

Leidraad met betrekking tot statistische aspecten bij een protocol

METC Erasmus MC, 6 Maart 2014

De METC toetst alle aangeboden onderzoeksvoorstellen aan de WMO en mede aan ICH/GCP richtlijnen. Dit ongeacht of het een geneesmiddelen onderzoek is, of anderszins. Hieronder worden enkele onderwerpen besproken die vaak onvoldoende duidelijk in de protocollen worden weergegeven. Mogelijk zijn de besproken punten nuttig bij het schrijven van een protocol, ook al betreft het geen Randomised Controlled Trial (RCT). Het is uitdrukkelijk niet de bedoeling van deze leidraad om alle statistische aspecten van onderzoeken uitputtend te beschrijven. Daarover zijn er voldoende naslagwerken beschikbaar. Deze leidraad gaat meer over hoe het gebruik van de statistische methoden en technieken dient te worden gedocumenteerd in het onderzoeksprotocol. Er worden enige ICH/GCP guidelines genoemd in de tekst. Deze zijn te vinden op de EMA website (onder "document search"): <http://www.ema.europa.eu/>

Randomisatie

- *Concealed allocation*: Een zeer belangrijk aspect bij gerandomiseerd onderzoek is hoe de volgende interventietoewijzing vooraf "verborgen" is. Dit om te voorkomen dat voorkennis over de interventietoewijzing op enigerlei wijze de selectie of inclusie van proefpersonen kan beïnvloeden (zie Schulz KF, Grimes DA. "Allocation concealment in randomised trials". Lancet 2002;359:614-618). De manier waarop dit gerealiseerd wordt, dient in het protocol te worden aangegeven. In het algemeen zal een externe partij de randomisatiesequentie op een deskundige wijze dienen vast te stellen, en deze volgorde dient geheim te blijven voor alle medewerkers die op enigerlei wijze betrokken zijn bij de selectie of inclusie van de proefpersonen. In het geval van een open studie mag pas na het ondertekenen van het informed consent formulier door de proefpersoon of gemachtigde de betreffende randomisatieuitslag bekend worden.
- *Geblokke/gestratificeerde randomisatie*: Het is wenselijk de randomisatie geblokt in de tijd uit te voeren. Om de onvoorspelbaarheid te handhaven dient de blok grootte geheim te blijven voor de onderzoekers. Bij een multicenter trial dient in het algemeen een randomisatie met stratificatie op centrum te worden gedaan. Centrale randomisatie heeft daarbij de voorkeur boven het werken met een systeem van gesloten enveloppen. Bij een groot aantal stratificatiefactoren is een simpele gestratificeerde randomisatie niet goed mogelijk. "Minimisatie" kan dan worden overwogen. Hierbij worden nieuwe patiënten zodanig aan een interventie toegewezen dat de relevante baselinefactoren steeds zo goed mogelijk gebalanceerd zijn tussen de interventie groepen (zie Treasure T, MacRae KD. "Minimisation: the platinum standard for trials?" BMJ 1998;317:362-363). Dit kan alleen met centrale randomisatie op een computer. Op het internet zijn er diverse sites die deze mogelijkheid tot "online" randomisatie bieden.

Powerberekening

- De METC wil altijd een *onderbouwing* zien van de onderzoeksgrootte. Voor een enkele groep van proefpersonen kan dit bijvoorbeeld zijn op grond van een gewenste voldoende smalle

- breedte van het betrouwbaarheidsinterval van de primaire uitkomst, bijvoorbeeld een prevalentie, incidentie of een correlatiecoëfficiënt en dergelijke. De gehanteerde onderzoeksgrootte berekening moet overeenkomen met de in de statistische methoden vermelde primaire analyse. Uitsluitend de haalbaarheid van de studie kan in uitzonderlijke gevallen een overweging zijn bij het vaststellen van het aantal proefpersonen.
- Bij vergelijkend onderzoek dient de studiegrootte in het algemeen vastgesteld te worden met een *powerberekening* op basis van het veronderstelde, dan wel klinisch relevant geachte, verschil van de primaire uitkomst tussen de groepen. In het geval van een meer zachte primaire uitkomstvariabele, bijvoorbeeld Quality of Life, dient de powerberekening uit te gaan van de “*Minimal Important Difference*” (MID), indien deze beschikbaar is. Als een “effect size” wordt gebruikt bij de berekening dient de klinische relevantie daarvan aangegeven te worden. Er dient in het algemeen te worden uitgegaan van een twee-zijdige alpha. Dat is omdat de conclusie “geen verschil aangetoond” op basis van een eenzijdige toets slechts een halve waarheid is als een tweezijdige toetsing wel een significant verschil, maar onverhoopt in de “verkeerde” richting, zou opleveren. De meeste medische wetenschappelijke bladen accepteren om die reden alleen tweezijdige p-waarden. Naast de te hanteren toets, alpha en power (of beta), dienen alle veronderstellingen omtrent parameter uitkomsten, en de herkomst daarvan, in detail te worden vermeld in het protocol.
 - Bij de vergelijking van twee groepen met een dichotome uitkomst dient de power-berekening te worden gedaan aan de hand van de continuïteits gecorrigeerde chi-kwadraat toets, of eventueel de Fisher exact test in geval van zeer kleine aantallen.
 - Een powerberekening bij een “*non-inferiority*” of “*equivalence*” studie verloopt anders dan bij een “*superiority*” trial. Bij een superiority studie is het doel om aan te tonen dat het verschil ongelijk aan nul is (of een ratio ongelijk aan 1). Bij een non-inferiority studie is het doel om een klinisch relevant nadeel (een gedefinieerde “non-inferiority marge”) bij de experimentele groep redelijkerwijs uit te sluiten. Dit omdat er evidente voordelen verbonden zijn aan de experimentele behandeling, bijvoorbeeld een geringere invasiviteit of lagere kosten. Voor de powerberekening bij een dergelijke studie is het wellicht raadzaam om daarover advies in te winnen bij een ter zake deskundige statisticus. Voor nadere informatie over non-inferiority studies, zoals het belang van een *per-protocol* analyse ten opzichte van een *intention-to-treat* analyse die gebruikelijk is bij superiority trials, zie Piaggio G et al. “*Reporting of non-inferiority and equivalence randomized trials: an extension of the CONSORT statement*”. JAMA 2006; 295(10):1152-1160.

