



Notat: Observationel, matchet analyse med kvægdatabasen

29-12-2021

Svend Vendelbo Nielsen, konsulent
sven@teknologisk.dk, +45 72 20 17 41

Notatet, kort fortalt

Dette notat giver følgende konklusioner

- Virker de nuværende behandlingerne mod Mastitis?
 - Svar: Ja.
- Hvordan er effekten af de forskellige behandlinger mod Mastitis?
 - Svar: Goldbehandling virker bedst, men den fulde effekt ses først efter 100 dage. Behandlingen "Yverbetændelse" virker allerede efter 40 dage, men har ikke samme langtidseffekt som goldbehandling
- Afhænger effekten af behandlingen mode Mastitis af bakterietypen målt ved dyrkningsprøver?
 - Behandlingen "Yverbetændelse, akut" virker godt mod bakterierne "KoagulasenegativestafylokokkerCNS", "Strdyagalactia" og "Staphaureus". Goldbehandling virker godt mod alle bakterier.

Der er følgende forbehold

- Sammenligningen mellem behandlede og ubehandlede køer er lavet med en retrospektiv matchet analyse, som er mere usikkert end et randomiseret forsøg, fordi der kan være confounding variable.

Introduktion

En ko med Mastitis har et højere celletal i dens mælk end køer uden Mastitis. Et højt celletal sænker kvaliteten af mælken og er en vigtig indikator for dyrets helbred. I månedlige ydelseskontroller får en stor del af mælkeproducerende, danske køer målt sit celletal samt andre centrale nøgletal af mælken. Data fra disse ydelseskontroller giver os derfor en unik mulighed for at sammenholde køers grad af Mastitis med data om deres behandlings- og sygdomshistorik.

I dette notat besvarer vi spørgsmålene

- Virker de nuværende behandlingerne mod Mastitis?
- Hvordan er effekten af de forskellige behandlinger mod Mastitis?
- Afhænger effekten af behandlingen mode Mastitis af bakterietypen målt ved dyrkningsprøver?

Dette gøres ved at lave et observationelt studie hvor vi sammenligner de køer, som har modtaget behandling for Mastitis med de køer som ikke har. En iboende udfordring med observationelle studier er deres sårbarhed over for uobserverede, *confounding* faktorer. Det er for eksempel plausibelt, at de køer som blev behandlet mod Mastitis er mere syge end dem, som ikke blev behandlet, og det er vigtigt at tage højde for i analysen. Det gør vi ved hjælp af *matching* og *multiple regression*.

Med matching konstruerer vi retrospektivt en kontrol- og behandlingsgruppe, som er så identiske som muligt bortset fra at køerne i behandlingsgruppen modtog behandling og kontrolgruppen ikke modtog behandling. Hvis der er forskel på helbredsudsigterne mellem de to grupper, skyldes det formentlig behandlingen. Matching-metodens effektivitet afhænger af, hvor identisk man kan gøre de to grupper. I den indeværende analyse bør vi for eksempel være opmærksomme på, at beslutningen om at behandle en ko oftest blev taget ud fra en visuel vurdering af dyrets infektionsgrad, hvilket ikke er gemt i databasen. Derfor må vi formode, at dyrene i behandlingsgruppen oftere så mere inficerede ud end kontrolgruppen.

Med multiple regression tages der højde for, at ikke kun behandlingen kan forklare helbredsudfaldet for den enkelte ko. Det gøres ved at modellere helbredsudfaldet som den samlede sum af effekterne af alle faktorer.

Metoder

I matching proceduren blev alle behandlingstidspunkter for alle køer parret med ét kontroltidspunkt. Behandlingstidspunkterne opfyldte

- Behandlingen tilhørte gruppen "Yverlidelser".
- Behandlingen fandt sted mellem 2. januar 2018 og 1. november 2021.

Kontroltidspunkterne blev udvalgt mellem alle ydelseskontroller, som opfyldte

- For den pågældende ko var der ingen "Yverlidelser"-behandling inden for de seneste 30 dage.
- Der er en måling for celletal og mælkemængde for den pågældende ydelseskontrol.
- Kontroltidspunktet lå mellem 2. januar 2018 og 1. november 2021.

Vi udførte matching proceduren med *nearest neighbor*-metoden, hvor afstand mellem behandlingstidspunkter og kontrolkandidater blev udregnet med den såkaldte GLM-metode. Det vil sige at for hvert behandlingstidspunkt og kontroltidspunkt, blev der udregnet en sandsynlighed for at den pågældende ko blev behandlet på netop dette tidspunkt med logistisk regression.

Efterfølgende parrede proceduren behandlingstidspunkter og kontroltidspunkter der havde så ens sandsynligheder som muligt. Vi indførte også den ekstra restriktion at et par ikke måtte have celletal som var mere end 50,000 celler forskellige.

Til at estimere sandsynlighederne for behandlingerne med den logistiske regression, brugte vi variablene celletal ved seneste ydelseskontrol, økologisk eller konventionel status, race, alder, paritet, seneste kalvs fødselskode og forløbskode af seneste kælvning. Vi inkluderede også fedtprocent, proteinprocent og mælkemængde i mælkeprøven ved seneste ydelse. Derudover brugte vi det totale antal tidligere behandlinger stratificeret på behandlingsgrupperne reproduktionslidelser, yverlidelser, stofskiftelidelser, klov- og lemmelidelser, fordøjelseslidelser, luftvejslidelser samt kategorien øvrige lidelser.

Mediancelletallet stiger med flere millioner for køer i lely-besætninger i et kort tidsrum inden behandlingstidspunktet (se andet dokument). Vi forventer derfor at en kos celletal ved behandlingstidspunktet kan være betydeligt højere end dens celletal ved sidste ydelseskontrol. Dette er potentielt problematisk, fordi vi matcher de behandlede køer på baggrund af deres celletal ved seneste ydelseskontrol og hvis dettes celletal ikke svarer til celletallet ved behandlingstidspunktet, vil de behandlede køer oftest blive matchet med køer som er raskere end dem selv. For at vurdere og eventuelt mindske denne effekt, kigger vi på forskellige delmængder af det matchede datasæt hvor

- Behandlingstidspunktet ligger højst Y dage efter ydelseskontrol

Vi varierer Y mellem 0,1,2,5,10,30 og uendelig.

