

Forslag til håndtering av falske positive og negative lusetellinger ved lave lusetall



Forslag til håndtering av falske positive og negative lusetellinger ved lave lusetall

Innhold

Sammendrag	2
Bakgrunn	2
Materiale og metoder	3
Resultater	4
Diskusjon	6

Forfattere

Thijs Christiaan van Son
Anja Bråthen Kristoffersen
Hildegunn Viljugrein
Kari Olli Helgesen
Lars Qviller
Peder Jansen

Bestilt av Nærings og Fiskeridepartementet

ISSN 1890-3290

© Veterinærinstituttet 2016

Design omslag: Reine Linjer

Foto forside: R. Svendsen, UW photo

Sammendrag

Denne rapporten tar for seg et forslag til løsning for hvordan oppdrettsnæringen og myndighetene kan håndtere falske resultat av lakselustellinger. Det er spesielt lagt vekt på å gjøre seg mindre sårbar for falske positive tellinger. Grenseverdien for tellingene som er brukt i dette arbeidet er satt til et gjennomsnitt på 0,20 hunnlus pr. fisk på en lokalitet.

Falske positive lakselustellinger i denne sammenhengen vil si at man feilaktig konkluderer at en lokalitet har 0,20 voksne hunnlus eller mer i gjennomsnitt per fisk, når gjennomsnittet i realiteten er 0,20 (eller under). Den skisserte løsningen baserer seg på Monte Carlo-simuleringer av tenkte lusetellinger som har en negativ binomialfordeling med en gitt variabilitet estimert fra et kjent, reelt datasett.

Det konkluderes med at utfordringene med falske positive telleresultat kan løses ved å gi oppdrettsnæringen muligheten til å telle gjennomsnittsverdier over 0,20, men under 0,30, to ganger på rad, uten at dette regnes som brudd på vilkåret om å holde lave nivå av lakselus. Dersom det telles gjennomsnittsverdier over 0,20, men under 0,30 tre ganger på rad, anses det som at gjennomsnittlig lusetall i populasjonen er over grenseverdien på 0,20.

Hvis en lokalitet observerer et gjennomsnitt på 0,30 eller høyere én gang anses det også som at gjennomsnittlig lusetall i populasjonen ligger over grenseverdien på 0,20.

Sannsynligheten for å observere verdier mindre enn 0,20 når det forventede gjennomsnittet av hunnlus er fra 0,20 til 0,30 (altså falske negative tellinger) synker fra henholdsvis 55,7 % til 7,6 %.

Bakgrunn

Oppdrettsnæringen ble i 2015 tilbudt fem prosent økning i biomasseproduksjon av laks, ørret og regnbueørret mot at innehaverne av akvakulturtillatelser klarer å overholde strenge krav til lakseluskontroll. Tilbudet er regulert av, og kravene til lakseluskontroll er definert i 'Forskrift om auke av maksimalt tillaten biomasse for løyve med laks, aure og regnbogeaure i 2015 (FOR-2015-06-17-817)'. En økning i biomasseproduksjon forplikter lokaliteten til å holde nivåene av kjønnsmodne hunnlus under 0,20 i gjennomsnitt per fisk. Dette er et strengere krav enn hva som ligger i den generelle forskriften om lakselusbekjempelse som krever færre enn 0,50 kjønnsmodne hunnlus i gjennomsnitt per fisk. I tillegg reguleres tilbudet om økt biomasseproduksjon av at man maksimalt kan bruke to medikamentelle behandlinger mot lakselus per produksjonssyklus. Dersom man bryter vilkårene i forskriften fire eller flere ganger innenfor en periode på to produksjonssykluser, trekkes tilbudet om økt biomasseproduksjon tilbake.

Oppdrettsnæringen har meldt interesse for denne ordningen, men har uttrykt bekymring for såkalte falske positive lusetellinger, dvs. at man feilaktig konkluderer at en lokalitet har 0,20 hunnlus eller mer i gjennomsnitt per fisk, når den i realiteten ikke har det. På den andre siden ønsker myndighetene også å sikre seg mot den omvendte problemstillingen, nemlig falske negative lusetellinger. Falske negative lusetellinger oppstår når man konkluderer at en lokalitet ikke har over 0,20 hunnlus når den i realiteten har det.

