cell	1
31	Skin Basal Cell Tumors	1
32	Skin Squamous cell carcinoma	1
33	Skin Squamous cell papilloma (Acanthoma)	1
34	Spleen Composite Lymphosarcoma	1
35	Stomach (forestomach) Squamous cell papilloma (Acanthoma)	1
36	Stomach Squamous cell Papilloma (Acanthoma)	1
37	Testes Interstitial cell adenoma	1
38	Testicular Interstitial Cell Tumors	1
39	Thyroid CCell Adenomas	1
40	Thyroid CCell Adenomas and Carcinomas	1
41	Thyroid CCell Carcinomas	1
42	Thyroid FollicularCell Adenoma and Carcinoma	1
43	Thyroid FollicularCell Adenomas	1
44	Thyroid FollicularCell Adenomas and Carcinomas	1
45	Thyroid FollicularCell Carcinomas	1
46	Thyroid gland ccell adenoma	1
47	Thyroid gland Ccell carcinoma	1
48	Uterus Squamous cell carcinoma	1
49	Vagina Benign granular cell tumor	1

In die lijst zien we wel degelijk lymfomen opgenomen, en het gros herken ik uit de Panzacchi studie. Als ik selecteer op de studies die in Portier zijn opgenomen dan zie ik de volgende lijst:

Tumor	n	
<chr>	<int>	
1	Hepatocellular Adenomas	1
2	Hepatocellular Adenomas and Carcinomas	1
3	Hepatocellular Carcinomas	1
4	Malignant Lymphomas	1
5	Pancreas Islet Cell Adenomas	1
6	Pancreas Islet Cell Adenomas or Carcinomas	1
7	Pancreas Islet Cell Carcinomas	1
8	Skin Basal Cell Tumors	1
9	Spleen Composite Lymphosarcoma	1
10	Testicular Interstitial Cell Tumors	1
11	Thyroid CCell Adenomas	1
12	Thyroid CCell Adenomas and Carcinomas	1
13	Thyroid CCell Carcinomas	1
14	Thyroid FollicularCell Adenoma and Carcinoma	1
15	Thyroid FollicularCell Adenomas	1
16	Thyroid FollicularCell Adenomas and Carcinomas	1
17	Thyroid FollicularCell Carcinomas	1

Hier zien we *malignant lymphomas* duidelijk terugkomen en dit is waarschijnlijk de kankersoort die het dichtst bij NHL ligt. Deze beschrijving is dus niet in Panzacchi opgenomen, maar daar lezen we benamingen zoals *Haemolymphoreticular tissues (Lymphomas)*. Het zou heel goed kunnen zijn dat hier nu data ligt die we kunnen gebruiken en deze bevinding bepaalt ook direct onze volgende selectie. Ik zal daarom alle soorten selecteren waarin lymfoom beschreven staat wat maakt dat ik in totaal tien kankersoorten

selecteer. Dit geeft mij op zijn minst een idee van wat de verschillende studies hebben gevonden. Ik krijg dan 96 rijen aan data met in totaal zeven studies:

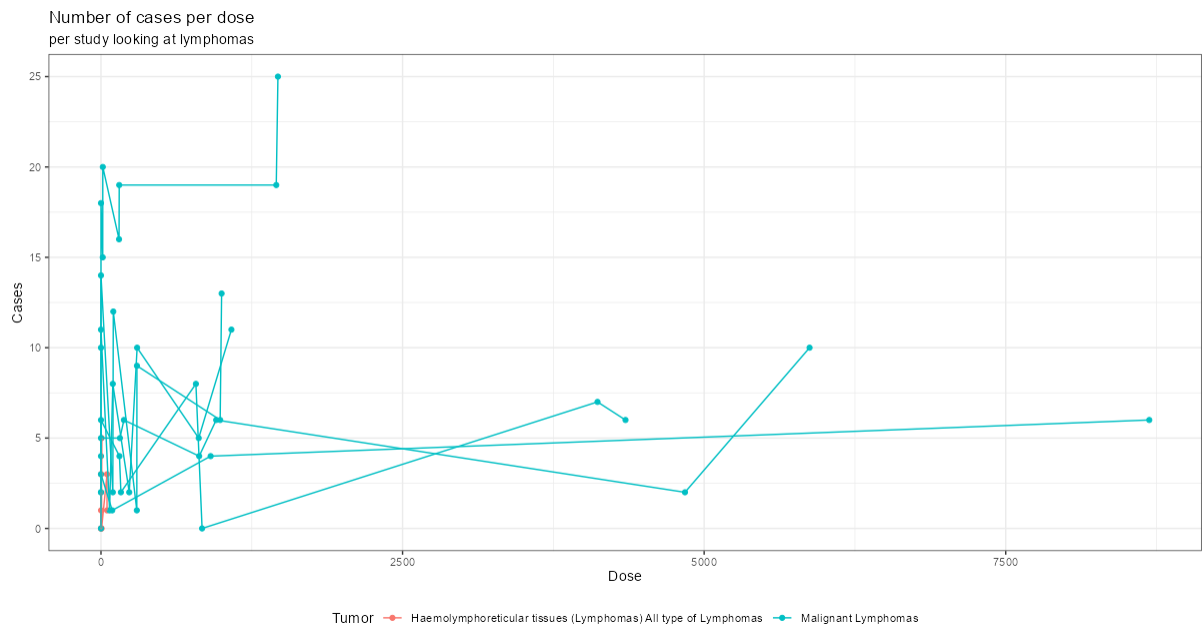
```
1 Panzacchi
2 Atkinson_a
3 Knezevich and Hogan
4 Kumar
5 Sugimoto
6 Takahashi
7 Wood_a
```

Als ik kijk hoeveel studies er zijn voor elke van de tien kankersoorten dan zie ik de volgende resultaten:

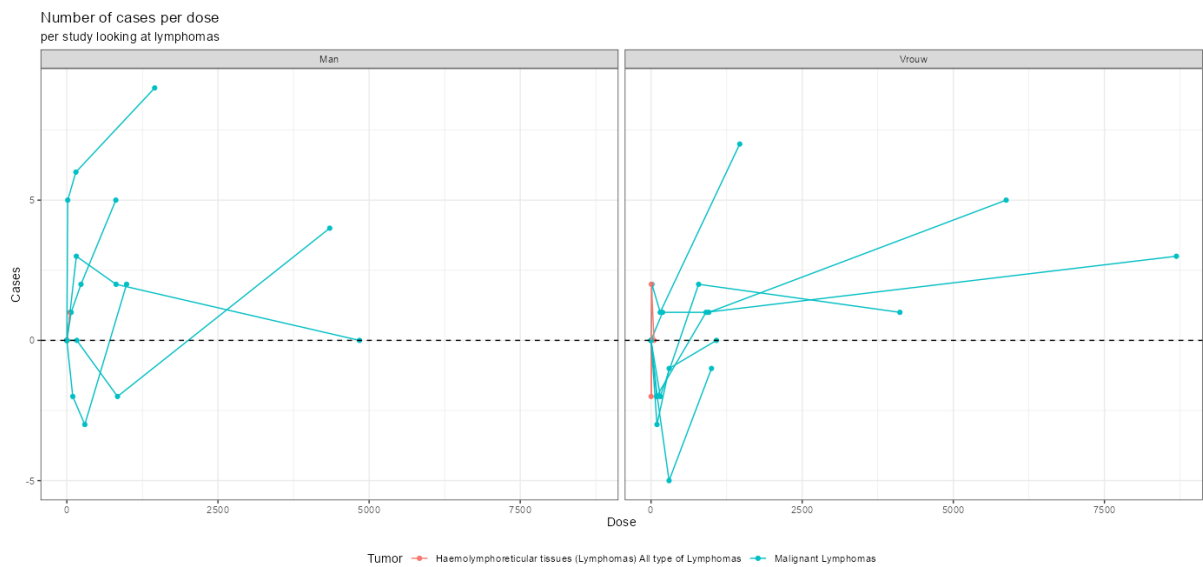
Tumor	Studie
<chr>	<chr>
1 Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas	Panzacchi
2 Haemolymphoreticular tissues (Lymphomas) Immunoblastic lymphomas	Panzacchi
3 Haemolymphoreticular tissues (Lymphomas) Lymphocytic lymphomas	Panzacchi
4 Haemolymphoreticular tissues Lymphomas	Panzacchi
5 Lymph node Fibroma	Panzacchi
6 Lymph node Hemangioma	Panzacchi
7 Lymphoblastic leukaemias	Panzacchi
8 Malignant Lymphomas	Atkinson_a
9 Malignant Lymphomas	Knezevich and Hog~
10 Malignant Lymphomas	Kumar
11 Malignant Lymphomas	Sugimoto
12 Malignant Lymphomas	Takahashi
13 Malignant Lymphomas	Wood_a
14 Spleen Composite Lymphosarcoma	Knezevich and Hog~

We zien dus, wederom, dat het gros van de lymfomen in Panzacchi gemeten zijn. En we zien ook dat het type *malignant lymphomas* in meerder studies gemeten is. Misschien kunnen we *malignant lymphomas* combineren met *Haemolymphoreticular tissues (Lymphomas) All type of Lymphomas*. Voordat we dit doen wil ik bovenstaande resultaten grafisch weergeven om een idee te krijgen van de relaties en wat het nu exact is waar we naar kijken. Het resultaat is **Figuur 89**, maar dat helpt ons niet echt. Wat direct opvalt is dat de dosering bij de Panzacchi studie veel lager lijkt dan de rest. Daarom is het wellicht beter om per studie de verandering te laten zien ten opzichte van de controlegroep. Ook zal ik dit per geslacht tonen. We krijgen dan **Figuur 90**, maar deze figuur maakt het niet direct gemakkelijker om een verband te zien. Wat we wel zien is nog duidelijker hoe de dosering uit de Panzacchi studie afsteekt tegen de rest¹³¹. Als we de waarden iets anders proberen te tonen, door middel van **Figuur 91**, dan zien we dat er net zo vaak een stijging is in het aantal lymfomen als dat er een daling is. Ook zien we, wederom, hoeveel verschillende doseringen er zijn meegenomen. Dit maakt het niet makkelijk om de data te modelleren, zoals ik in het [eerste rapport](#) al beschreef.

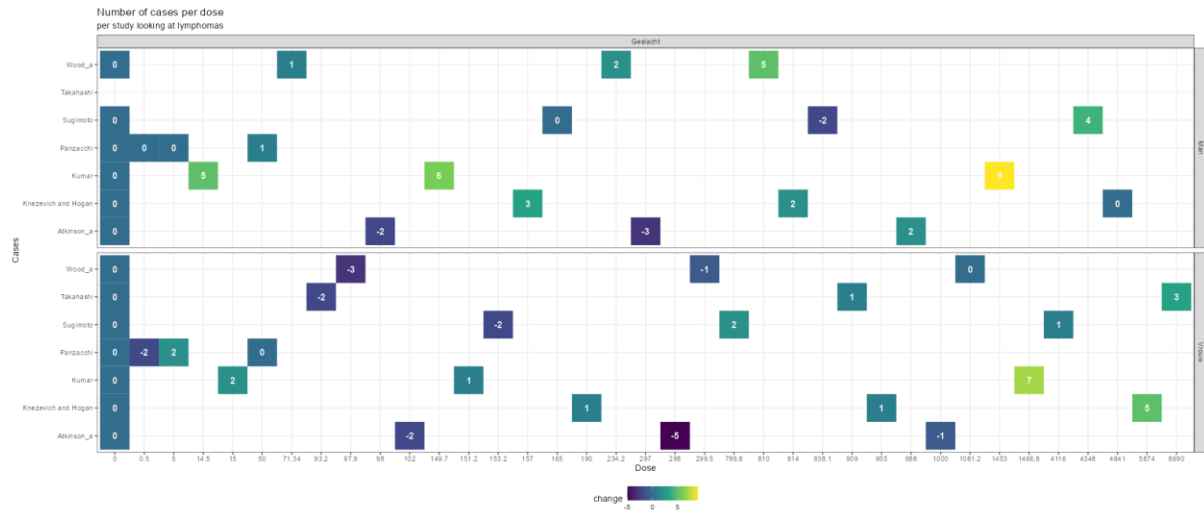
¹³¹ Of ik heb hier iets fout gedaan, dat kan natuurlijk ook en het is aan de lezer om dat te controleren.



Figuur 89. Aantal gevallen van kanker voor twee soorten lymphoma per dosering.



Figuur 90. Aantal gevallen van kanker voor twee soorten lymphoma per dosering en geslacht.



Figuur 91. Verschil tussen het aantal kankergevallen in de controlegroep en de glyphosaatgroepen per studie, geslacht en dosering voor lymfoomkanker.