For at måle effekten af en behandling, kiggede vi på forskellige responsvariable

- Om koen er i live efter X dage (kodet som "survival").
- Om koens celletal er faldet efter X dage (kodet som "improvement"). Afgang registreres som en negativ; dvs. hvis koen dør, siger vi at celletallet ikke er faldet.
- Om koens celletal er under 250,000 efter X dage (kodet som "healthy_after"). Afgang registreres som negativ.
- Om koens celletal er halveret efter X dage (kodet som "significant_improvement"). Afgang registreres som negativ.
- Om koen er død på andre måder end slagting efter X dage (kodet som "dead_not_butchered").

Vi varierede X mellem 40, 70, 100, 180 og 365.

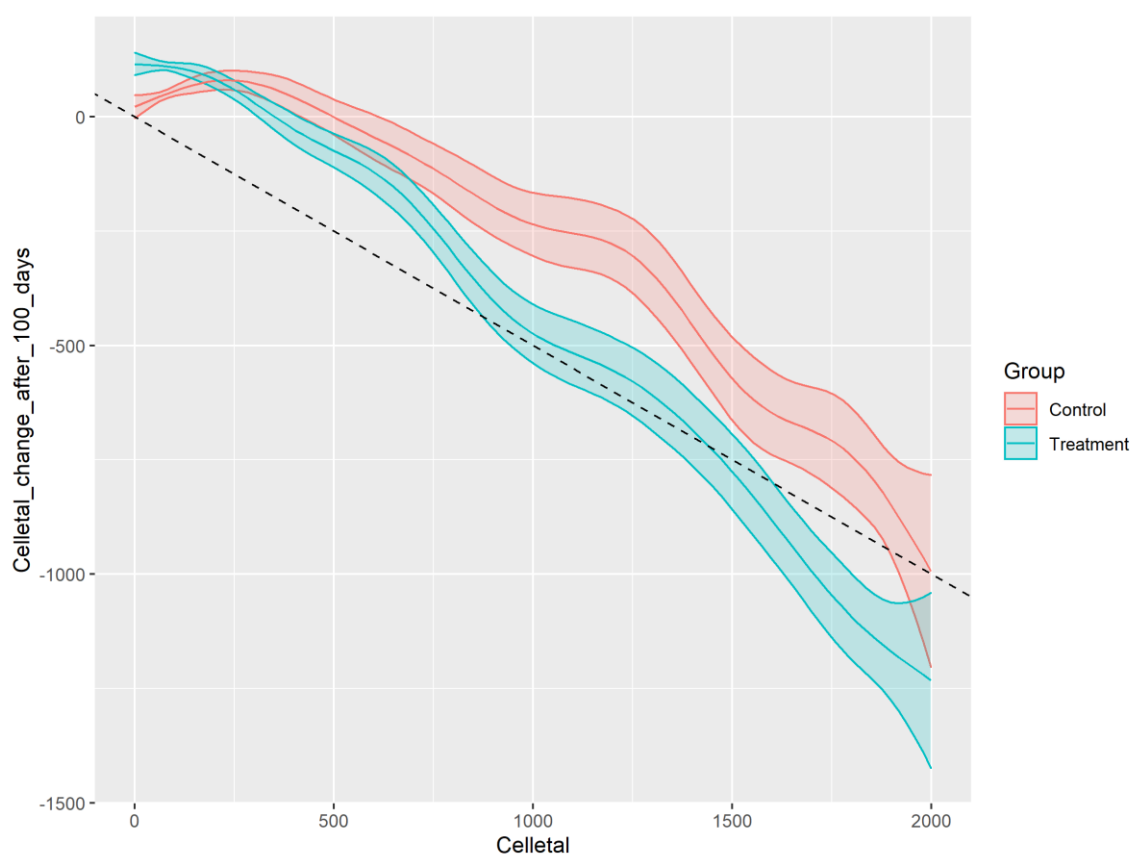
For at teste om en specifik behandling eller dyrkningskode har en gavnlige effekt, laver vi en delanalyse, hvor vi udelukkende kigger på de kontrol-behandlings-par, hvor den behandlede ko har fået behandlingen eller dyrkningskoden. Det betyder at delanalysen svarer på spørgsmålet "hvordan er udfaldet for denne behandlede ko (med behandling X og evt. dyrkningskode Y), sammenlignet med en tilsvarende ubehandlede ko. Delanalyserne kan dog ikke med samme sikkerhed svare på hvilken behandling har den største virkning på en specifik ko, fordi kontrolgruppen for én behandlingstype potentielt kan være anderledes end kontrolgruppen for en anden behandlingstype. Den multiple regression med de mange adjusting variable kunne i princippet tage højde for de forskellige kontrolgrupper, men det ville formentligt fungere bedre at lave et nyt matchet datasæt hvor parrene eksempelvis er lavet mellem køer behandlet med A og køer behandlet med B. Det er dog uden for denne rapport's omfang.

Grundet den store mængde data stratificerede vi matching proceduren efter 180 underinddelinger af alder. Der er lavet to separate matchede datasæt baseret på en tilfældig opsplitning af data. Selve analysen er kun lavet på den ene halvdel af data. Det er gjort for at bevare muligheden for at efterteste konklusioner med formelle tests.

Data exploration

Vi forventer at en kos celletal fluktuerer over tid. Det betyder at vi ikke ved om et fald i celletal skyldes en eventuel behandling eller bare den naturlige

fluktuation. For at give os en idé om denne fluktuationens størrelsesorden, fittede vi en kontinuert kurve, hvor x-aksen var celletallet, og y-aksen var ændringen i celletal efter 100 dage. Vi brugte det matchede datasæt og lavede en separat kurve for de behandlede og ubehandlede køer (se Figur 1). Fitningen blev lavet med R-pakken mgcv hvor celletalsændringen antoges at være normalfordelt. I mgcv-regressionen blev der adjusted for alle de variable der var brugt til at udregne sandsynligheden for behandling (se Metoder). Kurverne fortæller os at celletal over 500,000 naturligt sænkes til under 500,000 (i gennemsnit) uden indgreb. I behandlede køer falder celletallet allerede i gennemsnit, hvis startsudgangspunktet var over 350,000. Også størrelsen på faldet er større for behandlede køer – ned til en gennemsnitsværdi på 50% for behandlede køer.