I denne rapporten har vi skissert en løsning for å sikre at en kan unngå falske positive tellinger når grenseverdien er så lav som 0,2 voksne hunnlus pr fisk i gjennomsnitt. Ved bruk av Monte Carlo-simuleringer skisserer vi mulige scenarier som reduserer risikoen for falske tellinger. I tillegg ser vi på hvilket observert lusegjennomsnitt fra en enkelt telling som med høy sikkerhet angir at grensen på 0,20 hunnlus er overskredet, dvs. når vi anser at man med rimelig sikkerhet har en ekte positiv telling. Våre simuleringer bygger på og kan anses som en videreføring av Norsk Regnesentrals rapport om falske positive tellinger som ble gjort på oppdrag fra Sjømat Norge (1).

Definisjoner av essensielle uttrykk brukt i rapporten*Falske positive tellinger:*

At man observerer at en lokalitet har 0,20 hunn lus eller mer i gjennomsnitt per fisk, når den i realiteten har mindre.

Falske negative tellinger:

At man observerer at en lokalitet har mindre enn 0,20 hunn lus i gjennomsnitt per fisk, når den i realiteten har mer.

Forventet/reelt gjennomsnitt av lakselus:

Den virkelige, reelle verdien av gjennomsnittlig antall hunn lus på en lokalitet antatt i simuleringene.

Observert gjennomsnitt av lakselus:

Det faktiske gjennomsnittet som observeres i simuleringene ved et tenkt uttrekk av fisk (for eksempel 100 fisk fra fem merder) ved en lokalitet.

Materiale og metoder

Det ble gjort Monte Carlo-simuleringer for å teste sannsynligheten for å overskride grensen på 0,20 hunn lus (dvs. sannsynligheten for å oppnå falske positive lusetellinger). De samme simuleringene ble også brukt til å se på falske negative lusetellinger. Simuleringene ble basert på fordelingsparametere fra et datasett vi jobber med i andre sammenhenger knyttet til forskning og forvaltning. Dette datasettet inneholder tellinger av hunn lus på hver enkelt fisk innen ulike merder og på ulike lokaliteter. Fra dette datasettet valgte vi data med gjennomsnitt hunn lus mellom 0,05 og 0,40, fordi vi er interessert i sannsynligheten for falske positive når gjennomsnittlig antall hunn lus er rundt 0,20.

Datasettet ble sjekket for tilstedeværelse av overhyppighet av fisk med null lus (zero-inflation) ved å tilpasse en modell ved hjelp av funksjonen *zeroinfl* (2) i pakken *psci* (3) i statistikkprogrammet R (4). Modellestimatene viste tydelig at dataene ikke var zero-inflated noe som gjorde at vi kunne bruke en generalisert lineær modell (GLM) med negativ binomialfordeling (funksjonen *glm.nb* i pakken *MASS* (5) i R) for å estimere spredningsparameteren theta. Thetaen i en negativ binomialfordeling angir variabiliteten i datasettet og vi trenger et estimat av denne for å kunne gjøre våre simuleringer. GLM-modellen inneholdt kun skjæringspunkt (intercept), ingen prediktorer.

Som en kontroll ble thetaen estimert, på samme måte som beskrevet over, basert på andre utvalg av datasettet som lå tettere rundt 0,20, men dette hadde ingen effekt på dens verdi. I tillegg gjennomførte vi sensitivitetsanalyser på ulike verdier av theta for å vurdere hvordan dette påvirket våre simuleringer. Sensitivitetsanalysene ble gjennomført ved å kjøre simuleringene våre med verdier for theta over og under dens estimerte verdi.

Negativ binomialfordeling ble brukt da vi ved tidligere arbeider med lakselus har sett at lakselusdata ofte blir bedre tilpasset til modellen med denne fordelingen enn med Poissonfordeling. For Poissonfordeling antas variansen og forventningen å være lik. Dette stemmer dårlig med lusedata hvor variansen blir høyere enn forventningen. Allikevel har vi valgt å sammenlikne resultatene med hva vi hadde fått hvis vi hadde valgt en Poissonfordeling. En Poissonfordeling ville vært det naturlige valget dersom vi ikke hadde noen data som kunne brukes for å bestemme theta-verdien i en negativ binomialfordeling.