Bovenstaande analyse voelt onbevredigend, omdat ik de data van de dierproeven niet zo makkelijk kan combineren met de bevindingen uit de menselijke studies. Maar voordat we die stap gaan wagen wil ik eerst nog kijken naar alle vormen van leukemie¹³², nierkanker en hersenkanker. Als ik zoek naar alle mogelijke vormen van leukemie dan zie ik de volgende lijst:

```
# A tibble: 4 x 2
  Tumor          n
  <chr>          <int>
1 Haemolymphe... 1
2 Lymphoblast... 1
3 Monocytic le... 1
4 Myeloid leuk... 1
```

Ik kan nu gaan kijken welke studies naar deze vormen van leukemie hebben gekeken, maar ik zie al snel dat dit alleen Panzacchi is. Daarmee is een vergelijking met Portier niet mogelijk. Om er zeker van te zijn dat ik niks vergeten ben, roep ik alle kankersoorten op die in de Portier analyse zijn meegenomen. Ik zie hier inderdaad geen leukemie staan:

```
# A tibble: 44 x 1
  Tumor
  <chr>
1 Adrenal Cortical Adenomas
2 Adrenal Cortical Adenomas and Carcinomas
3 Adrenal Cortical Carcinomas
4 Adrenal Pheochromocytomas
5 AlveolarBronchiolar Adenomas
6 AlveolarBronchiolar Adenomas and Carcinomas
7 AlveolarBronchiolar Carcinomas
8 Harderian Gland Adenomas
9 Harderian Gland Adenomas and Carcinomas
10 Harderian Gland Carcinomas
```

¹³² Hoewel deze kankersoort echt anders is dan beenmergkanker, wordt leukemie wel sterk aangehaald in de Panzacchi studie. Dat alleen maakt het al interessant om hier dieper in te duiken. En we hebben natuurlijk gezien dat er bij de studies onder mensen onderzoek is gedaan naar *hairy cell leukemia*.

```

11 Hemangiomas
12 Hemangiosarcomas
13 Hepatocellular Adenomas
14 Hepatocellular Adenomas and Carcinomas
15 Hepatocellular Carcinomas
16 Kidney Adenomas
17 Kidney Adenomas (original pathology)
18 Kidney Adenomas and Carcinomas
19 Kidney Carcinomas
20 Liver Carcinomas
21 Liver Neoplastic Nodules
22 Liver Nodules and Carcinomas
23 Malignant Lymphomas
24 Mammary Gland Adenocarcinomas
25 Mammary Gland Adenomas
26 Mammary Gland Adenomas and Adenocarcinomas
27 Pancreas Islet Cell Adenomas
28 Pancreas Islet Cell Adenomas or Carcinomas
29 Pancreas Islet Cell Carcinomas
30 Pituitary Adenomas
31 Pituitary Adenomas and Carcinomas
32 Pituitary Carcinomas
33 Skin Basal Cell Tumors
34 Skin Epithelioma (Keratoacanthomas)
35 Skin Keratoacanthomas
36 Spleen Composite Lymphosarcoma
37 Testicular Interstitial Cell Tumors
38 Thyroid CCell Adenomas
39 Thyroid CCell Adenomas and Carcinomas
40 Thyroid CCell Carcinomas
41 Thyroid FollicularCell Adenoma and Carcinoma
42 Thyroid FollicularCell Adenomas
43 Thyroid FollicularCell Adenomas and Carcinomas
44 Thyroid FollicularCell Carcinomas
    
```

Ik hoef daarom ook niet verder te zoeken en/of grafieken te maken. Dan rest mij nog op zoek te gaan naar nierkanker en kanker in het centrale zenuwstelsel, zoals bij de hersenen. Eerst ga ik beginnen bij nierkanker en een kleine zoektocht¹³³ toont mij de volgende lijst:

```

# A tibble: 15 x 2
  Tumor                                n
  <chr>                                <int>
1 Adrenal Cortical Adenomas           1
2 Adrenal Cortical Adenomas and Carcinomas 1
3 Adrenal Cortical Carcinomas         1
4 Adrenal Pheochromocytomas           1
5 Adrenal glands Complex pheochromocytoma 1
6 Adrenal glands Cortical Adenoma     1
7 Adrenal glands Cortical adenocarcinoma 1
8 Adrenal glands Pheochromoblastoma   1
9 Adrenal glands Pheochromocytomas    1
10 Kidney Adenomas                     1
11 Kidney Adenomas (original pathology) 1
12 Kidney Adenomas and Carcinomas     1
13 Kidney Carcinomas                   1
14 Kidneys Malignant mesenchymal tumor 1
15 Kidneys Tubular cell carcinoma     1
    
```

Nu is het zo dat de *adrenal glands* boven op de nieren zitten en daarom ook wel bijnier worden genoemd. Ik zie in de dierstudies uiteraard niet specifiek *Wilm's tumor* staan, maar met bovenstaande lijst zie ik misschien wel een link naar de nierkanker die we bij kinderen hebben gezien. Ik wil dus graag met deze lijst verder en zien hoeveel studies we per

¹³³ Ik zoek hier zowel op de woorden 'renal' als 'kidney'.

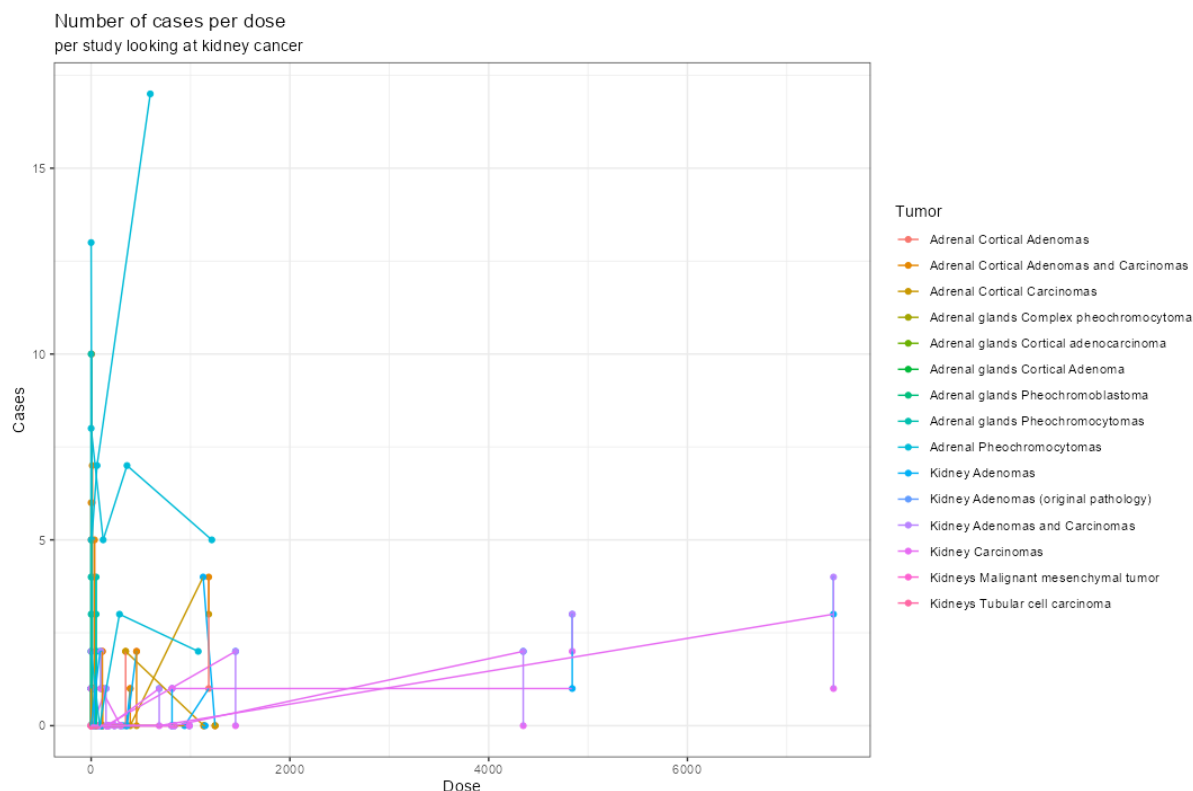
tumorsoort hebben. In feite herhaal ik nu wat ik bij de lymfomen heb gedaan, maar bij de vormen van leukemie niet kon vanwege een tekort aan studies. Een lijst van alles studies die naar (bij)nierkanker hebben gekeken laat een langere lijst zien dan bij de lymfomen:

```
# Groups:   Tumor [15]
  Tumor
  <chr>
1 Adrenal glands Complex pheochromocytoma Panzacchi
2 Adrenal glands Cortical adenocarcinoma Panzacchi
3 Adrenal glands Cortical Adenoma Panzacchi
4 Adrenal Cortical Adenomas Atkinson_b
5 Adrenal Cortical Adenomas Enemoto
6 Adrenal Cortical Adenomas Stout and Ruecker
7 Adrenal Cortical Adenomas Lankas
8 Adrenal glands Pheochromoblastoma Panzacchi
9 Adrenal Cortical Adenomas and Carcinomas Atkinson_b
10 Adrenal Cortical Adenomas and Carcinomas Enemoto
11 Adrenal Cortical Adenomas and Carcinomas Stout and Ruecker
12 Adrenal Cortical Adenomas and Carcinomas Lankas
13 Adrenal Cortical Carcinomas Atkinson_b
14 Adrenal Cortical Carcinomas Enemoto
15 Adrenal Cortical Carcinomas Stout and Ruecker
16 Adrenal Cortical Carcinomas Lankas
17 Adrenal glands Pheochromocytomas Panzacchi
18 Adrenal Pheochromocytomas Brammer
19 Adrenal Pheochromocytomas Suresh
20 Adrenal Pheochromocytomas Wood_b
21 Kidney Adenomas Atkinson_a
22 Kidney Adenomas Atkinson_b
23 Kidney Adenomas Enemoto
24 Kidney Adenomas Knezevich and Hogan
25 Kidney Adenomas Kumar
26 Kidney Adenomas Lankas
27 Kidney Adenomas Stout and Ruecker
28 Kidney Adenomas Sugimoto
29 Kidney Adenomas Takahashi
30 Kidney Adenomas Wood_a
31 Kidney Adenomas (original pathology) Knezevich and Hogan
32 Kidney Adenomas and Carcinomas Atkinson_a
33 Kidney Adenomas and Carcinomas Knezevich and Hogan
34 Kidney Adenomas and Carcinomas Kumar
35 Kidney Adenomas and Carcinomas Sugimoto
36 Kidney Adenomas and Carcinomas Takahashi
37 Kidney Adenomas and Carcinomas Wood_a
38 Kidney Carcinomas Atkinson_a
39 Kidney Carcinomas Knezevich and Hogan
40 Kidney Carcinomas Kumar
41 Kidney Carcinomas Sugimoto
42 Kidney Carcinomas Takahashi
43 Kidney Carcinomas Wood_a
44 Kidneys Malignant mesenchymal tumor Panzacchi
45 Kidneys Tubular cell carcinoma Panzacchi
```

Samengevat zijn het dus de volgende studies die we moeten gaan opnemen:

```
# A tibble: 14 x 1
  Studie
  <chr>
1 Panzacchi
2 Atkinson_b
3 Enemoto
4 Stout and Ruecker
5 Lankas
6 Brammer
7 Suresh
8 Wood_b
9 Atkinson_a
10 Knezevich and Hogan
11 Kumar
12 Sugimoto
13 Takahashi
14 Wood_a
```

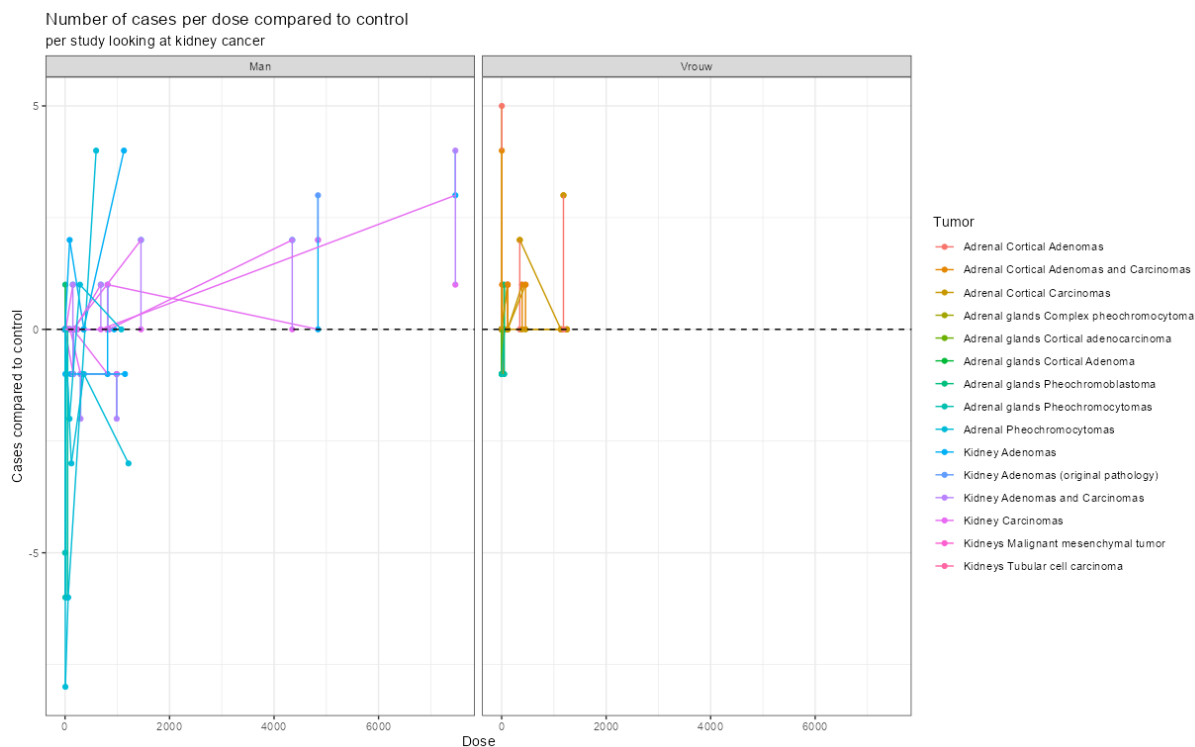
Het is ook wel een wat complexere lijst, zoals we kunnen zien in **Figuur 92** en **Figuur 93**. Ik wil nu ook graag een herhaling maken van **Figuur 91**, maar dat gaat deze keer niet zo makkelijk. We hebben namelijk een grotere verscheidenheid aan histologie dan bij de lymfomen én we hebben meer studies.



Figuur 92. Aantal tumoren gevonden per glyfosaat dosis over alle studies heen die keken naar (bij)nier kanker.