Figur 1: X akser viser celletallet målt i tusinder og y-aksen viser ændringen i celletal efter 100 dage (også målt i tusinder). Den stiplede linje angiver celletalshalveringen.

Resultater

I Figur 2 ses odds ratios for forskellige responsvariable, som sammenligner behandlede køer med ubehandlede køer. En odds ratio større end 1 betyder at behandlingen øger sandsynligheden for responsvariablen. Det ses at

- behandlede dyr har højere odds for at overleve (panel 1), men til gengæld har de større risiko for at dø af andre grunde end slagting (panel 6).

- For alle tider ser vi at behandling giver højere odds for at være rask – dvs. have celletal under 250,000 (panel 2). Dog afhænger effekten af afstanden mellem behandlingstidspunktet og seneste ydelseskontrol.
- Det er først 100 dage efter behandlingstidspunktet at odds for at celletallet falder, er højere for behandlingsgruppen end kontrolgruppen (panel 3).
- Celletallet for dyrene i behandlingsgruppen halveres med større odds end for dyrene i kontrolgruppen efter 70 dage (panel 4).
- I 70 dage efter behandlingstidspunktet mangler stadig ydelseskontrolldata og afgangsdato for 5% af dyrene og for 40 dage er tallene mellem 10 og 26 % (Tabel 1). Det er ikke jævnt fordelt mellem behandlingsgruppe og kontrolgruppe (Tabel 1 og Figur 1, panel 5).

Tabel 1: Andelen af køer som ikke har haft ydelseskontrol eller afgang forskellige antal dage efter behandlingstidspunktet (beh.)

Dage efter beh.	Kontrolgruppe	Behandlingsgruppe
40	0.262	0.099
70	0.05	0.043
100	0.039	0.039
180	0.038	0.037
365	0.038	0.036

Tabel 2: Andelen af køer som har et lavere celletal efter forskellige antal efter behandlingstidspunktet sammenlignet med celletallet ved behandlingstidspunktet. Jo højere tal, jo bedre har køerne det.

Dage efter beh.	Kontrolgruppe	Behandlingsgruppe
40	0.375	0.145
70	0.363	0.294
100	0.393	0.504
180	0.421	0.506
365	0.387	0.414

Tabel 3: Antallet af behandlinger i det matchede datasæt.

Behandling	Antal
Goldningsbehandling	353530
Yverbetændelse	141510
Intern pattelukning	68819
Startvac vaccination	60037
Yverbetændelse akut	22561
Mælkeprøve udtaget	14577
Andre	32035

Tabel 4: Antallet af forskellige dyrkningskoder i det matchede datasæt, hvor afstanden mellem behandlingstidspunkt og seneste ydelseskontrol er højst 10 dage.

Dyrkningskode	Antal
KoagulasenegativstafylokokkerCNS	28642
Struberis	16198
Ecoli	11437
Staphaureus	10899
Enterococcuspp	6400
Strdysgalactia	5689
Dyrkningsnegativ	5006
Mælkeprøveandenbakterie	3343
Micrococcuspp	1725
Klebsiellaspp	1560
Øvrigeenterobacteriaceae	807
Trueperellapyogenes	418

På Figur 3 og 4 kan vi se effekten af behandling fordelt på de 6 mest almindelige underbehandlinger af yverlidelser (Se tabel 3).

- Begge figurer viser at goldningsbehandling har en stor effekt ved 100-dagesmærket.
- Når vi kun begrænser afstanden mellem behandlingstidspunktet og seneste ydelseskontrol til 10 dage, har Anitbiotikabehandlingerne "Yverbetændelse" og "Yverbtændelse akut" en mindre effekt end hvis grænsen var 10 dage.
- Antibiotikabehandlingerne udviser ikke samme forsinkede positive effekt som Goldningsbehandlingen. Dog aftager den positive effekt hurtigere.

På Figur 6 og 7 kan vi se effekten af behandlingen stratificeret på både de 6 mest almindelige behandlingstyper og dyrkningskoder (Tabel 4). Der er ikke så mange signifikante resultater, som for de andre grafer, så der er kun plottet de signifikante resultater (med signifikansniveau 5 %, ikke justeret for multiple testing).

- Goldningsbehandlingen udviser en forsinket positiv effekt for stort set alle dyrkningskoder, når responsen er hvorvidt celletallet halveres (Figur 6). Det er kun de mest sjældne dyrkningskoder, hvor der ikke er evidens for denne forsinkede positive effekt af goldbehandling.
- Specifikt for dyrkningskoderne "KoagulasenegativstafylokokkerCNS", "Strdysgalactia" og "Staphaureus" er der en signifikant, positiv effekt af behandlingen "Yverbetændelse, akut".

På Figur 8 vises det samme som Figur 2, blot er der her kun set på køer yngre end 1200 dage for at undersøge om de manglende observationer på de afgåede dyr har en effekt på resultaterne. Der var ikke nok afgåede dyr i datasæt til at estimere odds ratio for overlevelse.

- Behandlingerne har stort set samme effekt for de unge dyr som for de gamle dyr.

Diskussion

Efter 100 dage er effekten af yverlidelsesbehandlinger er størst. Dette signal er klart for alle vores undersøgte responser (undtagen overlevelse) og for alle valg af begrænsninger for afstanden mellem behandlingstidspunkt og seneste ydelseskontrol.

Behandlingens effekt afhænger meget af, hvor mange dage man kigger frem (Figur 2). Det kan næppe forklares med at de behandlede køer som ikke har fået en ydelseskontrol efter 40 og 70 dage skulle være særligt mere raske end de ubehandlede køer, da der ikke er nok behandlede køer med manglende ydelseskontrol (se Tabel 1 og 2). Det ses også at være uafhængigt af dyrets alder (Figur 8). Derfor tyder det på, at der er tale om et reelt mønster i data.