For å simulere en lakselustelling brukte vi funksjonen *rnegbin* (*MASS/R*). Ved hjelp av *rnegbin* genererte vi randomiserte lakselustellinger på 100 individer (for eksempel 20 fisk i fem merder) med forventninger mellom 0,05 og 0,40 og theta som estimert i GLM-modellen med negativ binomialfordeling nevnt ovenfor.

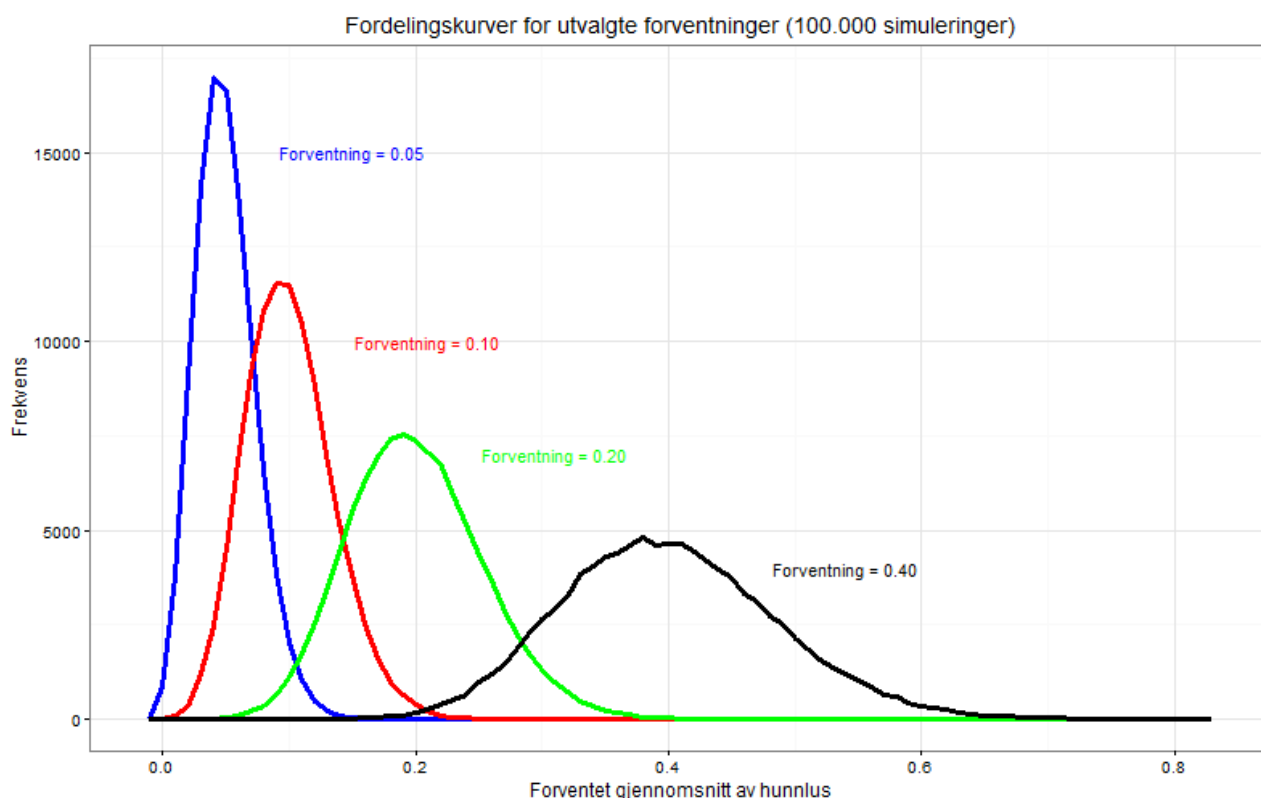
Her representerer forventningene reelle gjennomsnittsverdier av hunnlus på en lokalitet. En simulert lakselustelling ble avsluttet med å ta gjennomsnittet av de hundre hunnlusverdiene. Dette gjennomsnittet representerer verdiene som rapporters ukentlig til myndighetene. For å si noe om forventet variabilitet i disse gjennomsnittstallene ble de simulerte lakselustellingene repetert 100.000 ganger.