Dat betekent dat ik gegevens moet gaan groeperen en dat kan ik doen per studie óf per kankersoort. Omdat ik niet graag kankersoorten combineer doe ik het per studie om zo het verschil met de controle groep te combineren. Uiteraard is dat niet eenvoudig, want elke studie heeft wel een andere dosering. Een aantal keer proberen laat eigenlijk zien dat het niet mogelijk is om een dergelijk figuur te maken, omdat we dan over verschillende assen moeten groeperen waardoor we ook weer belangrijke informatie kwijtrafen (tumor en dosering bijvoorbeeld). Het is daarom beter om de data op te splitsen, maar het resultaat is overweldigend (zie **Figuur 94**). De kleuren verraden echter dat er over alles studies heen weinig gebeurt en ik heb al eerder geschreven over de problemen bij het statistisch samenvoegen. Mij rest dus alleen nog maar te kijken naar de tumoren van het centrale zenuwstelsel. Zo schreef ik al dat bijvoorbeeld *CNS malignant granular cell tumor* een

mogelijke kankersoort is die we kunnen verbinden aan de bevindingen uit de menselijke studies. Wat ik dus nu eigenlijk nu aan het doen ben is zien of de bevindingen, zoals getoond in **Figuur 9**, ook gevonden kunnen worden in de dierstudies. Maar als ik **Figuur 9** trouw wil volgen, dan zou ik eigenlijk ook op zoek moeten gaan naar de bevindingen voor *Thyroid cancer*: schildklierkanker. En hoewel deze tumor maar één keer is gezien wil ik wel volledig zijn, dus ga ik voor beiden.



Figuur 93. Aantal tumoren gevonden per glyfosaat dosis en per geslacht over alle studies heen die keken naar (bij)nier kanker.



Figuur 94. Verschil in kankergevallen, vergeleken met de controlegroep, voor alle studies die keken naar (bij)nierkanker voor zowel mannetjes als vrouwtjes.

Om te beginnen start ik dus met hersenkanker, of kanker in het centrale zenuwstelsel¹³⁴. Als ik zoek op hersenen, of centraal of centraal zenuwstelsel, dan krijg ik de volgende tumoren te zien:

```
Tumor      n
<chr>      <int>
1 CNS Brain Astrocytoma malignant 1
```

¹³⁴ Staat absoluut niet synoniem voor elkaar, maar het helpt mij in de zoektocht.

2	CNS Brain Oligodendroglioma malignant	1
3	CNS Meninges Benign granular cell tumour	1
4	CNS Meninges Benign meningioma	1
5	CNS Meninges Malignant granular cell tumor	1

En dit zijn allemaal soorten gezien in Panzacchi wat maakt dat we er niet veel aan hebben, omdat we weten dat deze bevindingen sowieso niet significant zijn. We kunnen deze kankersoort dus eigenlijk laten voor wat het is. Een extra check in de 44 kankersoorten die zijn meegenomen in de Portier analyse toont ook geen kanker van de hersenen. Wel zien we *Harderian Gland* beschreven, maar dat is een traanklier. Ook zien we een hoop schildkliertumoren beschreven staan, dus het is nu tijd om daar naar te kijken. We zien dan de volgende soorten beschreven staan:

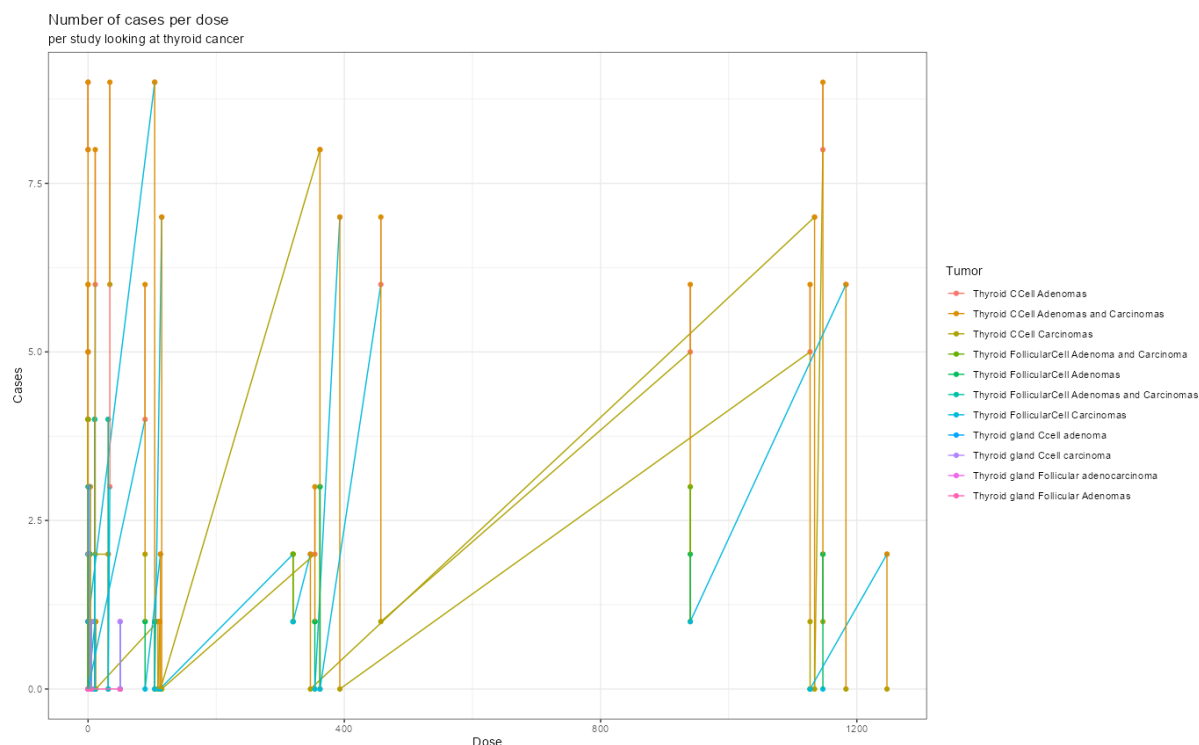
```
# A tibble: 11 x 2
  Tumor                                n
  <chr>                                <int>
1 Thyroid CCell Adenomas                1
2 Thyroid CCell Adenomas and Carcinomas 1
3 Thyroid CCell Carcinomas              1
4 Thyroid FollicularCell Adenoma and Carcinoma 1
5 Thyroid FollicularCell Adenomas        1
6 Thyroid FollicularCell Adenomas and Carcinomas 1
7 Thyroid FollicularCell Carcinomas      1
8 Thyroid gland Ccell adenoma           1
9 Thyroid gland ccell carcinoma         1
10 Thyroid gland Follicular Adenomas     1
11 Thyroid gland Follicular adenocarcinoma 1
```

En deze soorten zijn in de volgende studies opgenomen:

```
# A tibble: 28 x 2
# Groups:   Tumor [11]
  Tumor                                Studie
  <chr>                                <chr>
1 Thyroid CCell Adenomas                Atkinson_b
2 Thyroid CCell Adenomas                Enemoto
3 Thyroid CCell Adenomas                Lankas
4 Thyroid CCell Adenomas                Stout and Ruecker
5 Thyroid CCell Adenomas and Carcinomas Atkinson_b
6 Thyroid CCell Adenomas and Carcinomas Enemoto
7 Thyroid CCell Adenomas and Carcinomas Lankas
8 Thyroid CCell Adenomas and Carcinomas Stout and Ruecker
9 Thyroid CCell Carcinomas              Atkinson_b
10 Thyroid CCell Carcinomas             Enemoto
11 Thyroid CCell Carcinomas             Lankas
12 Thyroid CCell Carcinomas             Stout and Ruecker
13 Thyroid FollicularCell Adenoma and Carcinoma Atkinson_b
14 Thyroid FollicularCell Adenoma and Carcinoma Enemoto
15 Thyroid FollicularCell Adenoma and Carcinoma Stout and Ruecker
16 Thyroid FollicularCell Adenomas        Atkinson_b
17 Thyroid FollicularCell Adenomas        Enemoto
18 Thyroid FollicularCell Adenomas        Lankas
19 Thyroid FollicularCell Adenomas        Stout and Ruecker
20 Thyroid gland Follicular Adenomas     Panzacchi
21 Thyroid FollicularCell Adenomas and Carcinomas Lankas
22 Thyroid FollicularCell Carcinomas      Atkinson_b
23 Thyroid FollicularCell Carcinomas      Enemoto
24 Thyroid FollicularCell Carcinomas      Lankas
25 Thyroid FollicularCell Carcinomas      Stout and Ruecker
26 Thyroid gland Ccell adenoma           Panzacchi
27 Thyroid gland ccell carcinoma         Panzacchi
28 Thyroid gland Follicular adenocarcinoma Panzacchi
```

Dat zijn er dus uiteindelijk vijf en de resultaten staan grafisch afgebeeld in **Figuur 95**, **Figuur 96** en **Figuur 97**:

```
# A tibble: 5 x 1
  Studie
  <chr>
1 Atkinson_b
2 Enemoto
3 Lankas
4 Stout and Ruecker
5 Panzacchi
```



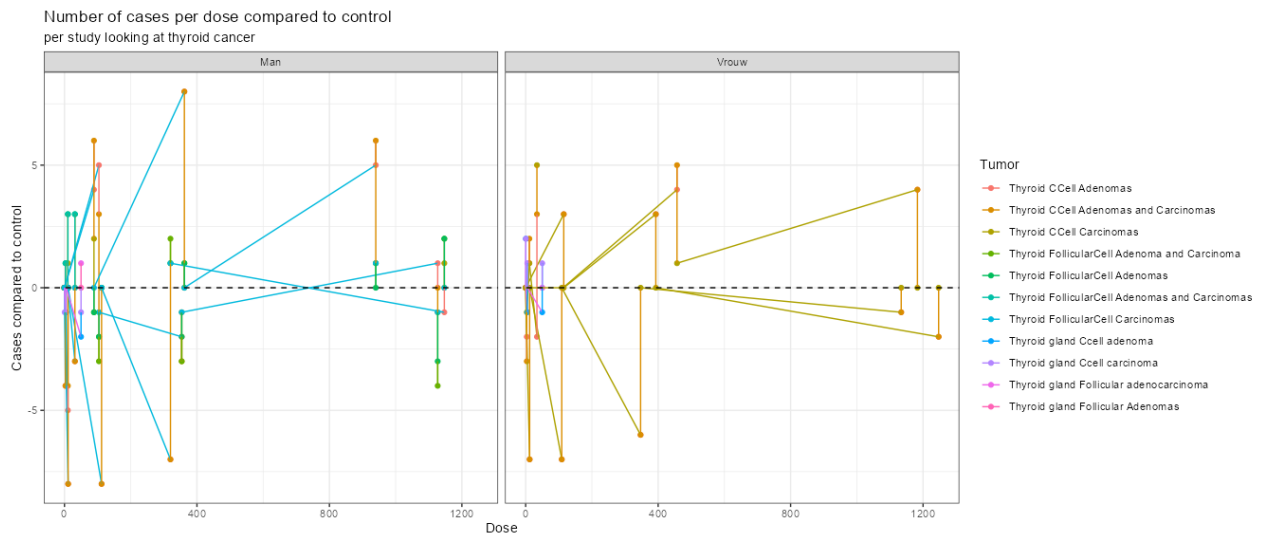
Figuur 95. Aantal tumoren gevonden per glyfosaat dosis over alle studies heen die keken naar schildklierkanker.

We kunnen heel duidelijk zien dat er geen duidelijke relatie is tussen glyfosaatdoseringen en het ontstaan van schildklierkanker. Hoewel er in sommige studies wel degelijk een toename is in schildklierkanker is dat geen vaste relatie die overall wordt gevonden.

