



iFELT

FHF PROSJEKT: 901836

22.03.2024

KONTAKTPERSON: ARNFINN AUNSMO, SALMALYTICS AS (arnfinn@salmalytics.no)

Veileder for automatisk deteksjon av merket laks til bruk for forskning og evaluering av helsefremmende tiltak i felt.

Paul J. Midtlyng¹, Audhild Blomsø², Ingar S. Nerbø², Anne Bakke Fylling³, Eystein Skjerve¹, Arnfinn Aunsmo¹

¹Salmalytics AS, ²Optoscale AS, ³Previwo AS

Bakgrunn

Formålet med denne veilederen er å beskrive og gi innsikt i design og gjennomføring av feltstudier i sjø hvor målet er å avdekke effekt av helsefremmende tiltak (vaksinering, behandlinger etc.) som gjennomføres før utsett. Metoden egner seg for prosjekter der en behandlet testgruppe merkes med fettfinneklipp som kan gjenkjennes automatisk ved bruk av kamera blant umerket (ikke fettfinneklippet) kontrollfisk. Veilederen utgjør første leveranse fra det FHF-finansierte prosjektet «**Studiedesign for kontinuerlig evaluering i felt - iFelt**» (prosjektnr. 901836) som startet våren 2023.

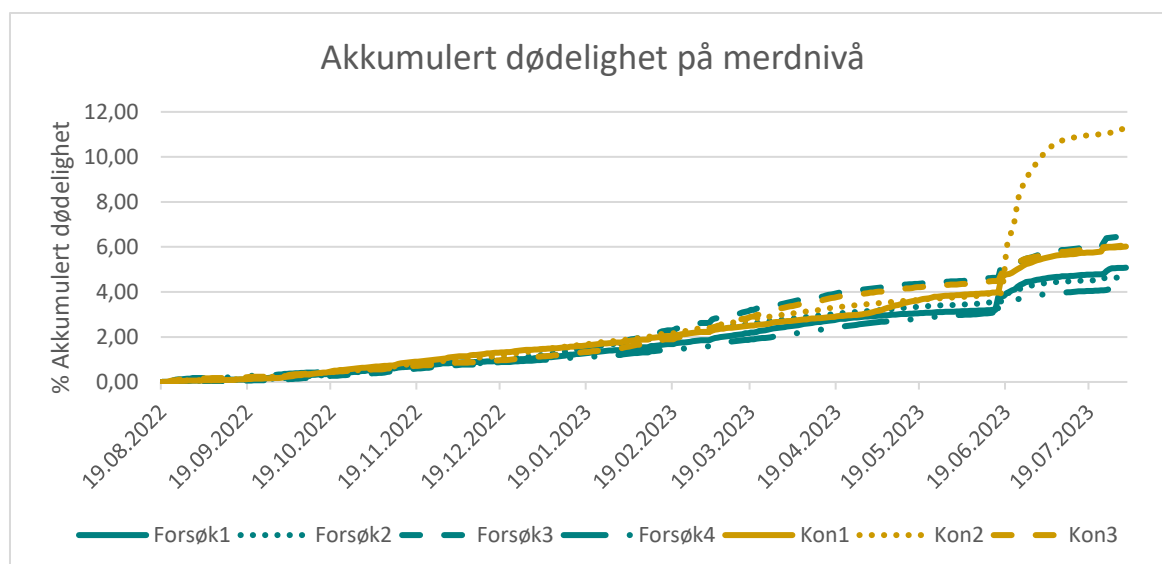
Hvorfor feltstudier?

Feltstudier og -forsøk er nødvendige for å undersøke om biomedisinske tiltak gir ønsket effekt under reelle oppdrettsbetingelser. Hvor store er de faktiske effektene og gir tiltakene uønskede bi-virkninger? Resultater fra feltforhold er nødvendig for å beregne både medisinsk og økonomisk nytte av et tiltak når man tar både effekter, bi-virkninger og kostnader med i en endelig analyse. Ved registrering av nye legemidler (for både mennesker og dyr) er det et lovregulert krav om kliniske utprøvnings, dvs. studier med bruk av det endelige legemiddelet og på virkelige pasienter.

Godt designede feltstudier med tilhørende analyser er nødvendige for å avdekke og kvantifisere effekter og bi-virkninger av legemidler og andre tiltak. Uten godt designede og gjennomførte feltstudier har man begrenset grunnlag for å vurdere om tiltak er fornuftige eller om man sløser bort ressurser. I verste fall kan dårlig evaluerte legemidler gi manglende effekt og betydelige bi-virkninger med fravær av ønsket effekt, redusert produksjon og fiskevelferdsmessige utfordringer som resultat.

Hvorfor har det vært vanskelig å gjennomføre gode feltforsøk?

Mange feltforsøk er basert sammenligning av behandlede merder med kontrollmerder, men det har alltid vært stilt spørsmål ved dette som designprinsipp. Merder inneholder smolt med forskjellig bakgrunn og historikk, det er svært vanskelig å lage et design med sammenlignbare behandlede merder og kontrollmerder. Antallet merder blir alltid lavt, og det vil nesten alltid oppstå hendelser i enkeltmerder som kan påvirke sammenligningsgrunnlaget. Spesielt ved håndtering av fisken i forbindelse med for eksempel avlusning, kan tilfeldige effekter skape store forskjeller mellom merder som i utgangspunktet var ganske like. Et eksempel på dette er vist nedenfor, hvor det var høyere dødelighet i en av merdene etter avlusning sammenlignet med de andre (Figur 1).



Figur 1. Akkumulert dødelighet fra et feltoppsett med 4 forsøksmerder og 3 kontrollmerder hvor kontrollmerd 2 får økt dødelighet etter avlusning (Kilde: Feltforsøk i regi av Previwo).

