



Universiteit
Leiden
The Netherlands

Mixed models for correlated compositional data: applied to microbiome studies in Indonesia

Martin, I.

Citation

Martin, I. (2019, October 8). *Mixed models for correlated compositional data: applied to microbiome studies in Indonesia*. Retrieved from <https://hdl.handle.net/1887/79254>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/79254>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/79254> holds various files of this Leiden University dissertation.

Author: Martin, I.

Title: Mixed models for correlated compositional data: applied to microbiome studies in Indonesia

Issue Date: 2019-10-08

Samenvatting

Snelle verstedelijking gaat bijna altijd gepaard met een overgang van infectieziekten naar niet-overdraagbare ontstekingsziekten als dominante oorzaak van morbiditeit. Dit is zeer waarschijnlijk toe te wijzen aan veranderingen in leefstijl en omgevingsfactoren. Om het precieze mechanisme van deze transitie beter te begrijpen is er een populatiestudie uitgevoerd op Flores, een eiland in Indonesië. Dit gebied was geschikt voor deze studie omdat een groot aantal mensen is besmet met parasitaire wormen en er snelle veranderingen in leefstijl plaatsvonden. Hoewel worminfecties een negatieve invloed kunnen hebben op de menselijke gezondheid, wordt de aanwezigheid van wormen ook geassocieerd met een sterke regulatie van de afweerreactie en dit verklaart de lage prevalentie van ontstekingsziekten in gebieden waar veel mensen besmet zijn met wormen. In een gerandomiseerde, dubbelblind-uitgevoerde studie is er klinische en biomedische data verzameld om het complexe biologische mechanisme te onderzoeken dat aan deze associatie ten grondslag ligt. Tijdens deze studie werd de ene helft van de onderzoeksgroep behandeld tegen worminfecties, de andere helft kreeg een placebo. In dit proefschrift wordt de relatie tussen worminfecties, de samenstelling van het darmmicrobioom en de afweerreactie beschreven. Omdat de beschikbare analysemethoden niet geschikt bleken voor de complexiteit van de verzamelde data in deze populatiestudie, is een ander aandachtspunt in dit proefschrift de ontwikkeling van statistische methodologie.

Hoofdstuk 1 beschrijft een algemene introductie van het proefschrift met betrekking tot de verzamelde data, de onderzoeksvragen en de statistische methoden die beschikbaar zijn om deze data te analyseren en de vragen te beantwoorden. Er wordt ook kort beschreven hoe de microbioomprofielen zijn verkregen. De methoden onderliggend aan deze profielen leggen een compositioneel structuur op de microbioomdata op, waar rekening mee moet worden gehouden in het modelleren. Daarnaast zijn er meerdere datapunten van dezelfde persoon en dit leidt tot correlatie tussen de metingen. Tot slot wordt er een overzicht gegeven van de statistische hulpmiddelen die worden gebruikt om microbioomdata te analyseren, gevolgd door een discussie van de uitdagingen die komen kijken bij het correct modelleren van dit type data in een onderzoeksopzet met herhaalde metingen.

Hoofdstuk 2 beschrijft de toepassing van een recent ontwikkelde statistische methode om de microbiomdata, verzameld in de populatiestudie, te modelleren. Een aanname van dit model is dat de microbiomdata voortkomen uit een multinomiale verdeling. Om rekening te houden met de aanwezigheid van extra variatie, wordt de aanname gedaan dat de parameters van de multinomiale verdeling random effecten zijn die volgen uit een geconjugeerde verdeling. Echter, dit model is alleen geldig voor onafhankelijke waarnemingen van een multinomiale verdeling. Hierdoor kan er geen rekening worden gehouden met de correlatiestructuur in onze data door de herhaalde metingen. Om die reden zijn de analyses voor elk tijdstip apart uitgevoerd. Terwijl de data verkregen voorafgaand aan de behandeling is gebruikt om het effect van worminfecties op de samenstelling van het darmmicrobioom te analyseren, is de na afloop van de behandeling verkregen data gebruikt om het effect van de anti-wormen behandeling op de samenstelling van het darmmicrobioom te bestuderen. Om te onderzoeken of de behandeling een verschillend effect heeft op personen die voor aanvang van de behandeling met wormen besmet waren, ten opzichte van personen die niet besmet waren, is er een interactieterm tussen de status van infectie en de behandeling toegevoegd aan het model. Het bleek dat alleen in de groep die behandeld was met anti-wormenmedicatie en die zowel voor aanvang van de behandeling, als na afloop van de behandeling besmet was met wormen, er een significant verschil was in de verhouding van Bacteroidetes ten opzichte van Firmicutes en de verhouding van Actinobacteria ten opzichte van Firmicutes in vergelijking met de andere groepen. De methode die hier is toegepast heeft de beperking dat er maar één tijdstip kan worden geanalyseerd, en daarom kan de verandering van de samenstelling van het microbiom niet over het verloop van een bepaalde periode worden bestudeerd. Hoofdstuk 3 beschrijft de ontwikkeling van een model dat zich hierop richt.