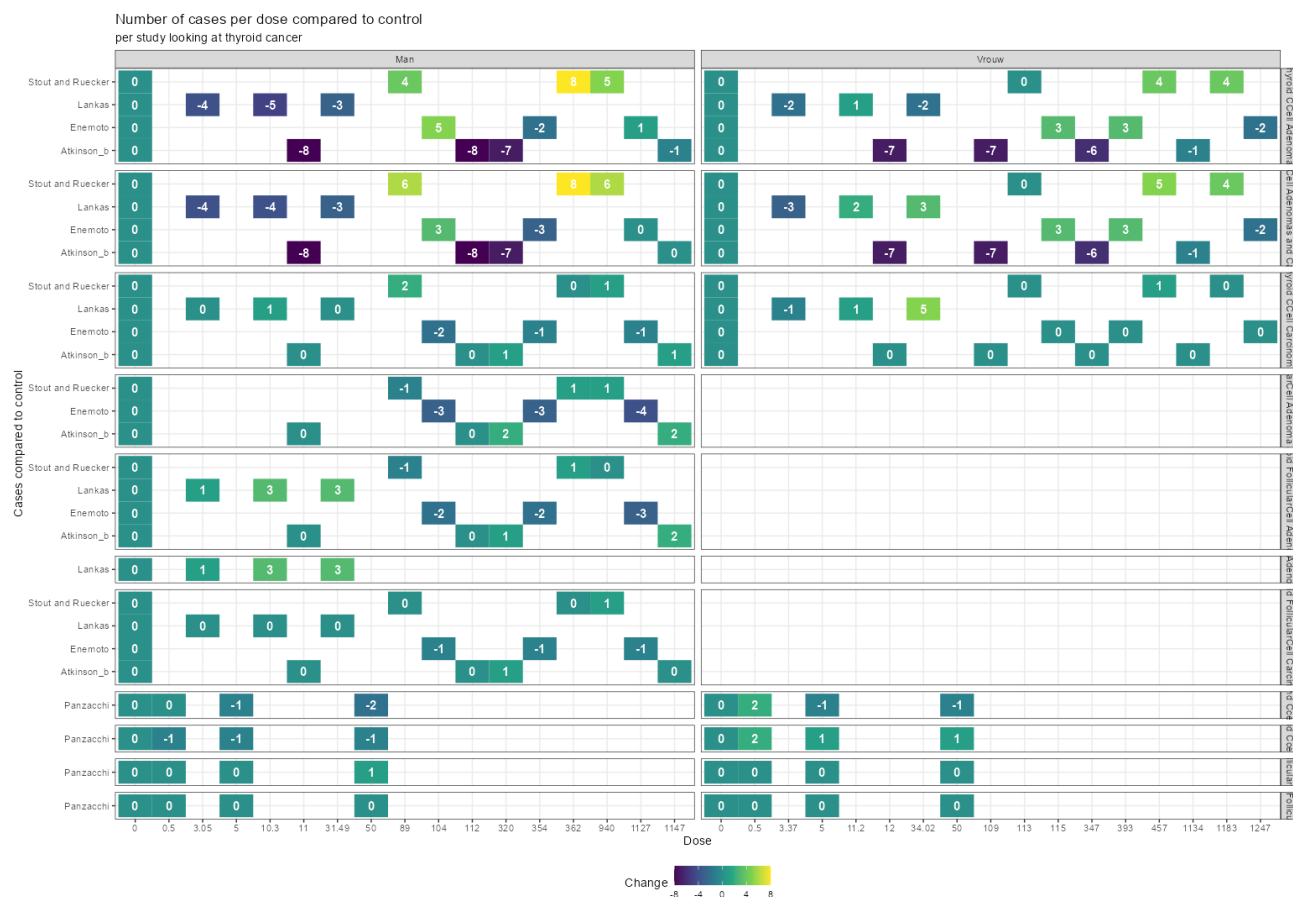
Er rest mij nog één analyse en dat is een analyse die zich specifiek richt op de controlegroep. Ik heb daar ook in mijn eerste rapport aardig wat tijd en moeite in gestoken, juist omdat de controlegroep verre van kankervrij is. En ik heb steeds vaker het idee, en lees dat ook terug in studies zoals Panzacchi, dat het observeren van maar één enkele tumor in de glyfosaatgroep(en)¹³⁵ al wordt gezien als iets noemenswaardigs. Dat dit ook als iets

¹³⁵ Ongeacht welke dosis dan ook.

noemenswaardigs wordt gepresenteerd zien we ook in Panzacchi waar we een aantal keren dezelfde p-waarde terug vinden. Op geen enkele manier lees ik de benodigde voorzichtigheid terug in het analyseren van dergelijke data.



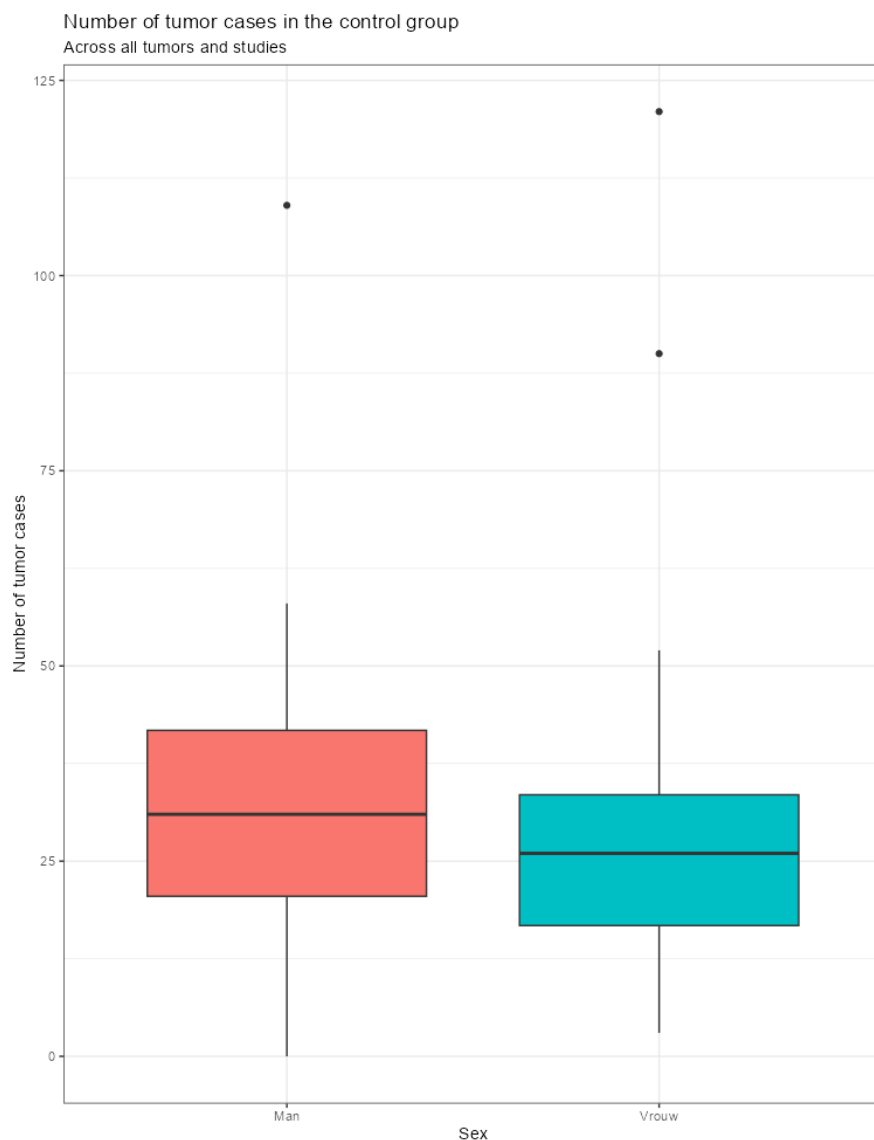
Figuur 96. Aantal tumoren gevonden per glyfosaat dosis en per geslacht over alle studies heen die keken naar schildklierkanker.



Figuur 97. Verschil in kankergevallen, vergeleken met de controlegroep, voor alle studies die keken naar schildklierkanker voor zowel mannetjes als vrouwtjes.

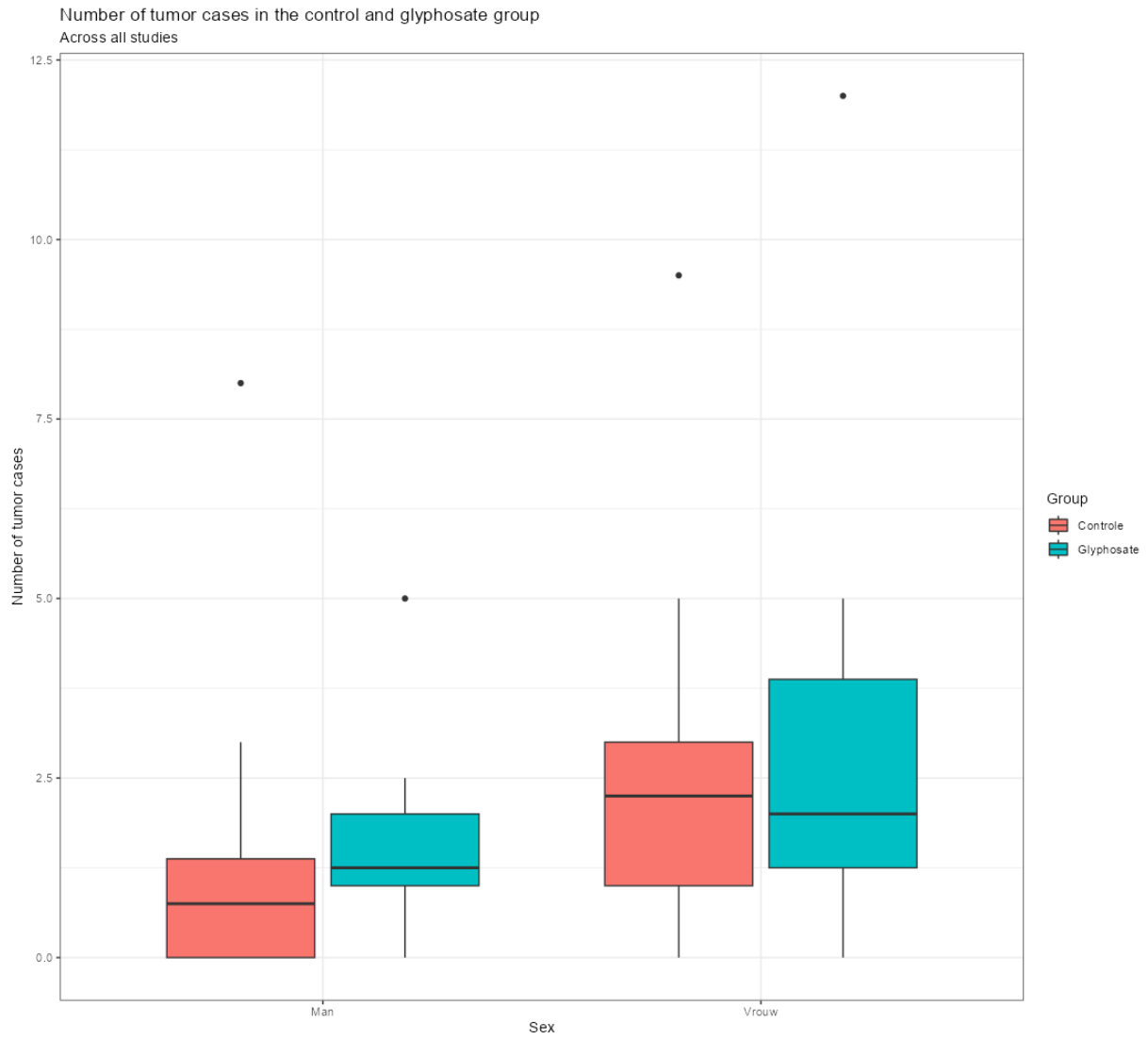
Ik wil nu dus nog een keer, en dan dieper, kijken in wat de data van de controlegroepen mij te bieden hebben. Niet alleen is elke controlegroep in principe hetzelfde (géén glyfosaat), maar het geeft ook een basisidee van wat men ongeveer aan tumoren mag verwachten in het leven van een rat of muis. Een simpele *boxplot* grafiek (**Figuur 98**) toont de hoeveelheid tumoren in de controlegroep, ongeacht studie of soort tumor, maar wel per geslacht.

Natuurlijk zegt het getal zelf niet zo heel veel, want wat als er in de glyfosaatgroepen veel meer kanker wordt gevonden? Wat we kunnen doen is alle doseringen groeperen om er zo een gemiddelde of mediaan van te nemen. Op die manier krijgen we enigszins een vergelijking tussen de controlegroep en glyfosaat, hoewel het wel een vergelijking is met veel haken en ogen. Maar om eerlijk te zijn, zitten aan bijna alle vergelijkingen in dit dossier wel haken en ogen. Op zijn minst kunnen we (proberen) de aantallen in de controlegroep enigszins (te) duiden.

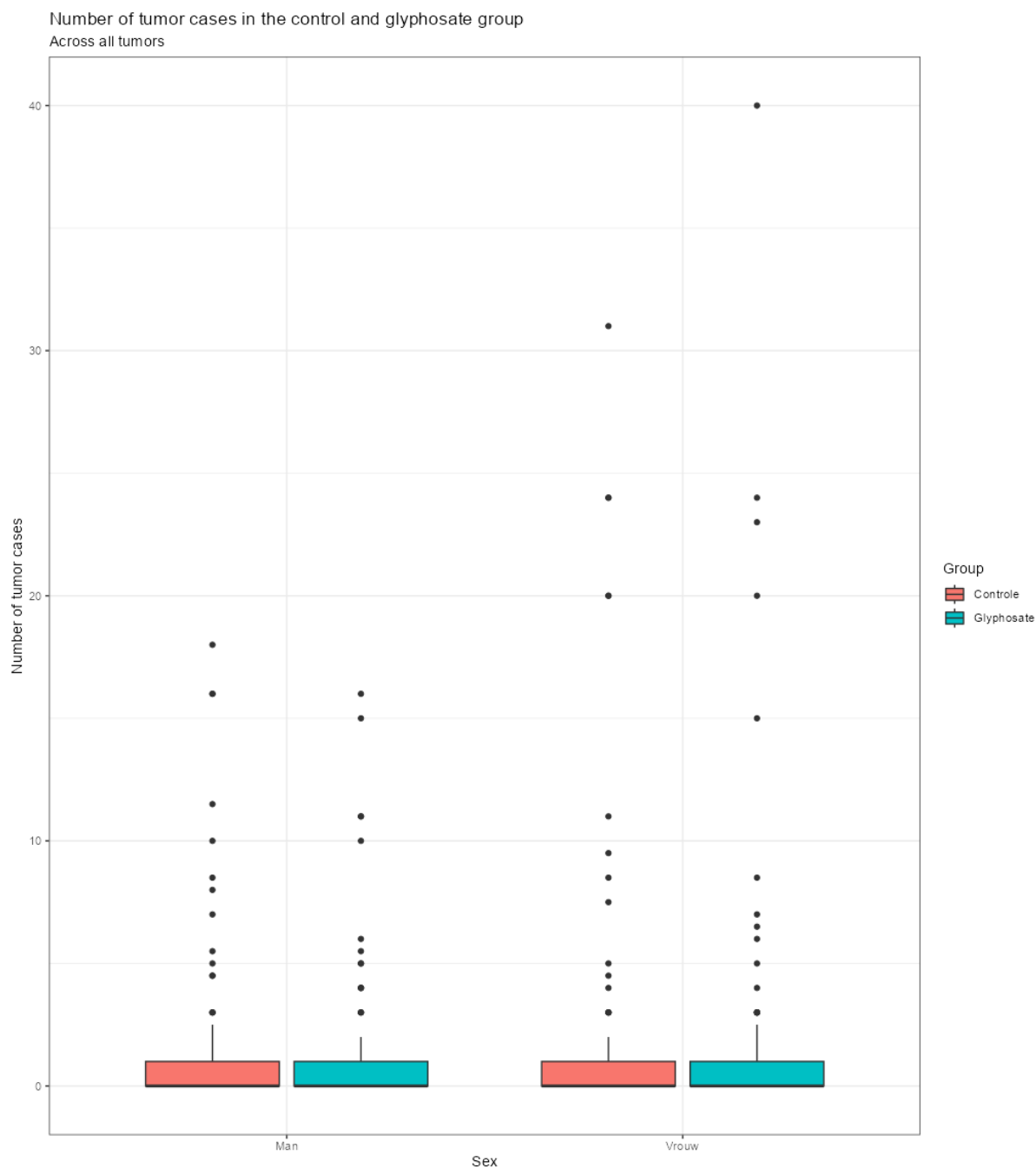


Figuur 98. Spreiding van aantal kankergevallen in de controlegroep, per geslacht, ongeacht kankersoort.

