

Doktorgradsarbeid «Development of antibacterial drug resistance – impact of the bacterial infection treatment – artikkel V.

Undertegnede har fått to mandater:

- 1) Å vurdere om det i artikkel fem er sannsynlig at det foreligger bevisst manipulering av data i et forsøk på å tilpasse resultater til en hypotese, eller om det er sannsynlig at de feil som foreligger har sin forklaring i unøyaktigheter fra ph.-kandidatens side.
- 2) Å lese hele avhandlingen for å danne seg et bilde av helheten i arbeidet.

Tre dokumenter ligger til grunn: 1) avhandlingen, 2) et excelark med tittelen «Prosjekt 554: ETEC, SPEDGRIS (Februar 2012) som er lagt ved (vedlegg a), 3) et excelark med samme tittel som b), men hvor kun CRP verdiene ligger inne (vedlegg b). Variablene i vedlegg a er Lab.nr, Analysedato, Merket, Hemolyse, Na, K, Cl, Urea, CRP, Tprot, Alb-elfo, Alpha 1, Alpha 2, Beta 1, Beta 2. Vedlegg b inneholder foruten CRP verdiene også utregningene som er foretatt og for å gjennomføre disse er det satt inn verdier i de tilfellene hvor verdier mangler eller er lik 0.

Det framgår av artikkel V, Authors' Contributions, at kandidaten var alene om å gjennomføre de statistiske analysene. Jeg går derfor ut fra at beregningene i vedlegg b er gjort av kandidaten.

CRP verdiene i artikkel V

I artikkel V under avsnittet «Blood sampling» står det: «The serum analyses were performed by the Central Laboratory, Norwegian School of Veterinary Science and comprised sodium (Na), potassium (K), chloride (Cl), urea, C-reactive protein (CRP) and total-protein (Tprot)». Det er altså Sentrallaboratoriet ved Norges veterinærhøgskole som har utført alle serumanalysene på nevnte substanser, inkludert CRP.

Serumverdiene er målt på tre tidspunkter: «Immediately postpartum 0 h prior to challenging», «6-8 h post-partum (prior to treatment)» og «30-32 h (prior to necropsy)», konferer side 21 i artikkel V. Det opereres for øvrig med andre tidspunkter i tabell teksten i Supplementary material 2 og i selve tabellen. Jeg vil betegne tidspunktene med A, B og C i overensstemmelse med definisjonen av «Host fold response to treatment» som betegnes Fold i mange av figurene. Definisjonen er $Fold = C * A / B * B$, konferer figur 1 i artikkel V. Beregning av Fold for en gris krever en serumverdi på alle tre tidspunkter.

Forsøket er satt opp med 10 behandlinger. Det er fem dyr (griser) i hver gruppe bortsett fra placebogruppa hvor det er 6 griser og i gruppa «Zoolac+ >MIC Baytil 10%» hvor det er 4 griser og gruppa «Zoolac+ >MIC Baytril 10%+Metacam» hvor det er tre griser. I alt er det 48 dyr som hver skulle vært målt tre ganger, i alt $48 * 3 = 144$ målinger. Ved gjennomgang av vedlegg b så mangler det så mangler det CRP verdi, eller CRP=0 for 38 målinger. Dette involverer 28 dyr. Av de 48 dyr som inngår i forsøket er det mulig å beregne Fold for kun 20 dyr.

Ifølge vedlegg b er det satt inn verdier for de manglende målinger og for de med verdien lik 0. Hvilken prosedyre som er benyttet for å erstatte manglende verdier er ikke nevnt i artikkelen. E-post

korrespondanse med kandidaten, via hovedveileder, har ikke gitt svar på spørsmålet. Det kan altså ikke dokumenteres hvordan erstatningsverdiene har framkommet.

Det finnes forskjellige måter å erstatte manglende verdier. Prosessen betegnes imputering (imputation) i statistikken. En måte ville være å bruke gjennomsnittet for de dyr, innen forsøket, hvor det foreligger måling. I vedlegg c, «imputering», har jeg sjekket i hvilken grad denne metode er blitt benyttet. Horisontalt er forsøksnr, fra 1 til 10. Nummereringen følger den som kom fra Sentrallaboratoriet, vedlegg a. Vertikalt er gris nr. De tallene som er uthevet og skrevet med kursiv er imputerte tall. Nederst er gjennomsnittet for de griser hvor det foreligger verdier (ikke imputerte). Hvis disse er skrevet med rødt, så betyr det at de imputerte tallene stemmer med gjennomsnittet av de ikke imputerte tallene. Ta forsøk nr 1 ved tidspunkt A. Gjennomsnittet av de ikke imputerte tallene er 0,13, men det er imputert inn 0,1 for to griser. Ved tidspunkt B burde det vært imputert inn 1,38, men det er satt inn 0,7. Ved tidspunkt C er det imputert inn 5,75, hvilket stemmer overens med gjennomsnittet for de ikke imputerte verdiene. I forsøk 5 manglet samtlige griser verdi ved tidspunkt B. Her er det satt inn 0,5 for samtlige griser. Prosessen som har ledet fram til erstatning av manglende verdier har ikke systematisk vært basert på gjennomsnitt av eksisterende verdier.