In hoofdstuk 3 wordt de ontwikkeling van een statistische methode beschreven die herhaalde metingen van de samenstelling van het microbiom modelleert, waarbij rekening wordt gehouden met de correlatiestructuur tussen de herhaalde metingen van een persoon. Dit is gedaan door een normaal verdeeld random effect te introduceren. Er worden drie verschillende covariantiestructuren voor de normaal verdeelde random effecten overwogen. Ten eerste stellen we een univariate en persoons-specifieke random effect vast, waar het random effect voor elke categorie van bacteriën per verschillend tijdstip hetzelfde is. Ten tweede, elke categorie heeft een verschillend random effect met een categorie-specifieke variantie. Als laatste wordt aangenomen dat er een multivariate random effect is met een gemeenschappelijke variantie voor alle categorieën. Aan de hand van een simulatiestudie is onderzocht in hoeverre deze voorgestelde methode in staat is om een schatting te maken van de fixed effects en de standaard deviaties van de random effects. Hieruit bleek dat de schattingen van de fixed effects niet worden beïnvloed door de keuze van de covariantiestructuur van de nor-

maal verdeelde random effecten. Met betrekking tot de dataset, de conclusie uit hoofdstuk 2 is bevestigd door het uitgebreidere model. Dat wil zeggen dat de groep die zowel voor aanvang van de behandeling, als na afloop van de behandeling besmet was met wormen, een verandering vertoonden in de verhouding Bacteroidetes ten opzichte van Firmicutes. Om de betrouwbaarheid van het model te beoordelen, hebben we de marginale correlaties uitgerekend tussen en binnen de categorieën over tijd. Hieruit blijkt dat de marginale correlaties goed worden geschat met behulp van het multivariate random effect, gebaseerd op de gemeenschappelijke variantie voor alle categorieën.

In de volgende hoofdstukken wordt de wisselwerking tussen worminfecties, de samenstelling van het darmmicrobioom en de afweerreactie, gekenmerkt door cytokinerespons in het bloed op antigenen, beschreven. Eerder onderzoek heeft aangetoond dat anti-wormenbehandeling leidt tot herstel van de verzwakte afweerreactie, geassocieerd met worminfecties. Er wordt tevens gedacht dat bepaalde darmbacteriën de afweerreactie beïnvloeden. Het doel van hoofdstuk 4 is om inzicht te verkrijgen in de mechanismen die een rol spelen bij deze wisselwerking waarbij de metingen van beide tijdstippen worden gebruikt. Een linear mixed model is toegepast op de data, met de cytokineproductie in reactie op bepaalde antigenen als uitkomstvariabele. Voor dit model zijn het bacteriële aandeel of de diversiteit en hun interactieterm met parasitaire wormen gebruikt als verklarende variabelen. We hebben onze analyse beperkt tot drie categorieën van bacteriën, namelijk Actinobacteria, Bacteroidetes en Firmicutes, omdat deze bacteriële groepen geassocieerd waren met worminfecties in de analyses in Hoofdstuk 2 en 3.

In deze studie hebben we aangetoond dat er een significante associatie is tussen de toename in het aandeel van Bacteroidetes en de afname in de concentratie van IL-10 in reactie op LPS in personen die niet besmet zijn met wormen. De associatie is minder sterk in personen die besmet zijn met wormen. Hoewel deze resultaten de hypothese bevestigen dat het verwijderen van worminfecties leidt tot een sterkere afweerreactie en dat darmbacteriën de afweerreactie beïnvloeden, zijn er ook een aantal beperkingen in de analyse te benoemen. Er wordt aangenomen dat elke bacterieel aandeel onafhankelijk is, en de associatie met de productie van cytokines wordt apart geanalyseerd. Hieruit volgt dat het kenmerk van de samenstelling van het microbioom wordt genegeerd. In aanvulling hierop wordt er ook geen rekening gehouden met de meetfout binnen de microbioom-data bij dit model. Wat weer kan leiden tot onder- of overschatten van de regressieparameters. Bovendien wordt de associatie tussen worminfecties en bacteriële aandelen niet gekwantificeerd in dit model. Om die reden is er in Hoofdstuk 5 een statistische methode ontwikkeld die probeert deze beperkingen te overwinnen.