De volgende figuur laat de mediaan zien per groep én per geslacht over alle studies heen (**Figuur 99**). Ik doe het nu expres alleen over de studies heen, want als ik het over de tumoren heen doen dan krijg ik een figuur die aan interpretatie te wensen overlaat (**Figuur 100**). **Figuur 99** laat verder zien dat bij de mannetjes er een stijging lijkt te zijn, maar bij de vrouwtjes niet. Verder zien we vooral een grote spreiding wat maakt dat de stelling dat glyfosaat altijd kanker veroorzaakt zeker niet opgaat.



Figuur 99. Boxplot van het aantal tumoren dat is gevonden bij mannetjes en vrouwtjes in de controlegroep en glyfosaatgroepen (ongeacht tumor en dosering).



Figuur 100. Boxplot van het aantal tumoren dat is gevonden bij mannetjes en vrouwtjes in de controlegroep en glyfosaatgroepen (ongeacht tumor en dosering).

De relatie tussen glyfosaat en lymfoom bij dierstudies.

Gecombineerd met de bevinding dat we niet kunnen bepalen wie wanneer kanker krijgt is het eigenlijk wel tijd om de analyses af te ronden en de rest van wat we hebben gevonden te bespreken in de discussie. Maar dit voelt onbevredigend, omdat we geen getal hebben uitgerekend voor de kans op lymfoom op basis van de dierproeven. Ik denk ook niet dat we een zinvol getal kunnen uitrekenen en dat zal ik hieronder uitleggen.

Zoals we in **Figuur 79** konden zien zijn er verschillende vormen van lymfoom meegenomen door Portier óf Panzacchi. Beide studies hebben hun problemen, dus hoopte ik ze beiden te kunnen combineren om te zien wat de waarde van deze informatie zou zijn, maar het is onduidelijk waar de overlap is. Zo lijkt data voor '*malignant lymphomas*' behulpzaam, maar deze data vinden we alleen in de Portier meta-analyse en op basis van die data heb ik toentertijd geen zinvolle conclusies kunnen trekken. Ik citeer uit mijn eigen rapport:

Het lijkt erop dat een dose-response relatie in de vorm van een regressie geen relatie laat zien tussen dosering en glyfosaat, maar een binomiaal model toont uiteindelijk toch een relatie op basis van een model dat ook soort en geslacht meeneemt. De relatie is voornamelijk zichtbaar in Swiss-Albino ratten, maar daar hebben we maar één studie van. Door de bocht genomen is de kans op kanker, in die populatie, na behandeling voor glyfosaat 8% hoger bij de vrouwen en 6% hoger bij de mannen (op basis van het model). Deze bevinding staat los van de gekozen prior. Samengevat betekent dit dat dat we een traditionele dose-response analyse los moeten laten.

Dit betekent dat we alleen kunnen uitwijken naar de Panzachi studie, maar ook deze kent een hoop tekortkomingen waaronder het niet corrigeren voor een grote hoeveelheid statistische toetsen. Maar zelfs met een toegenomen kans op vals-positieven zie ik in **Tabel 5** en **Tabel 6** alleen acute lymfatische leukemie terugkomen. Dit is niet hetzelfde als non-hodgkin lymfoom. Dan hebben we nog **Figuur 89**, **Figuur 90** en **Figuur 91** om ons te laten zien dat het aantal lymfomen zowel gelijk blijft, toeneemt of afneemt in vergelijking met de controle groep. Ik kan uiteindelijk dus dezelfde bevinding maken als die ik gemaakt heb in mijn eerste rapport en dat is er dat er geen zinvolle *dose-response* relatie kan worden opgetekend. De invloed van glyfosaat op kanker (althans lymfoom) is niet eenduidig evident én dat is ook de reden dat het optekenen van een formele meta-analyse in dit rapport geen zin. Niet alleen heb ik het al geprobeerd in mijn eerste rapport, maar er blijkt ook gewoonweg geen informatie te zijn die ik kan vertalen naar de menselijke studies.

Discussie

Dit rapport beschrijft de informatie die voorhanden was in alle gepubliceerde systematische reviews en meta-analyses die keken naar de relatie tussen kanker en pesticiden bij zowel dieren (ratten en muizen) als bij mensen. Het doel was om de bevindingen bij dieren af te zetten tegen die van mensen om zo te zien of bevindingen uit dierproeven zich laten vertalen naar wat gevonden is bij mensen. Mijn zoektocht onthulde dat in muizen en ratten er geen enkele meta-analyse voorhanden is rondom pesticiden, behalve voor glyfosaat, en die meta-analyse heb ik in mijn eerste rapport in dit dossier al onder handen genomen. Met als gevolg dat dit rapport niet over alle mogelijke pesticiden kan gaan, maar uitsluitend bericht over glyfosaat. Dit is al een bijzondere bevinding. En ondanks dat glyfosaat de meeste gebruikte (en waarschijnlijk ook de meest bekende) pesticide is, betekent deze bevinding nog steeds dat dit rapport maar een klein deel van het gehele onderzoeksveld kan beslaan. Gewoonweg omdat er geen overkoepelende studies zijn die kijken naar andere pesticiden dan glyfosaat. Dit gebrek aan duidelijk data staat haaks op de 'duidelijke' berichtgeving zoals ik die in de media tegenkom. In die berichten lees ik weinig over de moeite die iemand zou moeten doen om alle beschikbare data tot zich te nemen. Noch lees ik over de onzekerheden die optreden wanneer men tracht om data om te zetten in betekenisvolle informatie. Dat proces van wikken en wegen wordt te weinig genoemd, waardoor het foutieve idee bestaat dat dit dossier glashelder is.

Analyses op basis van deze database laten duidelijk zien hoe moeilijk het is om data van verschillende reviews te combineren. Deze exacte bevinding zag ik al eerder tijdens mijn werk rondom de relatie tussen pesticiden (waaronder glyfosaat) en de ziekte van Parkinson. Ook hier verschilden de gepubliceerde reviews in hun uitvoer, kwaliteit en rapportage. Zo bleken reviews vaak te verschillen in de gehanteerde statistiek (*odds-ratio's* of *risk-ratio's*) terwijl de schatting hetzelfde was. Verder was het niet altijd even duidelijk welk deel van de originele studie men nu wel of niet had opgenomen in de meta-analyse. Hierdoor werd het haast onmogelijk om de gepubliceerde meta-analyse(s) na te bootsen. En wanneer ik de bevindingen van de dierstudies wilde combineren met de studies onder mensen bleek er weinig tot geen uniformiteit te zijn in de beschreven histologie van elke kankersoort. Met als gevolg dat de bevindingen zich maar lastig (zo niet onmogelijk) laat combineren.

Het gebrek aan gepubliceerde data en codes is een serieus gemis in dit onderzoeksveld waardoor het opnemen van alle informatie in één enkele database niet alleen een enorme exercitie werd, maar nog steeds de nodige lacunes bevat. Niet alleen blijft er data ontbreken omdat ik niet alle PDF files kon downloaden, maar de reviews verschilden soms zo sterk in hun beschrijving dat ik de gaten niet kon opvullen met informatie uit andere publicaties. In tegenstelling tot mijn eerdere werk rondom de ziekte van Parkinson heb ik er daarom deze keer maar voor gekozen om niet tot in het extreme extra sensitiviteitsanalyses uit te voeren. Niet alleen neemt de kans op vals positieven toe bij elke extra analyse, maar ik had deze keer ook veel minder data tot mijn beschikking om die extra analyses naar behoren uit te voeren. Ondanks een grote verscheidenheid aan onderzoek en onderzochte kankersoorten bleek de informatiedichtheid niet voor elke kankersoort even groot te zijn. Waardoor we dus eigenlijk niet kunnen spreken over 'kanker' maar het expliciet moeten hebben over de specifieke kankersoort waarover we iets wel of niet kunnen en willen zeggen.

Uiteindelijk is dit, bij mijn weten, wel de grootst beschikbare database rondom glyfosaat en kanker bij mens én dier. Met deze database is in een oogopslag zichtbaar welke informatie door welke review wordt gerapporteerd én welke informatie niet. Ook kunnen we nu zien in hoeverre de reviews, maar ook de studies in en tussen de reviews, het met elkaar eens zijn. Maar boven alles kunnen we nagaan in hoeverre de bevindingen bij dieren zich (enigszins) laten vertalen bij mensen.

Wat ik heb gevonden bij mensen

Vanuit vier verschillende zoektochten heb ik in totaal 9318 publicaties onderzocht waardoor ik uiteindelijk een database heb weten op te bouwen met gegevens afkomstig uit 20 reviews en 42 publicaties. In totaal werden 30 unieke kankersoorten onderzocht, maar uiteindelijk bleek er voor het gros van deze soorten te weinig data te zijn om mee te rekenen. De meest onderzochte kankersoort was non-hodgkin lymfoom. Verder was er (op het eerste gezicht) voldoende informatie beschikbaar voor beenmergkanker, leukemie en nierkanker bij kinderen. Uiteindelijk bleek alleen voor non-hodgkin lymfoom voldoende data beschikbaar om een meta-analyse te doen. Maar ook hier was het nodig om een aantal keuzes te maken

zodat we appels met appels bleven vergelijken, waardoor we uiteindelijk veel minder gegevens overhielden dan voorheen gedacht.

Een berekening op basis van zes studies en acht *odds-ratio's* (wel/geen blootstelling) toonde een verhoogd risico op kanker van ongeveer 35% met een 95% betrouwbaarheidsinterval tussen de 14% en 61%. Ook het 95% voorspellingsinterval was statistisch significant, maar wijder en haast marginaal (0.02%; 83%). Een analyse voor de mate van blootstelling (*dose-response*) op basis van drie studies en 18 schattingen toonde een *odds-ratio* van 24.1% [-0.1%; 55.7%] met een 95% voorspellingsinterval tussen de -45.3% en 281%. Dit zijn getallen op basis waarvan je geen beleid mee kunt informeren.

Wat ik heb gevonden bij dieren

Voor dieren heb ik één (al bekende) meta-analyse (Portier *et al.*) gevonden en één extra dierproef: de Ramazzini studie die in verschillende media is aangehaald als (wederom) bewijs dat glyfosaat kankerverwekkend is. In totaal werden hier meer dan 100 unieke kankersoorten onderzocht, maar de manier waarop de kanker werd beschreven maakt het uiterst lastig om deze gegevens te combineren met resultaten verkregen bij mensen. De meest onderzochte kankersoort bij mensen, non-hodgkin lymfoom, vond ik niet terug als dierlijke variant in de dierstudies.

De Ramazzini studie (Panzacchi *et al.*) kent helaas dezelfde gebreken als de meta-analyse van Portier. Zo is er niet gecorrigeerd voor het uitvoeren van een grote verscheidenheid aan statistische testen. Sterker nog, er lijkt bewust te zijn ingezet op statistiek waarvoor minder informatie nodig is dan normaal om een p-waarde uit te rekenen. Daarmee lijkt de Ramazzini studie eerder op een screeningsstudie dan op onderzoek dat beleid moet ondersteunen. Wat verder, wederom, opvalt is dat het aantal kankergevallen bij de controlegroep niet per definitie kleiner is dan in de glyfosaatgroepen. Het komt dan wel voor dat er kanker wordt gezien in de glyfosaatgroepen die helemaal niet wordt gezien in de controlegroep, maar het komt ook voor dat een bepaald soort kanker in geen enkele groep wordt gezien. Ook is het mogelijk dat er in de glyfosaatgroepen minder van een bepaald soort kanker voorkomt dan in de controlegroep. Daarmee lijken de bevindingen sterk op die van Portier en is het onmogelijk om een *dose-response* relatie op te tekenen.

De vertaalslag van dier naar mens

Het is mij niet gelukt om de resultaten bij mensen en dieren met elkaar te verenigen. Zo is voor mensen alleen voldoende informatie voor de relatie tussen glyfosaat en non-hodgkin lymfoom die onder bepaalde omstandigheden statistisch significant lijkt, maar marginaal is en op persoon niveau niet kan worden ingezet. Deze data kunnen we echter niet combineren met gegevens van de dierstudies, omdat hier non-hodgkin lymfoom niet wordt gemeten. Kijken we naar de bevindingen rondom andere lymfomen dan zie ik geen noemenswaardige relatie behalve voor leukemie, maar leukemie zien we zelden terug in de studies bij mensen. Ook nu kan ik dus geen verbinding maken. Daarbovenop blijken de studies onderling ook nog eens flink te verschillen. Het is uiteindelijk dus gewoonweg niet mogelijk om een meta-analyse uit te voeren waarin ik de relatie tussen glyfosaat en een bepaald soort kanker per onderzoeksgroep laat zien.