For yderligere at vurdere sensitiviteten ved konklusionerne på Figur 2 lavede vi en sensitivitsanalyse for mulige confounding variable på behandlingseffekten med responsvariablen, der angiver om celletallet er halveret efter 100 dage. Afstanden mellem behandlingstidspunktet og seneste ydelseskontrol var max 5 dage. R-Udskriftet er vist i Figur 5. Sensitivitsanalysen viste at konklusionen om positiv effekt på behandling er robust over for confounding variable som ikke forklarer mere end 17% af variansen af både outcome (dvs. om celletallet er halveret efter 100 dage) og yverlidelsesbehandlingen (Figur 5). Det er en stor del af variansen som skal forklares af denne potentielle confounding variabel, og den mest oplagte confounding variabel – det visuelle tjek af koens yver og mælk – ville formodentlig virke i den modsatte retning. Dette sandsynliggør at de fundne signifikante positive effekter af behandling er reelle.

Når vi kigger nærmere på de forskellige underinddelinger af yverlidelsesbehandlinger, ser vi store forskelle mellem de forskellige behandlinger (Figur 3). Goldbehandlingen er umiddelbart den bedste behandling, da den giver størst odds for at et behandlingsforløb er gunstigt på den lange bane, sammenlignet med et ikke-behandlingsforløb. Den forsinkede, positive effekt af goldningsbehandlingen var også synlig i analysen, hvor alle underbehandlinger var samlet i en kategori (Figur 2).

Når afstanden mellem seneste ydelseskontrol og behandlingstidspunkt er højest 10 dage, ligner det at behandlingen "Yverbetændelse, akut" mindsker oddset for at blive rask (Figur 3). Dog er dette billede modsatrettet hvis denne afstand er højst to dage (Figur 4). Når afstanden er mindre, forventer vi, at de behandlede køer i højere grad bliver matchet op med ubehandlede køer, der er lige så syge som dem selv. Derfor giver det god mening, at effekten af behandlingen er mere positiv, når denne afstand er mindre. Det tyder altså på at ingen af behandlingerne i gennemsnit er værre end ikke at behandle.

Modsat goldningsbehandling har behandlingerne "Yverbetændelse" og "Yverbetændelse, akut" ikke en forsinket positiv effekt, men effekten aftager hurtigere (Figur 3 og 4). Det antyder at de to behandlinger er kortsigtede, sammenlignet med goldningsbehandling.

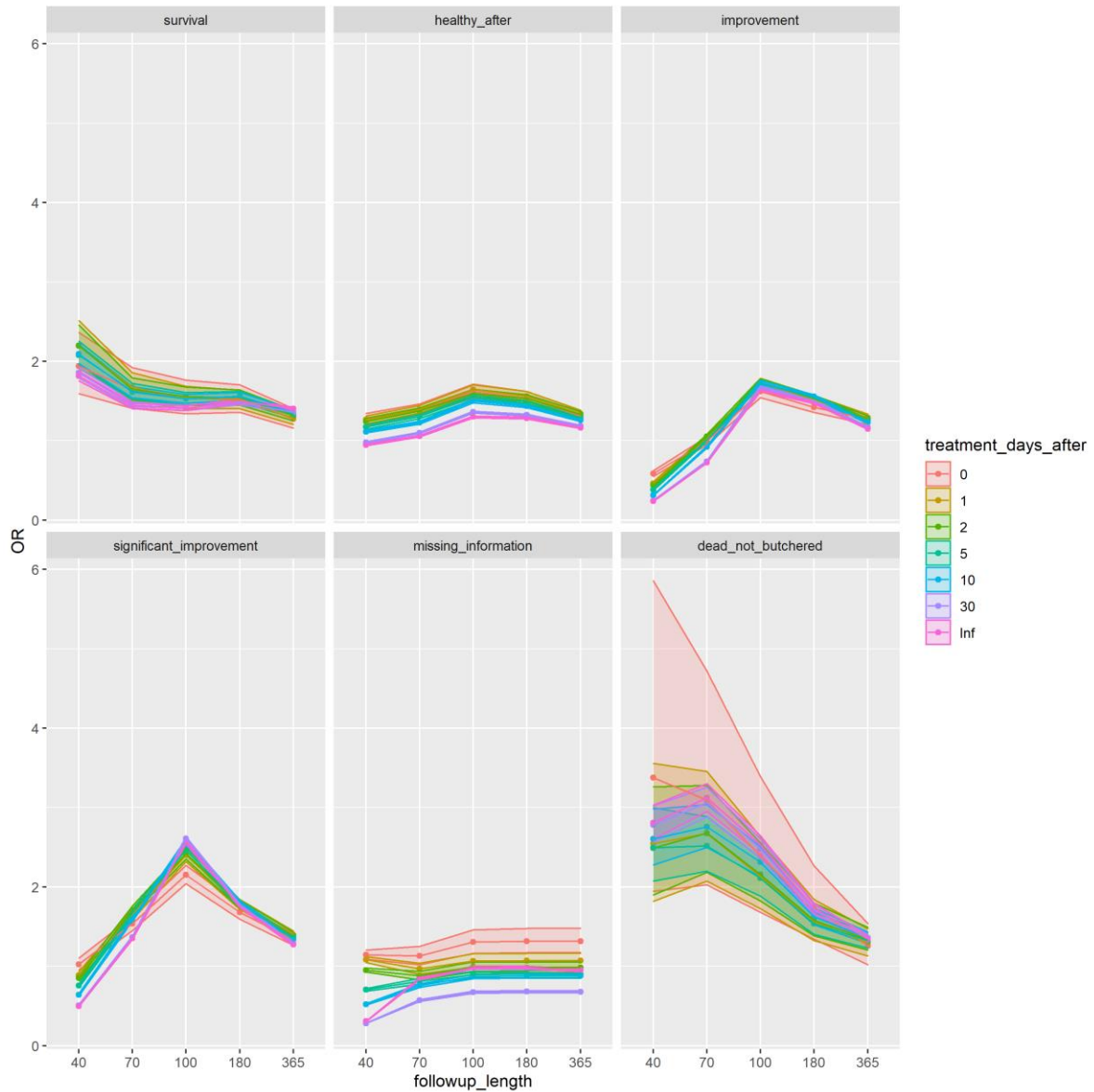
Goldningsbehandlingens forsinkede, positive effekt er synlig for alle de mest almindelige dyrkningskoder (Figur 6). Derudover var der for dyrkningskoderne "Koagulasenegative stafylokokker CNS", "Strdysgalactia" og "Staphaureus" en signifikant, positiv effekt af behandlingen "Yverbetændelse, akut" (Figur 6). Der er muligvis mange flere interaktioner mellem behandlinger og dyrkningsprøve, men der var ikke så meget statistisk styrke tilbage, når data er splittet så meget op. Flere steder lignede det at "Yverbetændelse" gav signifikant dårligere sygdomsudfald sammenlignet med de køer som ikke blev behandlet (Figur 6). Dette er dog nok den bias vi også så i Figur 3 på grund af afstanden mellem behandlingstidspunkt og seneste ydelseskontrol. Desværre var der for lidt data til at finde signifikante resultater når denne afstand var højst 2 dage.