Resultater

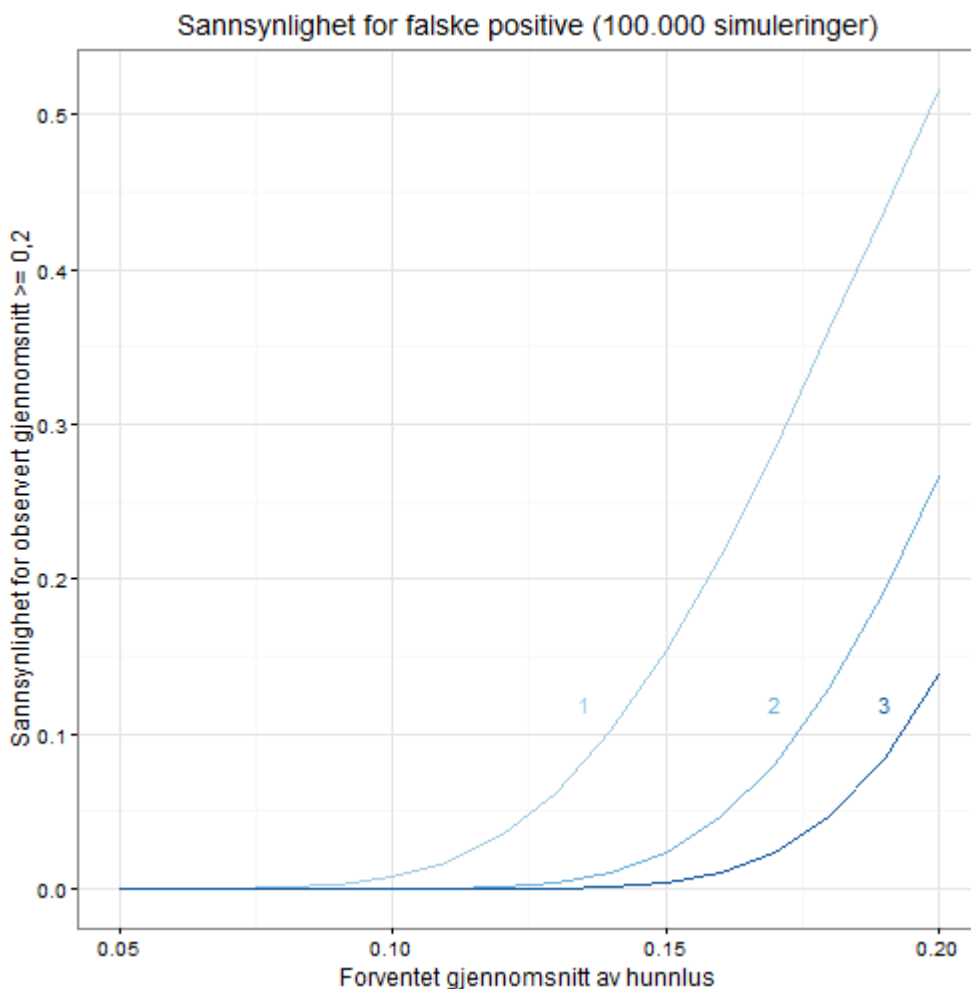
Den negativ binomialfordelte GLM-en på datasettet som inneholdt data med gjennomsnittlig hunnlus mellom 0,05 og 0,40 estimerte en $\theta = 0,5$. Denne verdien for θ ble derfor brukt i simuleringer av lakselustellinger. Vi gjennomførte sensitivitetsanalyser av simuleringene av lakselustellinger ved å endre verdiene for θ til 0,3 og 0,7 uten at dette påvirket våre simuleringer. Dette viser at simuleringene våre er robuste.

Hver enkelt simulert lakselustelling ble repetert 100.000 ganger for alle forventede gjennomsnittlige hunnlusverdier fra en startverdi = 0,05 og deretter med en økning på 0,01 fram til en sluttverdi = 0,40, dvs., 100.000 Monte Carlo-repeteringer ble gjennomført 36 ganger. Disse 100.000 repetisjonene gir et godt bilde av variabiliteten i fordelingen av simulerte observerte, gjennomsnittlig antall hunnlus ved ulike reelle (dvs. forventede) verdier for gjennomsnitt av hunnlus (Figur 1).