Beperkingen van dit onderzoek

De belangrijkste beperking in dit onderzoek is dit zelf geen *systematic review* is. Hoewel ik getracht heb mijn werk systematisch uit te voeren door al mijn stappen te documenteren én te delen, blijft het feit dat met dit onderzoek maar één onderzoeker gemoeid is. Het minimum voor een *systematic review* is twee en bij aanwezigheid van discrepantie zelfs drie.

Verder heb ik geen poging ondernomen om de kwaliteit van elke review te beoordelen. De reden om dit niet te doen, is omdat ik uit elke review de data van de opgenomen studies extraheer. Door deze tegen elkaar af te zetten kan ik mijns inziens beter beoordelen hoe kwalitatief sterk het landschap is. Zo mag worden verwacht dat als een andere review dezelfde studiegegevens tot zich neemt deze ook hetzelfde rapporteert, maar dat bleek vaak niet het geval te zijn. Verder mag gesteld worden dat ik met de inclusie van alleen meta-analyses niet alle relevante data tot mij heb genomen. Dat ik ook *systematic reviews* tot mij heb genomen heeft de database groter gemaakt, maar dat wil niet direct zeggen dat de kwaliteit van de data in de database is toegenomen.

Ten slotte heb ik geen poging gedaan om bestaande *frameworks* zoals [GRADE](#) toe te passen. Ik had er ook voor kunnen kiezen om alle data in een [GRADE software](#) omgeving in te laden en vanuit daar te bepalen of de studies kwalitatief hoogwaardig, betrouwbaar én

consistent genoeg zijn om beleid door te laten bepalen. Dit zijn stappen die ik andere onderzoekers wil aanmoedigen om te nemen zodra ze met deze data aan de slag gaan.

Het mag trouwens voor zichzelf spreken dat ik geen uitspraken doe over biologische mechanismes: dat is namelijk niet een onderzoeksveld waarvoor ik ook maar enige expertise geniet. Dit laat ik over aan de geïnteresseerde lezer.

Conclusie

Een zoektocht naar de mogelijke relatie tussen pesticiden en kanker moet zich beperken tot glyfosaat wanneer men tracht om dierstudies en studies onder mensen te combineren. Een statistische combinatie van gegevens bij mens en dier bleek uiteindelijk niet haalbaar door een gebrek aan combineerbare gegevens. Bij dieren lukt het helemaal niet om een zinvolle meta-analyse uit te voeren waardoor het ook niet mogelijk was om een relatie te leggen met de bevindingen bij mensen. Op basis van deze gegevens kan men in feite niet spreken over een relatie tussen pesticiden en kanker, maar moet men zich beperken tot glyfosaat en één bepaalde kankersoort: non-hodgkin lymfoom. Beide onderzoeksvelden (mens én dier) hebben verder te maken met serieuze methodologische en statistische tekortkomingen waardoor appels niet meer met appels vergeleken kunnen worden. Gevonden relaties lijken hierdoor eerder het resultaat van hoogst sensitieve statistische screeningsmethoden. Daarbovenop is er een serieus gebrek aan beschikbare data en codes om gepubliceerde gegevens na te rekenen.

Figuren

Figuur 1. Flow-chart eerste twee studies	26
Figuur 2. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat. Verschillende reviews hebben verschillende studies opgenomen.	31
Figuur 3. Het aantal gegevens per kolom in de database op basis van 20 reviews en 42 publicaties.	33
Figuur 4. Het aantal reviews per jaar.	34
Figuur 5. Het aantal studies gepubliceerd per jaar.	35
Figuur 6. Het aantal reviews dat naar een bepaalde risicofactor heeft gekeken.	36
Figuur 7. Het jaar waarin een review is gepubliceerd per kankersoort die is onderzocht in de review.	38
Figuur 8. Het jaar waarin een studie is gepubliceerd per kankersoort die is onderzocht in de review.	39
Figuur 9. Risico schattingen per kankersoort en of een schatting statistisch significant is of niet.	41
Figuur 10. Risicoschattingen per kankersoort (en per studie) in de review van Mink(32).	45
Figuur 11. Risicoschattingen uit de studie van De Roos(60) zoals gerapporteerd per review.	46
Figuur 12. Risicoschattingen per studie waarbij onderscheid wordt gemaakt tussen het type statistiek en type blootstelling.	48
Figuur 13. Risicoschattingen per studie waarbij onderscheid wordt gemaakt tussen wel of geen blootstelling.	49
Figuur 14. Risicoschattingen voor de studie van De Roos(60), zoals gerapporteerd per review en per type blootstelling.	51
Figuur 15. Schattingen voor de relatie tussen glyfosaat en non-hodgkin lymfoom zoals beschreven in elke review die de studie van Eriksson(64) heeft meegenomen.	52
Figuur 16. Risicoschattingen per studie en per hoeveelheid blootstelling. We kunnen zien dat er behoorlijk wat verschillen zijn in de mate van gemeten hoeveelheid blootstelling. Dit maakt het samenvoegen er niet makkelijker op.	53
Figuur 17. Risicoschattingen per type statistiek en of ze significant zijn of niet. Het gaat hier om elke individuele schatting, ongeacht studie, review, of kankersoort.	54
Figuur 18. De relatie tussen de gevonden risicoschattingen en het studie jaar waarin ze zijn gepubliceerd per soort meeting.	63
Figuur 19. De relatie tussen schattingen (odds-ratio's en risk-ratio's) en het jaar waarin de studies zijn gepubliceerd oer soort study design.	64
Figuur 20. Risicoschattingen per studie en study design waarbij de kleur aangeeft of een bevinding statistisch significant is of niet. De verticale lijnen laten de betrouwbaarheidsintervallen zien. De horizontale lijn is de lijn-van-geen-verschil.	65
Figuur 21. Risicoschattingen per studie, study design en mate van blootstelling waarbij de kleur aangeeft of een bevinding statistisch significant is of niet. De verticale lijnen laten de betrouwbaarheidsintervallen zien. De horizontale lijn is de lijn-van-geen-verschil.	65
Figuur 22. Risicoschattingen per studie en statistiek. De verticale lijn is de lijn-van-geen-verschil.	66
Figuur 23. Elke risicoschatting die is opgenomen in de database op basis van 20 reviews en 42 publicaties. De kolommen geven de specifieke statistiek weer en de kleur laat zien om welk soort study design het gaat.	67
Figuur 24. De verdeling van de odds-ratio's en risk-ratio's voor alle studies opgenomen in de database.	68
Figuur 25. De verdeling van de odds-ratio's en risk-ratio's voor alle studies opgenomen in de database per study design.	68
Figuur 26. De verdeling van de odds-ratio's en risk-ratio's per study design voor alle studies opgenomen in de database.	69
Figuur 27. Schattingen per mate van blootstelling zoals gerapporteerd in de studie van De Roos(60). De kleuren staan voor de informatie zoals aangeduid in elke review die deze studie heeft opgenomen.	70
Figuur 28. Schattingen per mate van blootstelling zoals gerapporteerd in de studie van Hardell(66). De kleuren staan voor de informatie zoals aangeduid in elke review die deze studies heeft opgenomen.	71
Figuur 29. Schattingen voor drie studies die de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL) hebben onderzocht. Per studie zien we de schattingen per review en of het duidelijk was of de schatting per geslacht is gedaan of niet.	72
Figuur 30. Risicoschattingen voor De Roos(60), Eriksson (64) en Hardell (66) zoals opgenomen in verschillende reviews, en per statistiek.	74
Figuur 31. Risicoschattingen voor De Roos(60), Eriksson(64) en Hardell(66) zoals opgenomen in verschillende reviews.	74
Figuur 32. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat op non-hodgkin lymfoom (NHL).	75

Figuur 33. Risicoschattingen voor alle case-control studies die hebben gekeken naar de invloed van glyfosaat. Ook getoond wordt de lengte van het 95% betrouwbaarheidsinterval en de totale studiegrootte (indient de data beschikbaar is).....	76
Figuur 34. Risicoschattingen voor alle case-control studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval.....	77
Figuur 35. Risicoschattingen voor alle studies die hebben gekeken naar de invloed van glyfosaat met daarbij de lengte van het 95% betrouwbaarheidsinterval.	78
Figuur 36. Risicoschattingen voor alle studies met daarbij vermeld de totale studiegrootte en de lengte van het 95% betrouwbaarheidsinterval.	79
Figuur 37. Risicoschattingen voor alle studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval per study design.	80
Figuur 38. Risicoschattingen voor alle studies afgezet tegen de lengte van het 95% betrouwbaarheidsinterval per soort statistiek.	81
Figuur 39. Risicoschattingen per studie zoals opgenomen in de reviews.	83
Figuur 40. Risicoschattingen met 95% betrouwbaarheidsintervallen per studies zoals opgenomen in elke review.....	84
Figuur 41. Risicoschattingen voor elke kankersoort in boxplot for maat.	85
Figuur 42. Risicoschattingen en lengte van het 95% betrouwbaarheidsinterval per kankersoort.	86
Figuur 43. Risicoschattingen en 95% betrouwbaarheidsintervallen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en leukemie (CLL/SLL). De kleur geeft statistische significantie aan.	87
Figuur 44. Risicoschattingen en 95% betrouwbaarheidsintervallen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en leukemie (CLL/SLL) per review opgenomen in de database. De kleur geeft de specifieke kankersoort aan.	88
Figuur 45. Risicoschattingen voor alle studies die hebben gekeken naar de relatie tussen glyfosaat en NHL of CLL/SLL of MM. De kleur geeft aan of een studie statistisch significant is of niet.	89
Figuur 46. Risicoschattingen met 95% betrouwbaarheidsintervallen per review jaar voor non-hodgkin lymfoom (NHL). De kleur geeft statistische significantie aan.	90
Figuur 47. Risicoschattingen met 95% betrouwbaarheidsinterval per studie opgenomen in de review van Acquavella(35) . Het betreft hier alleen non-hodgkin lymfoom (NHL).	91
Figuur 48. Risicoschattingen met 95% betrouwbaarheidsinterval per studie opgenomen in de review van Zhang(33). Het betreft hier alleen non-hodgkin lymfoom (NHL).	93
Figuur 49. Resultaten uit Zhang(33) zoals daar gepubliceerd. Het betreft de resultaten van twee meta-analyses: elk met een andere set aan AHS data.	94
Figuur 50. Berekening van de variatie van een odds-ratio op basis van gerapporteerde betrouwbaarheidsintervallen	94
Figuur 51. Meta-analyse voor glyfosaat en non-hodgkin lymfoom (NHL) op basis van de studies van Zhang(33). Ik probeer de eerste meta-analyse uit Zhang(33) na te bootsen door de studie van Andreotti(54) toe te voegen en die van De Roos(61) eruit te laten.	96
Figuur 52. Meta-analyse op basis van de review van Zhang, 2019. Het was de bedoeling om afbeelding A uit Figuur 49 te reproduceren, maar dat is niet helemaal gelukt.	97
Figuur 53. Risicoschattingen en de lengte van het 95% betrouwbaarheidsinterval per study design en de aan-of afwezigheid van correctiefactoren. De kleur geeft statistische significantie aan.	100
Figuur 54. Risicoschatting met 95% betrouwbaarheidsinterval per kankersoort, study design en de aan-of afwezigheid voor correctiefactoren.	101
Figuur 55. De relatie tussen de puntschatting van het effect van glyfosaat op kanker én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in odds-ratio's of risk ratio's. Twee studies (60, 67) springen eruit met een groter dan normale variatie.	106
Figuur 56. De relatie tussen de puntschatting van het effect van glyfosaat op kanker én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in odds-ratio's of risk ratio's. De kleur geeft het study design aan.	107
Figuur 57. De relatie tussen de puntschatting van een effect van pesticiden én de variatie rondom die puntschatting voor studies die hun bevindingen rapporteerden in odds-ratio's of risk ratio's. De kleur geeft de kankersoort aan.	108
Figuur 58. Schattingen en onzekerheden voor elke odds-ratio en risk-ratio die in de reviews is opgenomen. ...	110
Figuur 59. Odds-ratio's met 95% betrouwbaarheidsintervallen zoals verzameld over de reviews heen. De rode stippellijn is de lijn van geen-effect.	113