Klassifikationsmodel i fremtiden

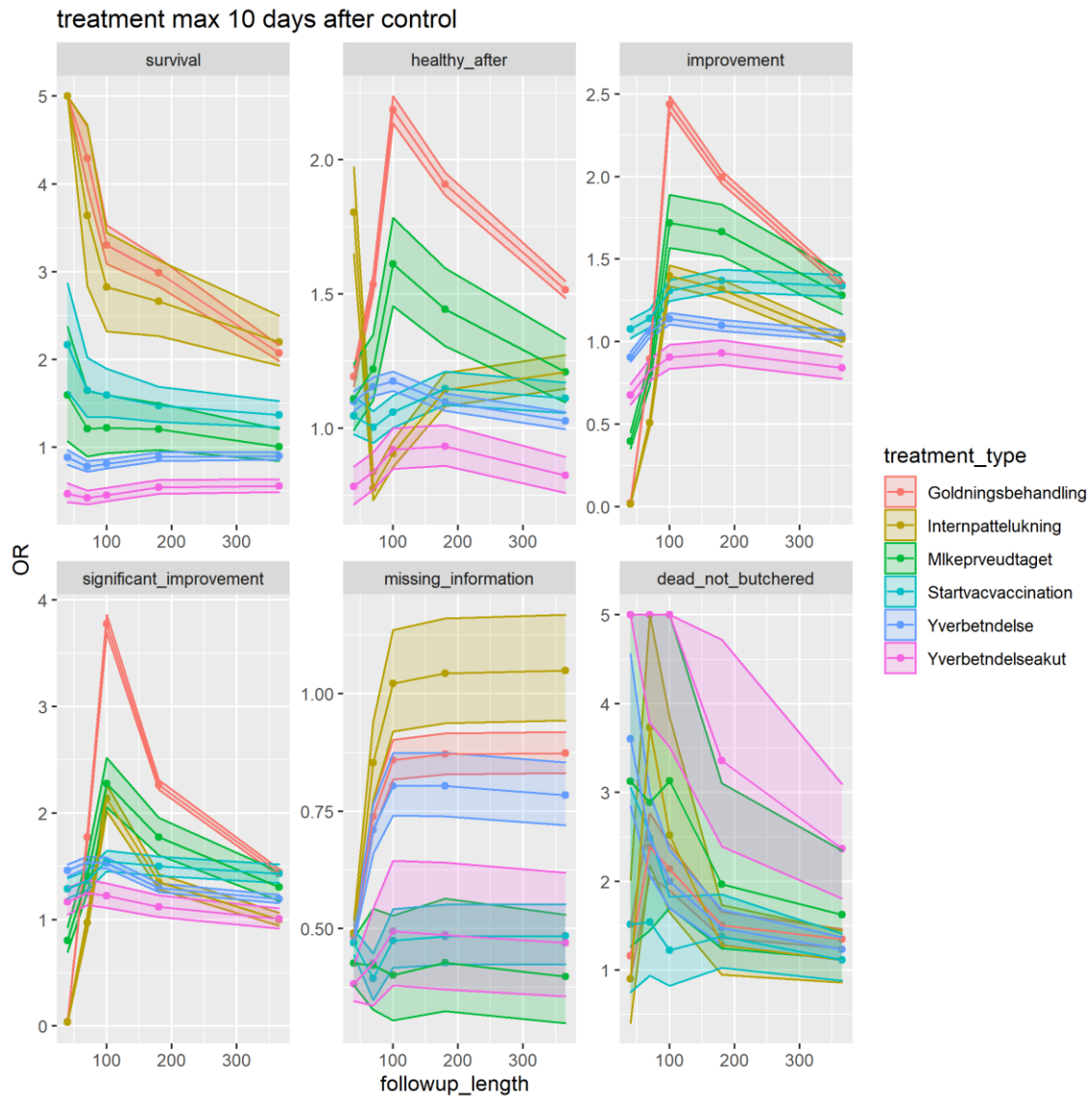
Analyserne i denne rapport kan relativt nemt bruges som en prædiktionsmodel, som kan forudsige virkningen af en behandling. For eksempel kan de logistiske regressioner for om celletallet er under 250,000 efter behandlingen bruges til at udregne sandsynligheden for om celletallet falder til under 250,000 for køer både i og udenfor det matchede datasæt.

I denne rapport har vi fokuseret på at resultaterne skulle være lette at fortolke ved for eksempel at vælge en logistisk regression, der ikke inkluderer interaktioner mellem behandlingen og andre variable. I en generel prædiktionsmodel kunne man sagtens bruge interaktioner samt ikke-lineære sammenhænge. Det åbner op for anbefalinger, der er mere specifikke for den enkelte ko.