For alle forventede hunnlusverdier mellom 0,05 og 0,20 har vi sett på sannsynligheten for å få falske positive lusetellinger over 0,20. Disse sannsynlighetene er også ganget opp to og tre ganger med seg selv for å illustrere sannsynligheten for å ha falske positive lusetellinger over 0,2 flere ganger på rad (Figur 2).



Figur 1. Fordelingskurver for utvalgte forventninger av gjennomsnittlig hunnlus på lokalitetsnivå basert på 100.000 Monte Carlo-simuleringer med $\theta = 0,5$.



Figur 2. Sannsynlighet for falske positive tellinger av hunnlus (dvs., observert gjennomsnitt av hunnlus $\geq 0,20$), gitt ulike forventede/reelle gjennomsnitt av hunnlus (x-akse). Det mest interessante er hva som skjer med sannsynligheten for falske positive når man nærmer seg 0,20. Tallene 1, 2 og 3 indikerer ulike sannsynlighetskurver for henholdsvis 1, 2 og 3 observasjoner på rad med falske positive tellinger av hunnlus gitt samme forventede/reelle gjennomsnitt (for eks., om man teller 0,20 hunnlus en, to eller tre ganger på rad).

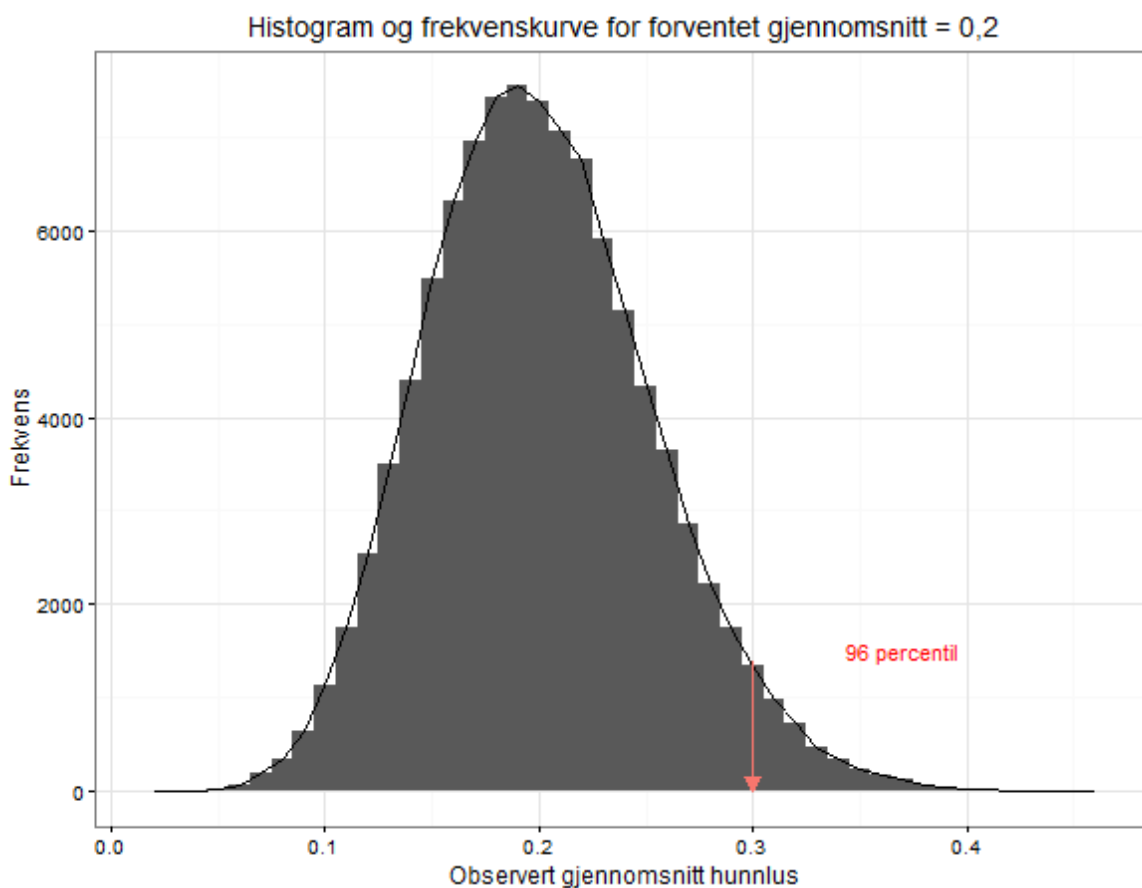
Et viktig spørsmål å svare på er når kan vi med høy sannsynlighet si at en observert gjennomsnittlig verdi for hunnlus er over grensen på 0,20. Dersom man tar utgangspunkt i fordelingskurven for de simulerte dataene når forventet gjennomsnitt av hunnlus = 0,20 kan vi se at en observert verdi på 0,30 ligger på 96 percentilen, dvs at bare fire prosent av de simulerte observasjonene har verdier over 0,30 (Figur 3). Vi anser denne verdien som god for å konkludere at gjennomsnittet med stor sannsynlighet er over 0,20.