Multipliciteiten in de vraagstelling

- In geval van een enkele evaluatie van een enkele primaire variabele bij het vergelijken van slechts twee groepen bestaat er geen multipliciteit in de vraagstelling en zijn er in principe geen problemen bij de statistische evaluatie. In alle andere gevallen (meer studiearmen, meer dan een enkele primaire uitkomst variabele, meerdere tijdstippen van meting of interim-analyses) dient te worden aangegeven op welke wijze ervoor zal worden gezorgd dat het percentage “vals positieve” bevindingen beneden de gestelde grens voor alpha (gebruikelijk 5%) blijft. Er zijn verschillende *multipliciteitscorrecties* met betrekking tot alpha. De eenvoudigste is de Bonferroni correctie. Er zijn weliswaar betere correcties, maar die hebben als nadeel dat zij lastig te gebruiken zijn bij de power berekeningen. Een te hanteren correctie van alpha dient ook te worden meegenomen bij het vaststellen van de power. Bijvoorbeeld, bij

- het voorhanden zijn van 2 primaire uitkomsten is voor ieder eindpunt de alpha gelijk aan 2.5% (=5% /2) met de Bonferroni correctie. De twee afzonderlijke powerberekeningen dienen beide deze aangepaste alpha te hanteren, wat tot een groter aantal proefpersonen zal leiden dan zonder correctie. Een soortgelijke redenering geldt voor een meer-armige studie waarin meerdere paarsgewijze vergelijkingen relevant geacht worden. De METC vindt het meestal niet nodig dat bij secundaire eindpunten voor multiplicitéit wordt gecorrigeerd.
- *Herhaalde metingen*: Bij veel studies vinden er na de baselinemeting op gezette tijden follow-up metingen plaats van de primaire variabele. Zonder nadere specificering kan zo het aantal primaire uitkomsten, zoals de diverse veranderingen ten opzichte van baseline, fors toenemen. Dat is niet gewenst. In een dergelijk geval dient in het protocol te worden aangegeven welke van de follow-up uitkomsten als primair zal gelden. Geldt bijvoorbeeld alleen een van de eerste follow-up metingen als primair, en zijn de latere vooral bedoeld als secundaire eindpunten om te bezien of het effect aanhoudt? Of moet het effect opbouwen of stabiliseren, en dient alleen een van de latere metingen als primaire uitkomst te worden beschouwd? Een alternatief kan bijvoorbeeld ook zijn om de “Area under the Curve” te kiezen als het enkele primaire eindpunt, indien relevant. Dit zijn verschillende manieren om de vraagstelling aan te scherpen en de multiplicitéits-problematiek te omzeilen. Belangrijk is dat in het protocol duidelijk wordt aangegeven hoe de prioriteit en de analysestrategie zullen zijn bij het evalueren van de herhaalde metingen. Dit betekent dat bij een geplande Repeated Measurements of Mixed Model Anova duidelijk wordt aangegeven wat het enkele primaire contrast zal zijn bij de evaluatie.
 - Uitgebreidere informatie over de multiplicitéitsproblematiek bij patiëntgebonden onderzoek is te vinden op de EMA website in het document genaamd “*Points to consider on multiplicity issues in clinical trials*”.