Mandatet gikk blant annet ut på å vurdere om det har forekommet bevisst manipulering av data med tanke på å tilpasse resultatene til en hypotese. En måte å nærme seg spørsmålet på er å lage en figur basert den nevnte gjennomsnittsmetoden for imputering og så sammenligne figurer. Dette er gjort i vedlegg d, «to figurer». Øverste figur er klippet inn fra vedlegg b. Nederste figur er laget av undertegnede basert på gjennomsnitts imputering. Forsøk nr 8 «høyt Baytril» skiller seg ut. Men forskjellen skyldes i hovedsak en regnefeil fra kandidatens side som framgår av regnearket, vedlegg b. Søylen som rager høyest, forsøk nr 3, er ganske lik med «gyldig» imputering og med kandidatens «imputering». Forskjellene framstår ikke store nok til å hevde bevisst manipulering for å understøtte en hypotese. På den annen side går det som er gjort langt utover det som kan tilskrives unøyaktigheter fra kandidatens side.

Overraskelsen blir imidlertid stor når man sammenholder figur 11 i artikkelen med figuren som kandidaten selv har laget, vedlegg b. I artikkelen stiger nå gruppa «Høybaytril zoo» opp til nærmere 40, mot 16 på regnearket. Det er også et par andre søyler som rager høyere i artikkelen enn i vedlegg b. Grunnlaget for figur 11 kjenner jeg ikke og de stemmer ikke med de verdier som kommer fra Sentrallaboratoriet med kandidatens «imputeringer». Hvis høye søyler understøtter den hypotesen som er framsatt, så framstår figur 11 i artikkel V som manipulert i retning av å understøtte en hypotese.

Natrium verdiene i artikkel V

Hvis man blar igjennom Na verdiene i vedlegg a, så ser man at det også her forekommer manglende verdier. For forsøk 5 så mangler samtlige dyr Na verdi på tidspunkt B. I og med at forsøk nr 5 inngår i figur 10, så må det være imputert. Hvordan vites ikke. Så vidt jeg kan se av regnearket så er det ingen verdier under 129 mmol/l. Dette er ikke forenlig med at gjennomsnittsverdier ligger lavere enn 129 mmol/l. I Supplementary material 2 er Na gjennomsnittet på tidspunkt C i placebogruppa 125,8 mmol/l. Vedlegg a viser følgende verdier for de 6 griser på tidspunkt C i placebogruppa: blank, blank, 142, 147, 145, 134. Det er uforståelig hvordan gjennomsnittet har blitt 125,8. Forsøk nr 10 viser følgende verdier på tidspunkt C (kun tre griser): 139, 141, 143. I Supplementary material 2 er gjennomsnittet 152. Dette er åpenbart feil. Det foreligger en rekke slike uoverensstemmelser mellom

de originale tallene fra Sentrallaboratoriet og Supplementary material 2 som samlet går langt utover det som kan tilskrives unøyaktigheter.

Gamma verdier

Gamma er valgt vilkårlig blant substansene hvor det var komplette tall ved tidspunkt A og tidspunkt C i placebogruppa. Gjennomsnittet ved tidspunkt A blir 0,55 og ved tidspunkt C 0,63. I Supplementary material 2 står det henholdsvis 0,5 og 0,8. Det er ingen grunn til å gjøre om 0,55 til 0,5 når det benyttes to desimaler lenger ned i tabellen.

Urea verdier

I artikkel V, side 6, framheves at Urea nivåene var signifikant forhøyet ved tredje tidspunkt (tidspunkt C). I Supplementary material 2 angis gjennomsnittsnivået i placebogruppa til 7,6 mmol/l. På originalarket fra Sentrallaboratoriet (vedlegg a) er det 4 griser i placebogruppa med måling på tidspunkt C. Gjennomsnittet for disse 4 er 3,45 mmol/l. For eksperiment gruppe 2 (Terramycin, Pfizer (Tetracycline)) angis gjennomsnittsnivået på tidspunkt C til 8,9 mmol/l. Originaltallene viser et gjennomsnitt på 3,42 mmol/l. En gjennomgang viser at for samtlige eksperimenter var det uoverensstemmelse mellom tall i artikkel og tall fra Sentrallaboratoriet ved tidspunkt C.

Vurdering

Det er skremmende lite som stemmer overens i artikkel V, Supplementary material 2 og det som er kommet fra Sentrallaboratoriet. Det er vitenskapelig uredelig ikke å ha opplyst om at det mangler verdier i store deler av eksperimentene og hva som er gjort for å bøte på det. Motivasjonen bak alle endringer av tall og «imputeringer» er vanskelig å forstå. CRP resultatene og urea resultatene kan tyde på at det er gjort for å understøtte en hypotese. På den annen side er alle tallene som er vurdert på siden av hovedhypotesen som dreide som om å studere «the expression of pathogenic ETEC borne pRAS1 transfer genes *in vivo* in piglets in response to antibiotic, probiotic and Non-steroidal Anti-inflammatory Drug (NSAID) treatments, while simultaneously monitoring selected blood serum proteins». Resultatene fra blodserum prøvene skulle altså kun monitoreres, men har fått en dominerende plass i artikkelen. Da jeg ikke har vurdert expression of tranfer genes finner ikke grunnlag for å hevde manipulasjon for å understøtte en hypotese, men det har foregått omfattende endringer i datamaterialet og uforståelige «imputeringer» som det er berettiget å betegne som fusk.

Dette dokumentet har 4 e-post vedlegg:

Vedlegg a - Kopi av 554-ETEC spedgris (leon Cantas)_mottatt_05_09_2013

Vedlegg b - Kopi av Cantas 130613

Vedlegg c - Imputering

Vedlegg d - to-figurer