Interesseområdet for falske negative tellinger blir begrenset til verdier av forventede gjennomsnitt av hunnlus mellom 0,20 og 0,30. I dette område vil det kunne forekomme at en lokalitet som reelt har 0,20 til 0,30 lus, vil kunne bli feilklassifisert ved gjentatt testing. I henhold til Figur 4 ser vi at sannsynligheten for én slik falsk negativ telling synker fra 55,7 % til 7,6 % når det forventede gjennomsnittet av hunnlus er henholdsvis 0,20 og 0,30. Sannsynligheten for at en lokalitet med reelt lusetall mellom 0,20 og 0,30 skal bli feilklassifisert på grunn av falske negative resultat ved gjentatte tellinger, må sees i sammenheng med at lakselus utvikler seg i forhold til vanntemperaturen. Ved varmt vann og ingen gjennomførte tiltak, vil lusetallet naturlig øke og bidra til at sannsynligheten for flere falske negative resultat over gjentatte tellinger blir lavt. Når det er kaldt i vannet vil lusebestanden naturlig minke over tid og sannsynligheten for falske negative over flere tellinger vil også her bli lavt. Falske negativer vil derfor i dette systemet ha liten praktisk betydning. I et temperaturuavhengig system vil sannsynligheten for falske negativer kunne fått større betydning.