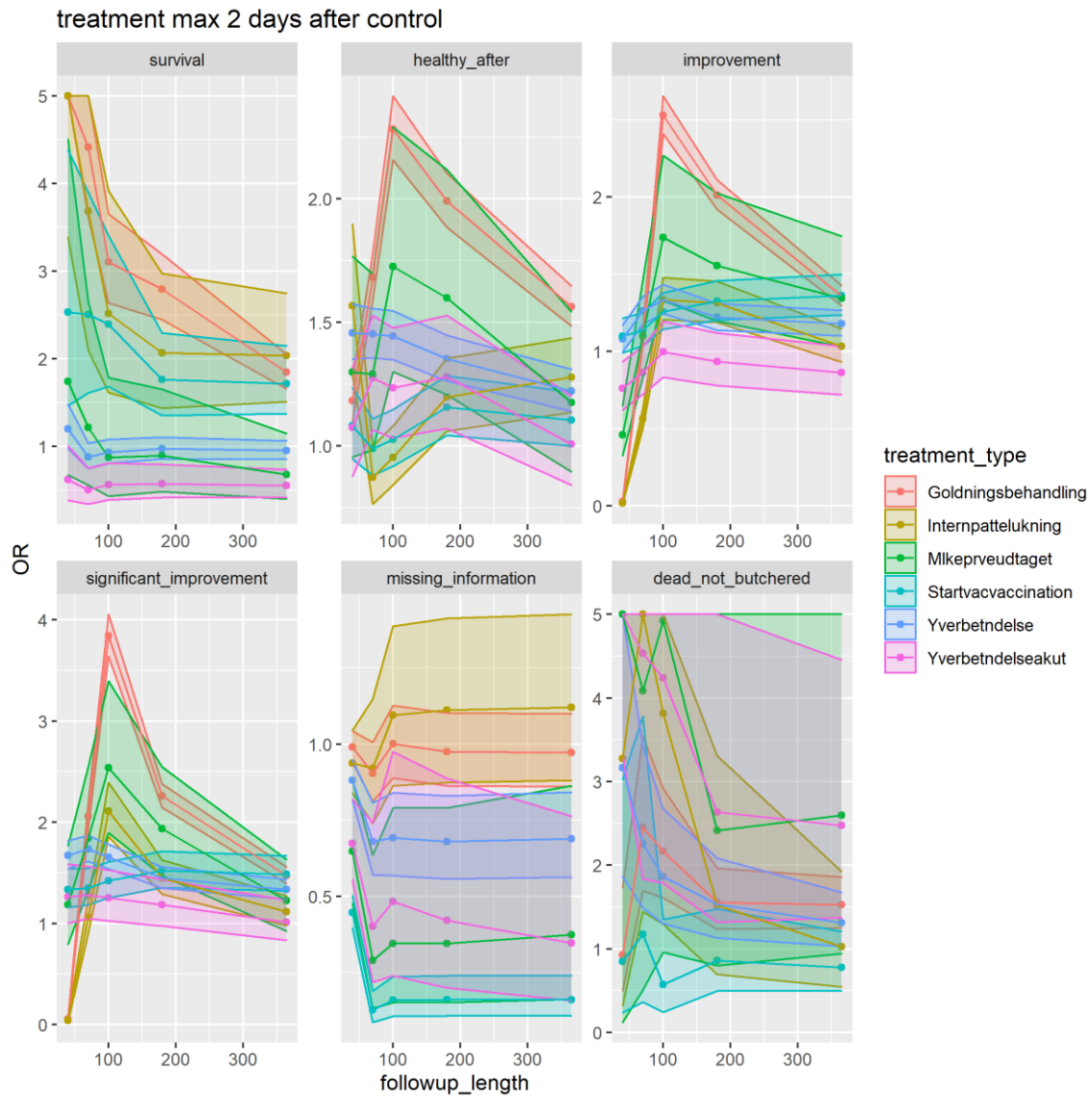
Ved at bruge det matchede datasæt fra denne analyse som træningsdata i prædiktionsmodellen får man flere fordele. For det første er det matchede datasæt allerede oprenset, og indeholder mange indførte variable om sygdomshistorikken tilgængelig til hver ydelseskontrol og behandlingstidspunkt. For det andet, og måske vigtigere, kan prædiktionsmodellen bygge videre på den adskillelse af sygdomsgrad og behandlingsvalg, som det matchede datasæt indeholder. Generelle prædiktionsmodeller såsom random forest kan i princippet godt selv adskille dem, men det er langt fra sikkert, at den gør det uden kolossale mængder data (og tid til at træne modellen). Hvis ikke sygdomsgrad og behandlingsvalg adskilles vil prædiktionsmodellen formodentligt sige at det er dårligt at behandle fordi behandlede dyr generelt er sygere end ikke-behandlede dyr. Ved at bruge det matchede datasæt, er der altså større sandsynlighed for at prædiktionsmodellen giver meningsfulde anbefalinger.



Figur 2: Estimer for odds ratio for behandlingens effekt målt med 6 forskellige responsvariabler. De 7 farver angiver hvor mange dage der højst er mellem den seneste ydelseskontrol og behandlingstidspunktet. De 5 punkter på X-aksen viser hvor mange der går indtil responsvariablen bliver bestemt. Alt i alt er der kørt $5 \cdot 6 \cdot 7 = 210$ forskellige logistiske regressionser som alle har givet en separat odds ratio.



Figur 3: Denne graf viser odds ratio for behandling stratificeret på forskellige behandlinger, responsvariabler samt længden på opfølgingsperioden. Graferne er trunkeret til at være mindre end 5. Denne graf er lavet på data, hvor der højst er 10 dage mellem seneste ydelseskontrol og behandlingstidspunktet.



Figur 4: Denne figur viser det samme som Figur 3 bortset fra at der her højst er 2 dage mellem behandlingstidspunkt og seneste ydelseskontrol.

Sensitivity Statistics:

Partial R2 of treatment with outcome: 0.0372

Robustness Value, $q = 1$: 0.1782

Robustness Value, $q = 1$, $\alpha = 0.05$: 0.1746

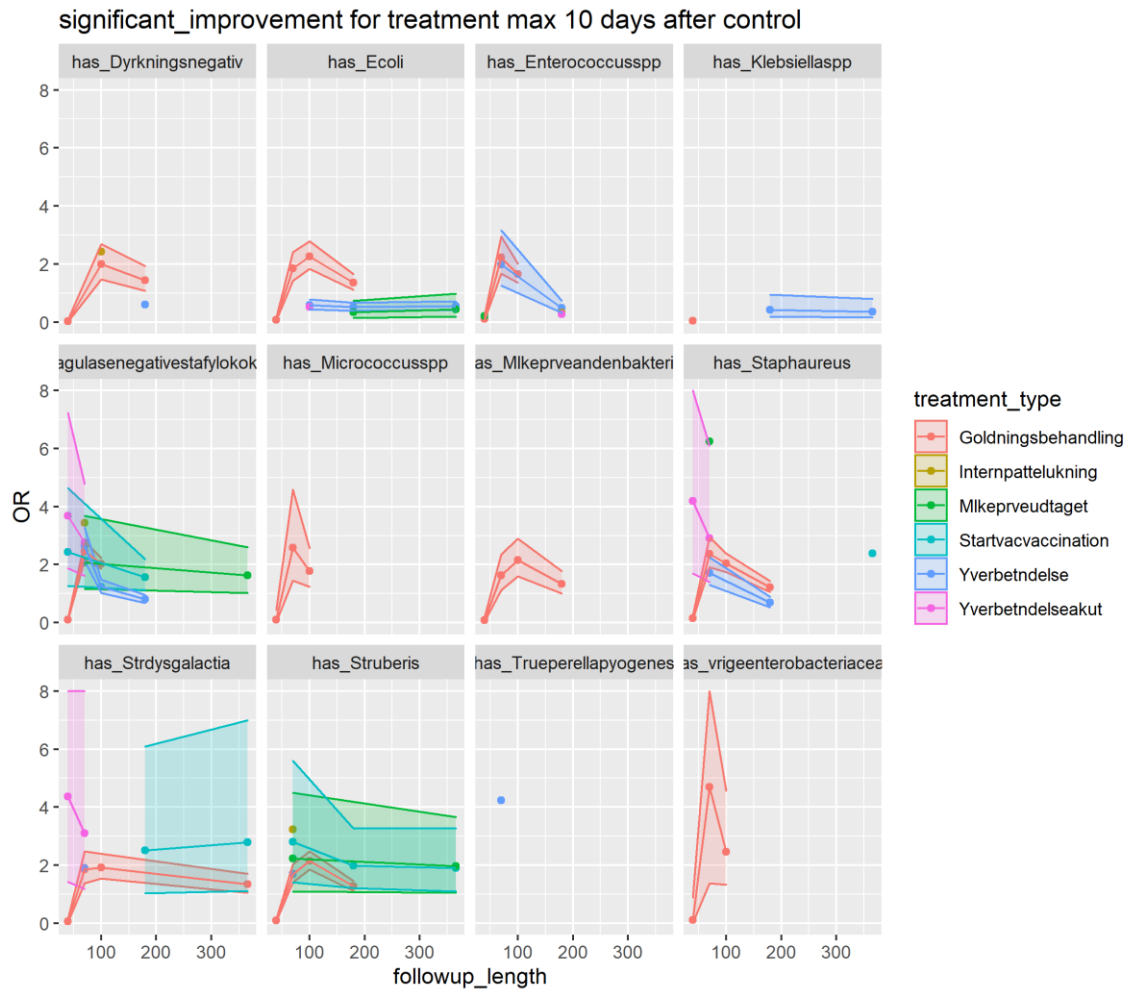
Verbal interpretation of sensitivity statistics:

-- Partial R2 of the treatment with the outcome: an extreme confounder (orthogonal to the covariates) that explains 100% of the residual variance of the outcome, would need to explain at least 3.72% of the residual variance of the treatment to fully account for the observed estimated effect.

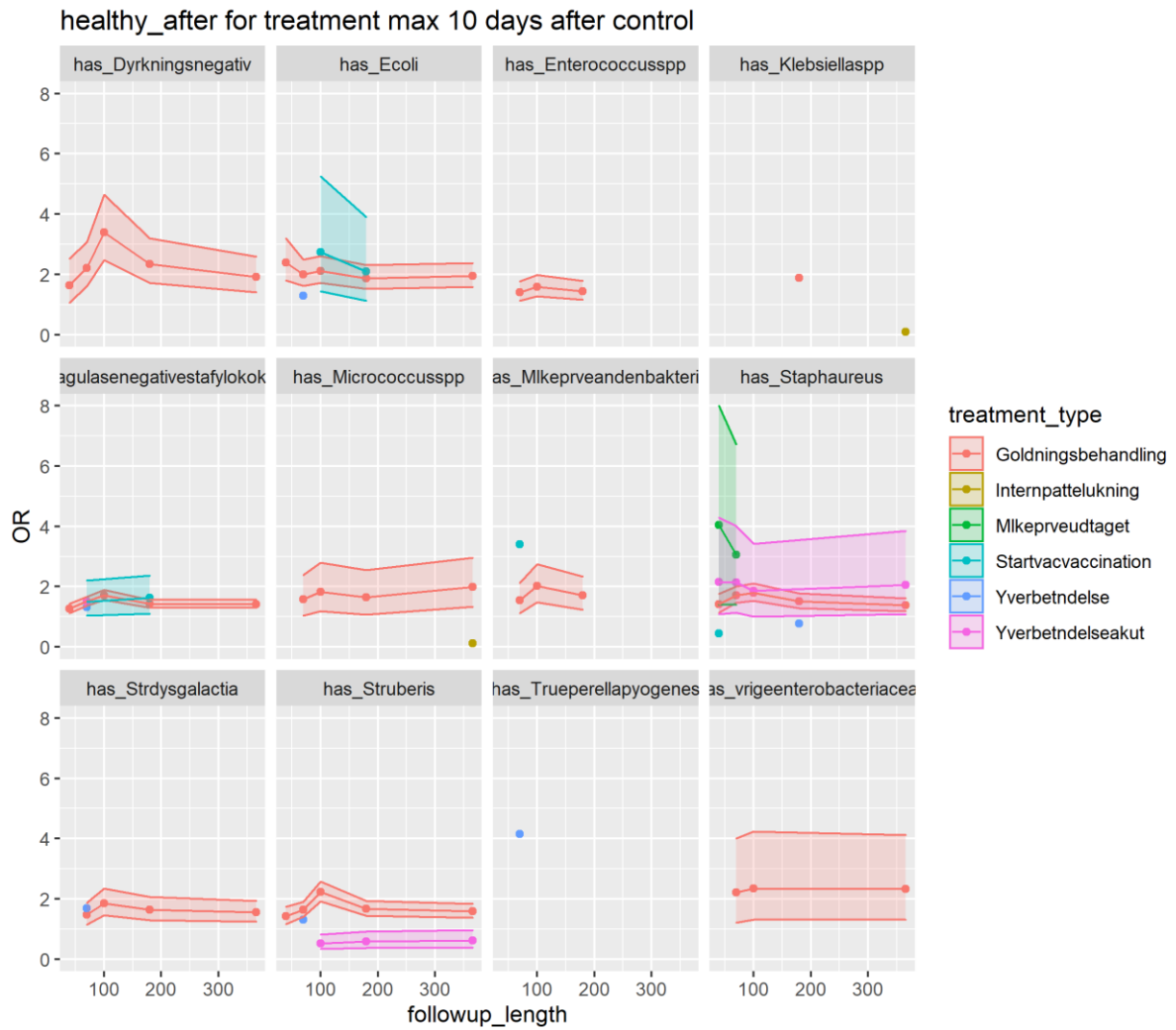
-- Robustness Value, $q = 1$: unobserved confounders (orthogonal to the covariates) that explain more than 17.82% of the residual variance of both the treatment and the outcome are strong enough to bring the point estimate to 0 (a bias of 100% of the original estimate). Conversely, unobserved confounders that do not explain more than 17.82% of the residual variance of both the treatment and the outcome are not strong enough to bring the point estimate to 0.