In Hoofdstuk 5 is ons doel om een statistisch model te bouwen voor de associatie tussen helminthinfecties enerzijds en zowel microbioom als cytokinerespons simultaan anderzijds, door alle bronnen van variatie in de data mee te nemen. Allereerst zijn de cytokinereponsen als continue uitkomsten en de multivariate teluitkomst van het darmmicrobioom gecorreleerd binnen een individu. Ten tweede worden er correlaties tussen dezelfde type observaties op verschillende tijdstippen verwacht. Ten slotte is er, specifiek voor microbioomdata, extra variabiliteit door overdispersie en meetfouten. We nemen aan dat de cytokineresponsen en de microbioomcomposities respectievelijk een normale en een multinomiale verdeling volgen. Een verzameling latente variabelen, waarvan aangenomen wordt dat ze een multivariate normale verdeling volgen, is in het model opgenomen om rekening te houden met additionele variatie in de data. De meetfout is gemodelleerd met een multidimensionaal normaal verdeeld random effect, m.a.w. elke categorie heeft zijn eigen random effect waarvan aangenomen wordt dat ze zijn gecorreleerd. Dit wordt gedaan voor meer flexibiliteit in het modelleren van de extra variatie in elke categorie. De gezamenlijke kansverdeling is geformuleerd en parameters zijn geschat door het maximaliseren van de gezamenlijke waarschijnlijkheidsfunctie met numeriek bepaalde kwadraturen. Een simulatiestudie is uitgevoerd om de prestatie van de schatters te onderzoeken voor zowel de fixed als de random effecten, in vergelijking met de methode geïntroduceerd in Hoofdstuk 4 (de naïeve methode). Het gezamenlijke model presteert beter dan de naïeve model. In de datatoepassing laten we zien dat de correlatie tussen de microbioomcompositie en de cytokineresponsen klein is. Wanneer we de marginale correlaties behorende bij het voorgestelde model uitrekenen, blijken deze niet in overeenstemming te zijn met de geobserveerde correlaties.

Hoofdstuk 6 beschrijft en vat de resultaten van de analyses die zijn uitgevoerd en de methoden gebruikt in Hoofdstuk 2 tot en met 5 samen. Het doel van dit hoofdstuk is om de bewijzen voor causaliteit van de geïdentificeerde associaties te evalueren tussen helminthinfecties, darmmicrobioomcompositie en cytokineresponsen, gebruikmakende van de data uit het gerandomiseerde gecontroleerde onderzoek. We hebben gevonden dat de behandeling een effect heeft op zowel de darmmicrobioomcompositie als cytokineresponsen via het verwijderen van de helminthinfectie, en dat de darmmicrobioom een directe effect heeft op cytokineresponsen. De gerichte acyclische grafen (DAGs) zijn gebruikt om de richting van het causale effect te visualiseren, en meerdere potentiële bronnen van vertekening zijn meegenomen in deze DAGs, namelijk niet-geobserveerde confounders en meetfouten. De statistische methoden ontwikkeld in dit proefschrift nemen de extra variatie veroorzaakt door niet-geobserveerde confounders en meetfouten mee via de inclusie van random effecten. Specifiek voor microbioomdata zijn er twee voor de hand liggende mogelijkheden voor verdelingsaannames op de randomeffecten, namelijk een geconjugeerde en een normale verdeling. In dit proefschrift hebben wij deze aannames verkend. Het blijkt dat modellen met

random effecten die een geconjugeerde verdeling hebben goed bij de microbiomdata passen, als je beschouwt hoe goed de marginale correlaties van het model de geobserveerde correlaties reflecteren. Gebaseerd op resultaten uit de literatuur en onze analyses, kunnen we concluderen dat behandeling een causaal effect heeft op helminthinfectie en dat een helminthinfectie directe effecten heeft op zowel de darmmicrobioom als de cytokineresponsen. Het blijkt dat de correlatie tussen darmmicrobioom en cytokineresponsen klein is, daarom is hun effect niet verder geëvalueerd.

Ten slotte, data uit gerandomiseerde gecontroleerde onderzoeken, zoals in dit proefschrift, zijn nuttig om causale verbanden tussen betrokken variabelen te onderzoeken. Verder leveren observaties die meermaals zijn gemeten informatie over hoe een specifieke uitkomst zich ontwikkelt over tijd. Helaas gebruiken de studies in dit proefschrift data over kleine deelsteekproeven uit een groter onderzoek, wat mogelijk leidt tot een verlaagd statistisch onderscheidend vermogen om effecten te detecteren. Een oplossing om deze limitatie te beperken is om verschillende bronnen met observaties te integreren.