Diskusjon

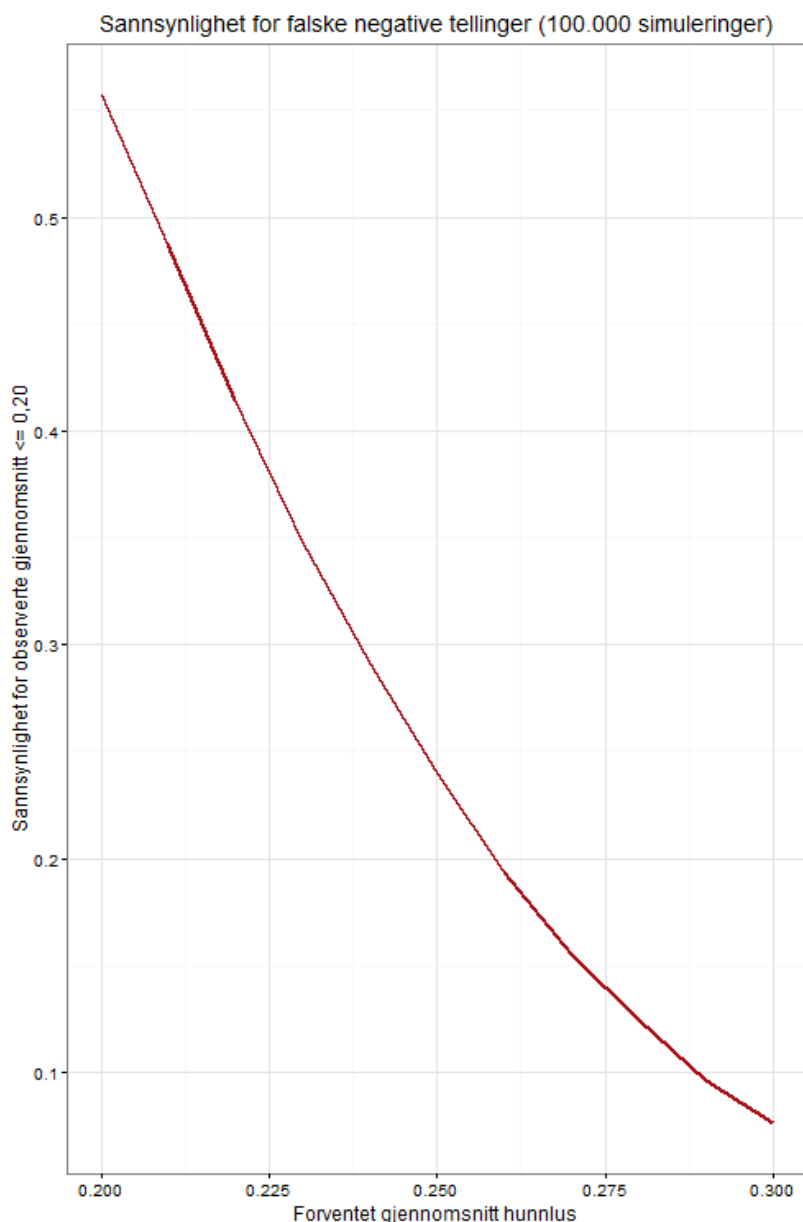
Monte Carlo-simuleringene våre viser at man kan redusere sannsynligheten for falske positive lusetellinger dersom man sier at man må observere lusetall over 0,20 hunn lus tre ganger på rad for at man med sikkerhet kan si at man er over grenseverdien på 0,20. Dersom man observerer et gjennomsnittlig lusetall på akkurat 0,20 én gang er sannsynligheten for en falsk positiv 51,4 %. Denne sannsynligheten reduseres til 26,4 % og 13,6 % dersom man observerer akkurat 0,20, henholdsvis to eller tre ganger på rad. Det er verd å merke seg at oppdretterne ifølge forskriften skal ha et gjennomsnitt som er mindre enn 0,20. Sannsynlighetene for falske positive reduseres betydelig allerede dersom man forventer et gjennomsnitt på 0,18 tre ganger på rad. Da blir sannsynligheten 4,8 % (se Figur 2). Det betyr at man kan forvente å observere en serie med tre falske positive på rad én gang hver 20. gang man teller slike tripler.

De samme simuleringene viser også at når man overstiger et observert lusetall på 0,30, er sannsynligheten så stor for at lusetallet faktisk er likt eller over 0,20 at man kan sette dette som grense for når betingelsene for god lusekontroll er brutt, selv etter bare én telling over 0,30. Begrunnelsen for dette er som følgende: Hele tiden gitt at man har et sant gjennomsnitt av hunn lus lik 0,20, er sannsynligheten for at man skal observere 0,30 4 % (100 - 96; Figur 3). Dette betyr i praksis at man kan forvente å observere 0,30 hver 25. telling (100/4). Dersom man da tenker man at i løpet av en produksjonssyklus kan ha opp mot 52 uker med luseproblemer (luseproblemene er sesongavhengige), vil det være sannsynlig at man observerer 0,30 to ganger (52/25) i løpet av en slik syklus, igjen gitt at det reelle gjennomsnittet er 0,20 alle de 52 ukene. I tillegg er det viktig å påpeke at sjansen for at man skal telle over 0,30 to ganger i løpet av en produksjonssyklus synker betraktelig med synkende reelt gjennomsnitt.



Figur 3. Fordelingskurven av 100.000 simulerte observerte hunn lus-gjennomsnitt når forventet/reelt gjennomsnitt av hunn lus = 0,20. Den røde pilen på 0,30 angir 96 percentilen for denne fordelingskurven, dvs. at man i 4 % av tilfellene vil observere en verdi over 0,30 selv om den reelle, forventede verdien er 0,20.