-- Robustness Value, $q = 1$, $\alpha = 0.05$: unobserved confounders (orthogonal to the covariates) that explain more than 17.46% of the residual variance of both the treatment and the outcome are strong enough to bring the estimate to a range where it is no longer 'statistically different' from 0 (a bias of 100% of the original estimate), at the significance level of $\alpha = 0.05$. Conversely, unobserved confounders that do not explain more than 17.46% of the residual variance of both the treatment and the outcome are not strong enough to bring the estimate to a range where it is no longer 'statistically different' from 0, at the significance level of $\alpha = 0.05$.

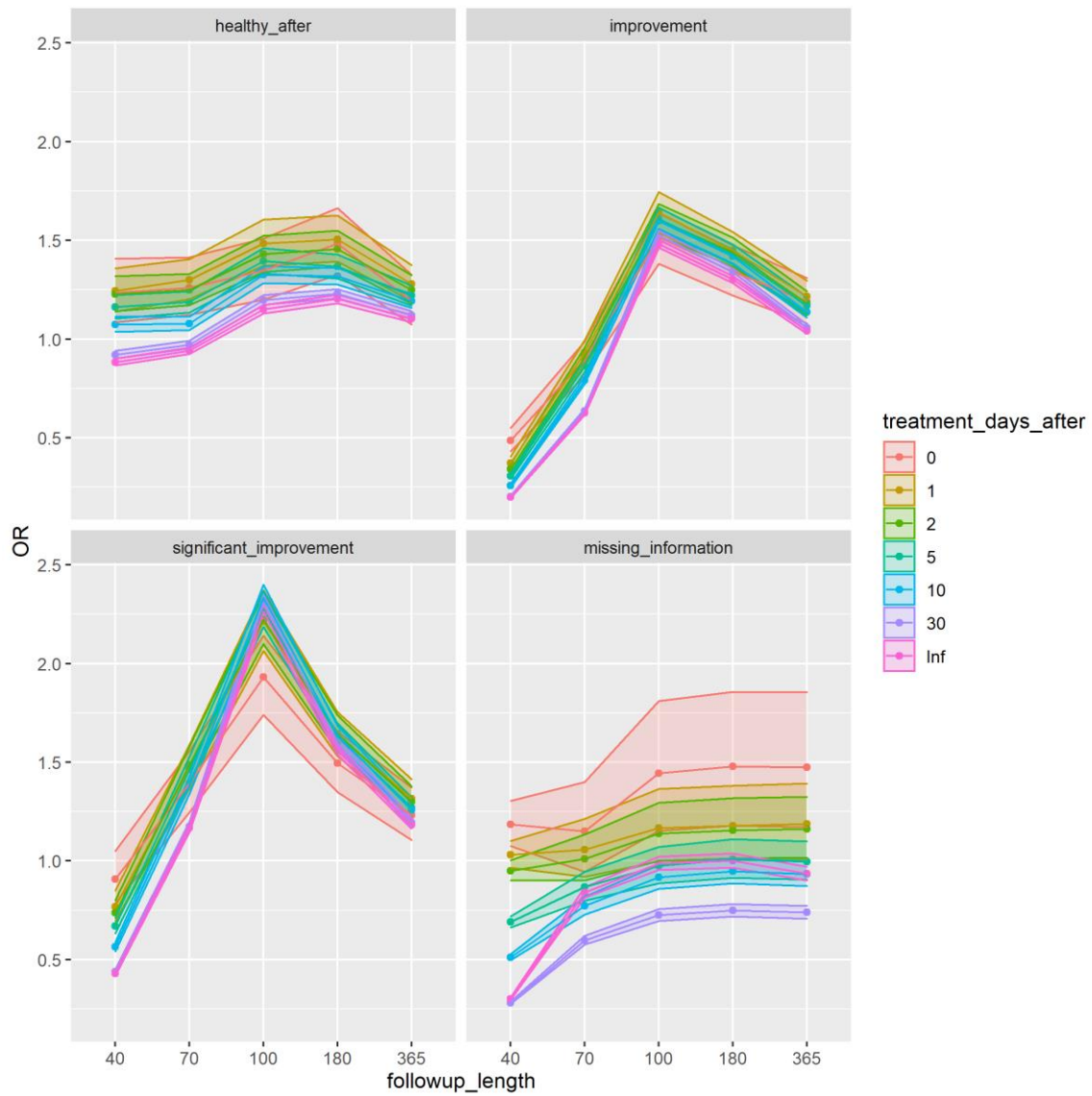
Figur 5: R-udskrift om sensitivtetsanalysen. De vigtigste tal står i de 4 øverste linjer. Se Diskussion for fortolkning.



Figur 6: Odds ratio for at behandling virker stratificeret på den specifikke behandlingstype, dyrkningskode og dage indtil followup. Responsvariablen er her hvorvidt celtallet er halveret ved followup ("significant_improvement"). Afstanden mellem seneste ydelseskontrol og behandlingstidspunkt er her højst 10 dage.



Figur 7: Samme graf som Figur 6 bortset fra at odds ratios omhandler responsvariablen, der angiver om dyret er rask (dvs. celletal under 250,000).



Figur 8: Dette plot viser en delmængde af de samme odds ratios som Figur 2 – her er dog kun inkluderet dyr under 1200 dage.

Referencer

Cinelli, C. and Hazlett, C. (2020), Making sense of sensitivity: extending omitted variable bias. *J. R. Stat. Soc. B*, 82: 39-67. <https://doi.org/10.1111/rssb.12348>

Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, Vol. 42, No. 8, pp. 1-28. <https://doi.org/10.18637/jss.v042.i08>