Figuur 60. Odds-ratio's met 95% betrouwbaarheidsintervallen. De grootte van elke schatting is afhankelijk van het gewicht in de meta-analyse. Zichtbaar is dat niet elke dubbeltelling eenzelfde gewicht heeft gekregen.	114
Figuur 61. Punt-schattingen van alle odds-ratio's afgezet tegen de lengte van de 95% betrouwbaarheidsintervallen. De grootte van elke schatting is afhankelijk van het gewicht in de meta-analyse. Duidelijk is dat het gewicht direct gebonden is aan de lengte van het 95% betrouwbaarheidsinterval.	115
Figuur 62. Baujat grafiek voor de meta-analyse van alle studies in de database die hebben gekeken naar de relatie tussen glyfosaat en kanker.	116
Figuur 63. Influence statistics voor de meta-analyse van alle studies met odds-ratio's die hebben gekeken naar de relatie tussen glyfosaat en kanker.	117
Figuur 64. Funnel plot van de geobserveerde studies. De plot laat zien dat de meeste bevindingen aan de rechterzijde en bovenaan de funnel zitten. Dit wijst op publication bias.	119
Figuur 65. Diagnostische plots rondom de aanwezigheid van publication bias. De figuur rechtsboven toont de funnel plot voor publication bias met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de publication bias teniet te doen.	120
Figuur 66. Risicoschattingen met 95% betrouwbaarheidsinterval voor elke studie opgenomen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM), hairy cell leukemie (HCL) en nierkanker (Wilm's tumor).	122
Figuur 67. Schattingen en varianties per schattingen voor non-hodgkin lymfoom (NHL), beenmergkanker (MM) en nierkanker (Wilm's tumor).	124
Figuur 68. De funnel plot voor publication bias met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de publication bias teniet te doen.	125
Figuur 69. Risicoschattingen voor glyfosaat en non-hodgkin lymfoom (NHL) afgezet tegen de variantieschattingen. De kleur geeft de gebruikte statistiek aan.	127
Figuur 70. Punt-schattingen en onzekerheidsschattingen voor de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL). De kleur geeft het type study design aan.	129
Figuur 71. Forest plot die de odds-ratio's laat zien voor de relatie tussen glyfosaat en non-hodgkin lymfoom (NHL) in de groep die is blootgesteld aan glyfosaat (in welke hoedanigheid dan ook).	131
Figuur 72. Baujat plot van de meta-analyse. De studies springen allemaal een beetje uit het veld, wat ook het gevolg is van het samenvoegen van weinig studies die ook nog eens verschillend zijn.	133
Figuur 73. Influence plot die laat zien welke van de 11 schattingen de grootste invloed heeft op het eindresultaat. Dat blijkt de schatting uit Cantor (57) te zijn wat we ook al zagen in Figuur 72	134
Figuur 74. Diagnostische plots rondom de aanwezigheid van publication bias. De figuur rechtsboven toont de funnel plot voor publication bias met geobserveerde gegevens (zwarte bolletjes) en gesimuleerde gegevens (witte bolletjes) in een dataset waarin gesimuleerde studies zijn toegevoegd om zo de publication bias teniet te doen.	135
Figuur 75. Odds-ratio's voor non-hodgkin lymfoom (NHL) en de verschillende maten van blootstelling uitgedrukt in (cumulatieve) dagen.	137
Figuur 76. Forest plot van een meta-analyse voor glyfosaat en non-hodgkin lymfoom (NHL) voor studies die de mate van blootstelling voor glyfosaat hebben gerapporteerd.	138
Figuur 77. Alle soorten kanker die zijn opgenomen in de review van Portier en in de Panzacchi studie. Duidelijk is het verschil in de hoeveelheid soorten kanker.	141
Figuur 78. Aantal kankersoorten zoals gezien in de Panzacchi studie bij mannetjes en vrouwtjes voor de 50 mg/kg dosering.	143
Figuur 79. Soorten lymfoom kanker en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.	146
Figuur 80. Soorten kanker die te maken hebben met leukemie en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.	146
Figuur 81. Soorten nierkanker en of deze zijn opgenomen in de Portier meta-analyse en/of de Panzacchi studie.	147
Figuur 82. Alle doseringen die zijn opgenomen. Duidelijk is dat de dosering gebruikt in de Panzacchi studie door geen enkele andere studie opgenomen in de Portier meta-analyse is gebruikt. Behalve de controle groep.	148
Figuur 83. Soort dieren dat is opgenomen. We kunnen zien dat de Panzacchi studie behoorlijk wat schattingen toevoegt aan de hoeveelheid die we al hadden vanuit de Portier meta-analyse.	148
Figuur 84. Verdeling van de veranderingen in het aantal tumoren, per tumorsoort en geslacht, wanneer vergeleken met de controlegroep.	151
Figuur 85. De verandering in het aantal tumoren, per tumor en geslacht, vergeleken met de controlegroep.	152

Figuur 86. De verandering in het aantal tumoren, per tumor en geslacht, vergeleken met de controlegroep. De dikte van de lijn geeft de frequentie aan.	153
Figuur 87. Verandering in aantal tumoren, per tumorsoort en geslacht, wanneer vergeleken met de controlegroep.	154
Figuur 88. Totaal aantal tumoren gezien per kankersoort en geslacht. Wanneer in totaal nul tumoren zijn gezien dan betekent dit dat de kankersoort wel wordt genoemd, maar niet is geobserveerd.	155
Figuur 89. Aantal gevallen van kanker voor twee soorten lymfoma per dosering.	161
Figuur 90. Aantal gevallen van kanker voor twee soorten lymfoma per dosering en geslacht.	161
Figuur 91. Verschil tussen het aantal kankergevallen in de controlegroep en de glyfosaatgroepen per studie, geslacht en dosering voor lymfoomkanker.	162
Figuur 92. Aantal tumoren gevonden per glyfosaat dosis over alle studies heen die keken naar (bij)nier kanker.	165
Figuur 93. Aantal tumoren gevonden per glyfosaat dosis en per geslacht over alle studies heen die keken naar (bij)nier kanker.	166
Figuur 94. Verschil in kankergevallen, vergeleken met de controlegroep, voor alle studies die keken naar (bij)nierkanker voor zowel mannetjes als vrouwtjes.	167
Figuur 95. Aantal tumoren gevonden per glyfosaat dosis over alle studies heen die keken naar schildklierkanker.	169
Figuur 96. Aantal tumoren gevonden per glyfosaat dosis en per geslacht over alle studies heen die keken naar schildklierkanker.	170
Figuur 97. Verschil in kankergevallen, vergeleken met de controlegroep, voor alle studies die keken naar schildklierkanker voor zowel mannetjes als vrouwtjes.	171
Figuur 98. Spreiding van aantal kankergevallen in de controlegroep, per geslacht, ongeacht kankersoort.	172
Figuur 99. Boxplot van het aantal tumoren dat is gevonden bij mannetjes en vrouwtjes in de controlegroep en glyfosaatgroepen (ongeacht tumor en dosering).	173
Figuur 100. Boxplot van het aantal tumoren dat is gevonden bij mannetjes en vrouwtjes in de controlegroep en glyfosaatgroepen (ongeacht tumor en dosering).	174

Tabellen

Tabel 1. Summary estimates voor Zhang(33) zoals gerapporteerd in de review én zoals berekend door mij. ...	98
Tabel 2. De schattingen rondom de heterogeniteit van de meta-analyse zoals gerapporteerd door twee statistische pakketten voor het statistiek programma R: metafor en meta.	104
Tabel 3. Summary estimate en prediction interval voor het risico op kanker door glyfosaat berekend vanuit beide pakketten voor gerapporteerde odds-ratio's.	121
Tabel 4. Gevallen per kankersoort en geslacht gezien in de Panzacchi studie.	150
Tabel 5. Alle statistisch significante verbanden per unieke kankersoort zoals gerapporteerd in Panzacchi, voor mannetjes en vrouwtjes apart.	156
Tabel 6. Statistisch significante resultaten zoals beschreven in de Panzacchi studie – nu per analysesoort.	157

Referenties

1. Portier CJ. A comprehensive analysis of the animal carcinogenicity data for glyphosate from chronic exposure rodent carcinogenicity studies. *Environ Health.* 2020;19(1):18.
2. Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health.* 2002;56(8):588-94.
3. Panzacchi S, Tibaldi E, De Angelis L, Falcioni L, Giovannini R, Gnudi F, et al. Carcinogenic effects of long-term exposure from prenatal life to glyphosate and glyphosate-based herbicides in Sprague-Dawley rats. *Environ Health.* 2025;24(1):36.
4. Panzacchi S, Mandrioli D, Manservigi F, Bua L, Falcioni L, Spinaci M, et al. The Ramazzini Institute 13-week study on glyphosate-based herbicides at human-equivalent dose in Sprague Dawley rats: study design and first in-life endpoints evaluation. *Environ Health.* 2018;17(1):52.
5. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.

6. Stroup DF. Meta-analysis of Observational Studies in Epidemiology A Proposal for Reporting. *JAMA*. 2000;283(15):2008-15.
7. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *Bmj*. 1997;315(7121):1533-7.
8. Egger M, Smith GD. Misleading meta-analysis. *Bmj*. 1995;310(6982):752-4.
9. Eysenck HJ. Meta-analysis and its problems. *Bmj*. 1994;309(6957):789-92.
10. Plumb AA, Halligan S, Mallett S. The choice and definition of summary measure for meta-analysis of clinical studies with binary outcomes: effect on clinical interpretation. *Br J Radiol*. 2020;93(1108):20190976.
11. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat Disord*. 2016;23:1-9.
12. Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. *BMJ Med*. 2022;1(1):e000071.
13. Uttley L, Quintana DS, Montgomery P, Carroll C, Page MJ, Falzon L, et al. The problems with systematic reviews: a living systematic review. *J Clin Epidemiol*. 2023;156:30-41.
14. Bae JM. Meta-epidemiology. *Epidemiol Health*. 2014;36:e2014019.
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj*. 2009;339:b2700.
16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Bmj*. 2015;350:g7647.
17. Sheehan MC, Lam J. Use of Systematic Review and Meta-Analysis in Environmental Health Epidemiology: a Systematic Review and Comparison with Guidelines. *Curr Environ Health Rep*. 2015;2(3):272-83.
18. Young SS, Acharjee MK, Das K. The reliability of an environmental epidemiology meta-analysis, a case study. *Regul Toxicol Pharmacol*. 2019;102:47-52.
19. Blair A, Burg J, Foran J, Gibb H, Greenland S, Morris R, et al. Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. *Regul Toxicol Pharmacol*. 1995;22(2):189-97.
20. Zocchetti C, Consonni D, Bertazzi PA. Relationship between prevalence rate ratios and odds ratios in cross-sectional studies. *Int J Epidemiol*. 1997;26(1):220-3.
21. Kerr S, Greenland S, Jeffrey K, Millington T, Bedston S, Ritchie L, et al. Understanding and reporting odds ratios as rate-ratio estimates in case-control studies. *J Glob Health*. 2023;13:04101.
22. Persoskie A, Ferrer RA. A Most Odd Ratio:: Interpreting and Describing Odds Ratios. *Am J Prev Med*. 2017;52(2):224-8.
23. Viera AJ. Odds ratios and risk ratios: what's the difference and why does it matter? *South Med J*. 2008;101(7):730-4.
24. Andrade C. Understanding relative risk, odds ratio, and related terms: as simple as it can get. *J Clin Psychiatry*. 2015;76(7):e857-61.
25. Alavi M, Hunt GE, Visentin DC, Watson R, Thapa DK, Cleary M. Using risk and odds ratios to assess effect size for meta-analysis outcome measures. *J Adv Nurs*. 2020;76(12):3231-4.
26. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *Bmj*. 2007;334(7589):349-51.
27. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Bmj*. 2003;327(7414):557-60.
28. Hedges LV. Effect sizes for experimental research. *Br J Math Stat Psychol*. 2026;79(1):31-45.
29. Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*. 2010;39(4):860-4.
30. Matthay EC, Hagan E, Gottlieb LM, Tan ML, Vlahov D, Adler N, Glymour MM. Powering population health research: Considerations for plausible and actionable effect sizes. *SSM Popul Health*. 2021;14:100789.
31. Friese M, Frankenbach J. p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychol Methods*. 2020;25(4):456-71.
32. Mink PJ, Mandel JS, Scurman BK, Lundin JI. Epidemiologic studies of glyphosate and cancer: a review. *Regul Toxicol Pharmacol*. 2012;63(3):440-52.

33. Zhang L, Rana I, Shaffer RM, Taioli E, Sheppard L. Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: A meta-analysis and supporting evidence. *Mutat Res Rev Mutat Res.* 2019;781:186-206.
34. Kabat GC, Price WJ, Tarone RE. On recent meta-analyses of exposure to glyphosate and risk of non-Hodgkin's lymphoma in humans. *Cancer Causes Control.* 2021;32(4):409-14.
35. Acquavella J. Epidemiologic studies of glyphosate and non-Hodgkin's lymphoma: A review with consideration of exposure frequency, systemic dose, and study quality. *Glob Epidemiol.* 2023;5:100101.
36. Chang ET, Delzell E. Systematic review and meta-analysis of glyphosate exposure and risk of lymphohematopoietic cancers. *J Environ Sci Health B.* 2016;51(6):402-34.
37. Acquavella J, Garabrant D, Marsh G, Sorahan T, Weed DL. Glyphosate epidemiology expert panel review: a weight of evidence systematic review of the relationship between glyphosate exposure and non-Hodgkin's lymphoma or multiple myeloma. *Crit Rev Toxicol.* 2016;46(sup1):28-43.
38. Alexander DD, Mink PJ, Adami HO, Chang ET, Cole P, Mandel JS, Trichopoulos D. The non-Hodgkin lymphomas: a review of the epidemiologic literature. *Int J Cancer.* 2007;120 Suppl 12:1-39.
39. Weisenburger DD. A Review and Update with Perspective of Evidence that the Herbicide Glyphosate (Roundup) is a Cause of Non-Hodgkin Lymphoma. *Clin Lymphoma Myeloma Leuk.* 2021;21(9):621-30.
40. Donato F, Pira E, Ciocan C, Boffetta P. Exposure to glyphosate and risk of non-Hodgkin lymphoma and multiple myeloma: an updated meta-analysis. *Med Lav.* 2020;111(1):63-73.
41. Boffetta P, Ciocan C, Zunarelli C, Pira E. Exposure to glyphosate and risk of non-Hodgkin lymphoma: an updated meta-analysis. *Med Lav.* 2021;112(3):194-9.
42. Khan A, Feulefack J, Sergi CM. Exposure to pesticides and pediatric Wilms' tumor. A meta-analysis on pre-conception and pregnancy parental exposure with an IARC/WHO commentary. *Hum Exp Toxicol.* 2022;41:9603271221136211.
43. Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. *Int J Environ Res Public Health.* 2014;11(4):4449-527.
44. Matich EK, Laryea JA, Seely KA, Stahr S, Su LJ, Hsu PC. Association between pesticide exposure and colorectal cancer risk and incidence: A systematic review. *Ecotoxicol Environ Saf.* 2021;219:112327.
45. Varghese JV, Sebastian EM, Iqbal T, Tom AA. Pesticide applicators and cancer: a systematic review. *Rev Environ Health.* 2021;36(4):467-76.
46. Stanganelli I, De Felici MB, Mandel VD, Caini S, Raimondi S, Corso F, et al. The association between pesticide use and cutaneous melanoma: a systematic review and meta-analysis. *J Eur Acad Dermatol Venereol.* 2020;34(4):691-708.
47. Goodman JE, Loftus CT, Zu K. 2,4-Dichlorophenoxyacetic acid and non-Hodgkin's lymphoma: results from the Agricultural Health Study and an updated meta-analysis. *Ann Epidemiol.* 2017;27(4):290-2.e5.
48. Silva JF, Mattos IE, Luz LL, Carmo CN, Aydos RD. Exposure to pesticides and prostate cancer: systematic review of the literature. *Rev Environ Health.* 2016;31(3):311-27.
49. Kim TH, Park SC, Kim JE, Yeon HJ, Kim JH, Park YS, et al. Exposure assessment for pesticide residues in agricultural products consumed in the Republic of Korea during 2016-2020. *Environ Sci Pollut Res Int.* 2023;30(2):4917-33.
50. Leon ME, Schinasi LH, Lebaillly P, Beane Freeman LE, Nordby KC, Ferro G, et al. Pesticide use and risk of non-Hodgkin lymphoid malignancies in agricultural cohorts from France, Norway and the USA: a pooled analysis from the AGRICOH consortium. *Int J Epidemiol.* 2019;48(5):1519-35.
51. Choudhary L, Monaghan M, Schweppe R, Franco AT, Goldner W, van Gerwen M. What Is the Impact of Glyphosate on the Thyroid? An Updated Review. *Biomedicines.* 2025;13(10).
52. Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, et al. Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. *Am J Epidemiol.* 2003;157(9):800-14.
53. Andreotti G, Freeman LE, Hou L, Coble J, Rusiecki J, Hoppin JA, et al. Agricultural pesticide use and pancreatic cancer risk in the Agricultural Health Study Cohort. *Int J Cancer.* 2009;124(10):2495-500.
54. Andreotti G, Koutros S, Hofmann JN, Sandler DP, Lubin JH, Lynch CF, et al. Glyphosate Use and Cancer Incidence in the Agricultural Health Study. *J Natl Cancer Inst.* 2018;110(5):509-16.