En god simulering av lusetellinger forutsetter et godt estimat av spredningsparameteren theta. Derfor har vi gjort sensitivitetsanalyser med hensyn på hva som skjer med våre simuleringer dersom theta endres til 0,3 og 0,7. Disse sensitivitetsanalysene viste at resultatene av simuleringene er lite følsomme for slike endringer i spredningsparameteren. Vi har også gjort tilsvarende simuleringer ved bruk av Poissonfordelte data. Disse viste at resultatene ble tilnærmet identiske med våre negativ binomialfordelte resultater. Dog er simuleringer basert på Poissonfordelte data noe mindre variable, noe som uttrykkes i en smalere fordeling. Mens en telling på 0,30 hunnlus ved et forventet gjennomsnitt på 0,20 ligger på 96 percentilen i våre simuleringer basert på negativ binomialfordelingen, lå tilsvarende telling ved 98 percentilen for simuleringer basert på Poissonfordeling. Dette viser at den negative binomialfordelingen evner å inkorporere mer av variabiliteten i lusetellingene enn Poissonfordelingen.



Figur 4. Sannsynlighet for falske negative tellinger av hunnlus (dvs., observert gjennomsnitt av hunnlus $\leq 0,20$), gitt ulike forventede/reelle gjennomsnitt av hunnlus mellom 0,20 og 0,30.

På bakgrunn av resultatene fra Monte Carlo-simuleringene anbefaler Veterinærinstituttet at en kan akseptere at en lusetelling overstiger grensen på 0,20 for gjennomsnittlig antall hunnlus per fisk ved to påfølgende tellinger, men overskridelser ved tre tellinger på rad må anses som at lusetallet i gjennomsnitt

er reelt høyere enn 0,20. På samme måte vil en observert verdi på 0,30 eller mer allerede etter én telling være å anses som at reelt nivå er mer enn 0,20.

Referanser

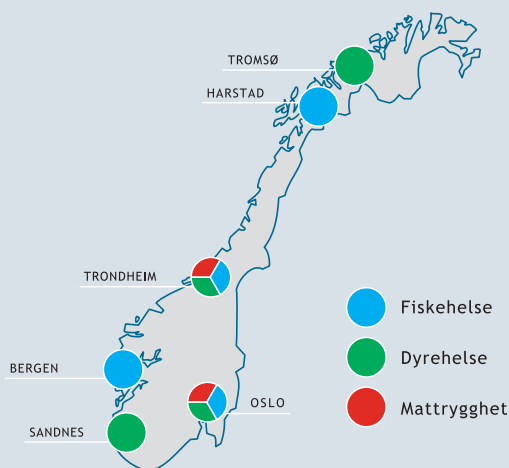
1. Løland A. Falske positive i lusetellinger? SAMBA/17/16, Norsk Regnesentral 2016.
2. Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R. J Stat Soft 27 (2008) URL <http://www.jstatsoft.org/v27/i08/>
3. Jackman S. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University. Stanford, California. R package version 1.4.9. URL <http://pscl.stanford.edu/> 2016.
4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016
5. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0 2002.

Faglig ambisiøs, fremtidsrettet og samspillende - for én helse!

Veterinærinstituttet er et nasjonalt forskningsinstitutt innen dyrehelse, fiskehelse, mattrygghet og fôrhygiene med uavhengig kunnskapsutvikling til myndighetene som primæroppgave.

Beredskap, diagnostikk, overvåking, referansefunksjoner, rådgivning og risikovurderinger er de viktigste virksomhetsområdene. Produkter og tjenester er resultater og rapporter fra forskning, analyser og diagnostikk, og utredninger og råd innen virksomhetsområdene. Veterinærinstituttet samarbeider med en rekke institusjoner i inn- og utland.

Veterinærinstituttet har hovedlaboratorium og administrasjon i Oslo, og regionale laboratorier i Sandnes, Bergen, Trondheim, Harstad og Tromsø.



Fiskehelse



Dyrehelse



Mattrygghet



Oslo
postmottak@vetinst.no

Trondheim
vit@vetinst.no

Sandnes
vis@vetinst.no

Bergen
post.vib@vetinst.no

Harstad
vih@vetinst.no

Tromsø
vitr@vetinst.no

www.vetinst.no



Veterinærinstituttet
Norwegian Veterinary Institute