Cross-over trials

- Een cross-over studie kan alleen tot betrouwbare conclusies leiden als het op voorhand aannemelijk is dat het effect in een periode niet beïnvloed kan worden door de behandeling in de voorgaande periode. Dit punt dient in het protocol te worden besproken. Hierbij moet worden opgemerkt dat het achteraf toetsen op een *carry-over effect* meestal niet goed mogelijk is omdat deze toets relatief grote aantallen vereist voor een zinvolle toepassing. In de statistische paragraaf dient bij dit type onderzoek te worden aangegeven hoe met eventuele periode-effecten zal worden omgegaan. In het algemeen zijn cross-over studies bij patiënten alleen mogelijk bij symptoombestrijdende interventies en bij farmacokinetiek onderzoek.

Statistische paragraaf

- In de statistische paragraaf dient elke vraagstelling te worden uitgewerkt wat betreft de analysestrategie. Een algemene samenvattende bespreking in technische termen, bijvoorbeeld als “Categorical data data will be analysed using the chi-square test and continuous data will be analysed using the t-test or Mann-Whitney test, as appropriate”, vindt de METC onvoldoende. Een dergelijke korte beschrijving is mogelijk acceptabel in een artikel, maar niet in een protocol. Het analyse plan moet volgens het door de METC verplicht gestelde *CCMO Model-onderzoeksprotocol* (zie de website www.ccmo.nl) gesplitst te worden

- in aparte gedeelten voor de primaire analyse en voor de secundaire uitkomsten en/of analyses.
- De analyse dient voor iedere vraagstelling apart zó gedetailleerd te worden beschreven dat iedere data-analist bij beschikking over de database tot dezelfde resultaten zal komen. Dit betekent dat naast de technieken ook beschreven moet worden hoe bijvoorbeeld zal worden omgegaan met missende data en het eventuele categoriseren van continue uitkomsten. Als er een alternatieve keuze gelaten wordt, bijvoorbeeld tussen de t-toets of de Mann-Whitney toets, dan dient een *objectief* criterium voor deze keuze te worden vermeld. Bijvoorbeeld het gebruik van de Shapiro-Wilk toets (niet de vaak gebruikte maar ongeschikte Kolmogorov-Smirnov toets) om te beoordelen of de data een normale verdeling vormen (al dan niet na een datatransformatie, zoals de logaritmische). Dit in plaats van het maken van een subjectieve keuze op grond van bepaalde grafieken.
 - *Multivariabele analyses*: Als voor een bepaalde uitkomst zowel een univariate als een multivariate methode zal worden uitgevoerd, zoals een Ancova of logistische- of Cox-regressie, dan dient duidelijk te worden aangegeven welke analyse a-priori het belangrijkste wordt geacht. Verder dient de keuze voor de te hanteren covariabelen, inclusief hoe deze te gebruiken, op voorhand te worden gespecificeerd in het protocol. Als het gebruik van een multivariate methode wordt opengelaten en later op basis van trial uitkomsten besloten kan worden, dan dient het objectieve criterium voor het toepassen daarvan te worden aangegeven in het protocol. Meer informatie wordt gegeven in het ICH document “*Points to consider on adjustment for baseline covariates*” op de EMA website.
 - *Subgroep en andere analyses*: Bevindingen bij vooraf gespecificeerde subgroep analyses, mits beperkt qua aantal, worden in het algemeen van meer waarde geacht dan “post-hoc” evaluaties. Ditzelfde geldt voor eventuele andere exploratieve analyses. Het is aan te raden alle relevant geachte analyses op voorhand te specificeren in het protocol.
 - Veel van de belangrijke bladen eisen tegenwoordig dat het studieprotocol wordt meegezonden bij een aangeboden manuscript. De statistische paragraaf in het protocol is een relevant onderdeel voor de reviewers. Dit is niet alleen om te evalueren of de beschreven statistische methoden adequaat zijn, maar ook om te bezien of de bevindingen die gepresenteerd worden niet “data-driven” of “post-hoc” zijn. De statistische paragraaf is een wezenlijk onderdeel van een protocol en verdient daarom de nodige aandacht.
 - Bij een dubbel-blinde studie is het toegestaan om enige details later in te vullen in het *statistical analysis plan* (SAP) voordat de code verbroken wordt. Bij een open studie daarentegen is dat niet mogelijk en dient het SAP al compleet te zijn voordat de studie aanvangt. Dit moet als complete statistische paragraaf in het protocol worden opgenomen, of eventueel in de vorm van een Appendix bij het protocol worden gevoegd als het om een zeer uitgebreid plan gaat.

Nadere informatie

Uitgebreide informatie over statistische aspecten van een RCT is te vinden op de EMA website in het ICH E9 document “*Statistical principles for clinical trials*”. Nuttige aanwijzingen over statistische aspecten zijn ook te vinden bij het blad *Annals of Internal Medicine* in de Information for Authors sectie onder *Statistical Guidelines*.

Een boek dat veel statistische zaken bij een RCT bespreekt is “*Clinical Trials; a practical approach*” van SJ Pocock (uitgever John Wiley & Sons).