55. Brown LM, Blair A, Gibson R, Everett GD, Cantor KP, Schuman LM, et al. Pesticide exposures and other agricultural risk factors for leukemia among men in Iowa and Minnesota. *Cancer Res.* 1990;50(20):6585-91.
56. Brown LM, Burmeister LF, Everett GD, Blair A. Pesticide exposures and multiple myeloma in Iowa men. *Cancer Causes Control.* 1993;4(2):153-6.
57. Cantor KP, Blair A, Everett G, Gibson R, Burmeister LF, Brown LM, et al. Pesticides and other agricultural risk factors for non-Hodgkin's lymphoma among men in Iowa and Minnesota. *Cancer Res.* 1992;52(9):2447-55.
58. Carreón T, Butler MA, Ruder AM, Waters MA, Davis-King KE, Calvert GM, et al. Gliomas and farm pesticide exposure in women: the Upper Midwest Health Study. *Environ Health Perspect.* 2005;113(5):546-51.
59. Cocco P, Satta G, Dubois S, Pili C, Pilleri M, Zucca M, et al. Lymphoma risk and occupational exposure to pesticides: results of the Epilymph study. *Occup Environ Med.* 2013;70(2):91-8.
60. De Roos AJ, Blair A, Rusiecki JA, Hoppin JA, Svec M, Dosemeci M, et al. Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. *Environ Health Perspect.* 2005;113(1):49-54.
61. De Roos AJ, Zahm SH, Cantor KP, Weisenburger DD, Holmes FF, Burmeister LF, Blair A. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occup Environ Med.* 2003;60(9):E11.
62. Dennis LK, Lynch CF, Sandler DP, Alavanja MC. Pesticide use and cutaneous melanoma in pesticide applicators in the agricultural health study. *Environ Health Perspect.* 2010;118(6):812-7.
63. Engel LS, Hill DA, Hoppin JA, Lubin JH, Lynch CF, Pierce J, et al. Pesticide use and breast cancer risk among farmers' wives in the agricultural health study. *Am J Epidemiol.* 2005;161(2):121-35.
64. Eriksson M, Hardell L, Carlberg M, Akerman M. Pesticide exposure as risk factor for non-Hodgkin lymphoma including histopathological subgroup analysis. *Int J Cancer.* 2008;123(7):1657-63.
65. Flower KB, Hoppin JA, Lynch CF, Blair A, Knott C, Shore DL, Sandler DP. Cancer risk and parental pesticide application in children of Agricultural Health Study participants. *Environ Health Perspect.* 2004;112(5):631-5.
66. Hardell L, Eriksson M, Nordstrom M. Exposure to pesticides as risk factor for non-Hodgkin's lymphoma and hairy cell leukemia: pooled analysis of two Swedish case-control studies. *Leuk Lymphoma.* 2002;43(5):1043-9.
67. Hardell L, Eriksson M. A case-control study of non-Hodgkin lymphoma and exposure to pesticides. *Cancer.* 1999;85(6):1353-60.
68. Hohenadel K, Harris SA, McLaughlin JR, Spinelli JJ, Pahwa P, Dosman JA, et al. Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. *Int J Environ Res Public Health.* 2011;8(6):2320-30.
69. Kachuri L, Demers PA, Blair A, Spinelli JJ, Pahwa M, McLaughlin JR, et al. Multiple pesticide exposures and the risk of multiple myeloma in Canadian men. *Int J Cancer.* 2013;133(8):1846-58.
70. Karunanayake CP, Spinelli JJ, McLaughlin JR, Dosman JA, Pahwa P, McDuffie HH. Hodgkin lymphoma and pesticides exposure in men: a Canadian case-control study. *J Agromedicine.* 2012;17(1):30-9.
71. Kaufman DW, Anderson TE, Issaragrisil S. Risk factors for leukemia in Thailand. *Ann Hematol.* 2009;88(11):1079-88.
72. Lee WJ, Colt JS, Heineman EF, McComb R, Weisenburger DD, Lijinsky W, Ward MH. Agricultural pesticide use and risk of glioma in Nebraska, United States. *Occup Environ Med.* 2005;62(11):786-92.
73. Lee WJ, Lijinsky W, Heineman EF, Markin RS, Weisenburger DD, Ward MH. Agricultural pesticide use and adenocarcinomas of the stomach and oesophagus. *Occup Environ Med.* 2004;61(9):743-9.
74. Lee WJ, Cantor KP, Berzofsky JA, Zahm SH, Blair A. Non-Hodgkin's lymphoma among asthmatics exposed to pesticides. *Int J Cancer.* 2004;111(2):298-302.
75. Lee WJ, Sandler DP, Blair A, Samanic C, Cross AJ, Alavanja MC. Pesticide use and colorectal cancer risk in the Agricultural Health Study. *Int J Cancer.* 2007;121(2):339-46.
76. McDuffie HH, Pahwa P, McLaughlin JR, Spinelli JJ, Fincham S, Dosman JA, et al. Non-Hodgkin's lymphoma and specific pesticide exposures in men: cross-Canada study of pesticides and health. *Cancer Epidemiol Biomarkers Prev.* 2001;10(11):1155-63.
77. Meloni F, Satta G, Padoan M, Montagna A, Pilia I, Argiolas A, et al. Occupational exposure to glyphosate and risk of lymphoma: results of an Italian multicenter case-control study. *Environ Health.* 2021;20(1):49.

78. Nordström M, Hardell L, Linde A, Schloss L, Näsman A. Elevated antibody levels to Epstein-Barr virus antigens in patients with hairy cell leukemia compared to controls in relation to exposure to pesticides, organic solvents, animals, and exhausts. *Oncol Res.* 1999;11(11-12):539-44.
79. Omidakhsh N, Heck JE, Cockburn M, Ling C, Hershman JM, Harari A. Thyroid Cancer and Pesticide Use in a Central California Agricultural Area: A Case Control Study. *J Clin Endocrinol Metab.* 2022;107(9):e3574-e82.
80. Orsi L, Delabre L, Monnereau A, Delval P, Berthou C, Fenaux P, et al. Occupational exposure to pesticides and lymphoid neoplasms among men: results of a French case-control study. *Occup Environ Med.* 2009;66(5):291-8.
81. Pahwa M, Beane Freeman LE, Spinelli JJ, Blair A, McLaughlin JR, Zahm SH, et al. Glyphosate use and associations with non-Hodgkin lymphoma major histological sub-types: findings from the North American Pooled Project. *Scand J Work Environ Health.* 2019;45(6):600-9.
82. Pahwa P, Karunanayake CP, Dosman JA, Spinelli JJ, McDuffie HH, McLaughlin JR. Multiple myeloma and exposure to pesticides: a Canadian case-control study. *J Agromedicine.* 2012;17(1):40-50.
83. Presutti R, Harris SA, Kachuri L, Spinelli JJ, Pahwa M, Blair A, et al. Pesticide exposures and the risk of multiple myeloma in men: An analysis of the North American Pooled Project. *Int J Cancer.* 2016;139(8):1703-14.
84. Ruder AM, Waters MA, Butler MA, Carreón T, Calvert GM, Davis-King KE, et al. Gliomas and farm pesticide exposure in men: the upper midwest health study. *Arch Environ Health.* 2004;59(12):650-7.
85. Sharpe CR, Franco EL, de Camargo B, Lopes LF, Barreto JH, Johnsson RR, Mauad MA. Parental exposures to pesticides and risk of Wilms' tumor in Brazil. *Am J Epidemiol.* 1995;141(3):210-7.
86. Sorahan T. Multiple myeloma and glyphosate use: a re-analysis of US Agricultural Health Study (AHS) data. *Int J Environ Res Public Health.* 2015;12(2):1548-59.
87. Tsai J, Kaye WE, Bove FJ. Wilms' tumor and exposures to residential and occupational hazardous chemicals. *Int J Hyg Environ Health.* 2006;209(1):57-64.
88. Cooney MA, Daniels JL, Ross JA, Breslow NE, Pollock BH, Olshan AF. Household pesticides and the risk of Wilms tumor. *Environ Health Perspect.* 2007;115(1):134-7.
89. Nordström M, Hardell L, Magnuson A, Hagberg H, Rask-Andersen A. Occupational exposures, animal exposure and smoking as risk factors for hairy cell leukaemia evaluated in a case-control study. *Br J Cancer.* 1998;77(11):2048-52.
90. Kim J, Leon ME, Schinasi LH, Baldi I, Lebailly P, Freeman LEB, et al. Exposure to pesticides and risk of Hodgkin lymphoma in an international consortium of agricultural cohorts (AGRICOH). *Cancer Causes Control.* 2023;34(11):995-1003.
91. Schüz J, Kaletsch U, Meinert R, Kaatsch P, Michaelis J. High-birth weight and other risk factors for Wilms tumour: results of a population-based case-control study. *Eur J Pediatr.* 2001;160(6):333-8.
92. Lee WJ, Blair A, Hoppin JA, Lubin JH, Rusiecki JA, Sandler DP, et al. Cancer incidence among pesticide applicators exposed to chlorpyrifos in the Agricultural Health Study. *J Natl Cancer Inst.* 2004;96(23):1781-9.
93. Woodward M. *Epidemiology: study design and data analysis*: CRC press; 2013.
94. Altman DG, Bland JM. How to obtain the confidence interval from a P value. *Bmj.* 2011;343.
95. Sedgwick P. Understanding confidence intervals. *BMJ.* 2014;349.
96. Morris JA, Gardner MJ. Statistics in medicine: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *British medical journal (Clinical research ed).* 1988;296(6632):1313.
97. Altman DG, Bland JM. How to obtain the P value from a confidence interval. *Bmj.* 2011;343.
98. Breslow N, Olshan A, Beckwith JB, Green DM. Epidemiology of Wilms tumor. *Medical and pediatric oncology.* 1993;21(3):172-81.
99. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods.* 2010;1(2):97-111.
100. Jackson D, Law M, Rücker G, Schwarzer G. The Hartung-Knapp modification for random-effects meta-analysis: A useful refinement but are there any residual concerns? *Stat Med.* 2017;36(25):3923-34.
101. van Aert RCM, Jackson D. A new justification of the Hartung-Knapp method for random-effects meta-analysis based on weighted least squares regression. *Res Synth Methods.* 2019;10(4):515-27.
102. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics.* 1999;55(3):732-7.

103. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in medicine*. 2002;21(18):2641-52.
104. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Bmj*. 1997;315(7109):629-34.
105. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *Journal of clinical epidemiology*. 2000;53(2):207-16.
106. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56(2):455-63.
107. IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open*. 2016;6(7):e010247.
108. Roth JV. Prediction interval analysis is underutilized and can be more helpful than just confidence interval analysis. *Journal of clinical monitoring and computing*. 2009;23(3):181.
109. Chiolero A, Santschi V, Burnand B, Platt RW, Paradis G. Meta-analyses: with confidence or prediction intervals? *European journal of epidemiology*. 2012;27(10):823-5.
110. Guolo A, Varin C. Random-effects meta-analysis: the number of studies matters. *Statistical methods in medical research*. 2017;26(3):1500-18.
111. Schaarschmidt F, Sill M, Hothorn LA. Poly-k-trend tests for survival adjusted analysis of tumor rates formulated as approximate multiple contrast test. *Journal of Biopharmaceutical Statistics*. 2008;18(5):934-48.