Denne økningen kunne ikke henføres til noen kjent årsak. Smittsomme sykdommer vil alltid ha forskjellig forløp mellom merder slik at merdene aldri blir helt sammenlignbare. Noen ganger er slike effekter store og synlige, men i de fleste tilfellene vil slike tilfeldige effekter være små, men kan likevel i stor grad påvirke resultat i en studie.

For å løse utfordringene med å sammenligne behandlet og ubehandlet fisk bør man bruke forsøk med behandlet og ubehandlet fisk i samme merd. Da oppnås følgende som er avgjørende for å avdekke reelle effekter av et legemiddel eller andre tiltak:

- En kan fordele fisk til behandlet og ubehandlet gruppe tilfeldig ("randomisere") slik at gruppene blir tilnærmet helt like.
- Miljøforhold og håndtering vil være helt likt for begge grupper
- En kan måle et stort antall individer slik at selv små effekter kan avdekkes.

Dersom en skal måle effekt av f.eks. en vaksine er en avhengig av at sykdommen opptrer, noe man som regel ikke kan «bestille». Bi-virkninger fra vaksiner eller andre ulemper ved inngrepet kan man imidlertid måle uavhengig av om sykdommen opptrer.

Til slike studier trenger man et gruppemerke som f.eks. klipping av fettfinne, eller også individmerke med PIT- tags hvor hver fisk har sitt eget nummer, som igjen er koblet til behandling eller ikke. Med gruppemerking får man svar på gjennomsnittseffekter for ei gruppe, med en god design vil dette gi gode svar på effekter og bi-virkninger. Med individmerke kan individuelle fisk identifiseres for måling av vekt eller andre variabler på forskjellige tidspunkt. Noen eksempler på vitenskapelige publikasjoner fra et fåtall tidligere norske studier finnes i litteraturlista. Avlesning av fisker som er gruppemerket eller individmerket kan dessuten skje automatisk ved behandling og ved slakting dersom egnet avlesningsutstyr er montert.

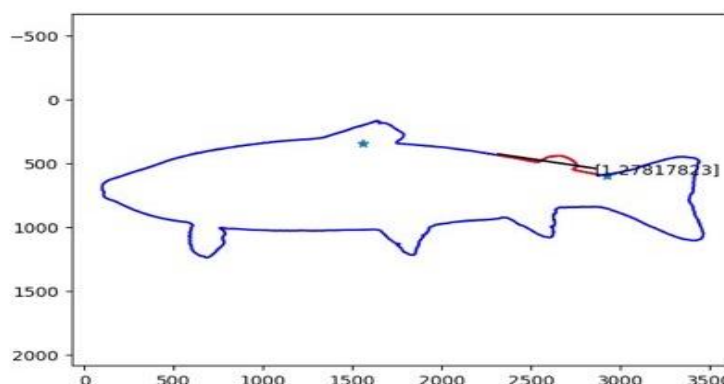
Nye muligheter med automatisert optisk gjenkjenning og kontinuerlig registrering

Forbedret kamerateknologi med tilhørende bildebehandling er i økende bruk i kommersielt lakseoppdrett. Fra å være et hjelpemiddel i manuell overvåkning av livet og føringen i merdene, tilbyr nye software-applikasjoner telling av lus, telling og klassifisering av sår, og estimering av vekt. Bruk av høyoppløselig fargekamera for bildeanalyse, og ulike teknikker for maskinlæring gjør det nå mulig å automatisere gjenkjenningen av relativt komplekse visuelle objekter og mønstre. Dette har vært utgangspunktet for FHF-prosjektet «iFelt» med mål om å undersøke og dokumentere forebygging av sår ved hjelp av en probiotisk bakteriekultur. Samtidig er kunnskap om- og bruk av automatisert teknologi for gjenkjenning av merkede fiskegrupper sentralt i prosjektet. Som del av prosjektet fettfinneklippes en mindre undergruppe laks samtidig som de behandles med bakteriekulturen. Merkingen foregår manuelt mens fisken er under anestesi i forbindelse med vaksinerings, og den merkede gruppen gjenforenes med umerket fisk umiddelbart. En tidligere studie viste at epitelcellene hos laks lukker såret etter fettfinneklipping i løpet av få timer (Andrews og medarb. 2015), noe som tilsier at metoden kan regnes som et «minimalt belastende» inngrep. Under løpende drift vil merdpopulasjonen overvåkes med kamera, mens programvaren fortløpende klassifiserer fisk som «merket» eller «umerket» basert på automatisert bildeanalyse. Samtidig registreres fiskens vektestimat samt eventuelle sår.

Automatisk bildeanalyse

Programvaren som følger med kamerasystemet bruker automatisert bildeanalyse med allerede definerte og validerte kriterier som godkjenner hvert bilde til beregning av henholdsvis vekt, registrering av sår på huden, eller telling av lakselus. En høykvalitets kontur av hver enkelt fisk er basis for at bildene kan brukes til automatisert registreringer av individets gruppetilhørighet med tilhørende estimering av vekt, forekomst av sår og lus.

Klassifiseringssystemet utviklet i prosjektet iFelt startet ved først å sortere et utvalg av bilder manuelt, hentet fra en annen pågående studie hvor halvparten av fisken var fettfinneklippet. Området mellom ryggfinnen og haleroten ble brukt for utvikling av en algoritme – en slags regneformel – som måler hvor mye konturen i dette området avviker fra en rett linje. Dette er illustrert i Figur 2. Gjennom denne prosessen kunne vi skille merket fisk fra de umerkede på en systematisk måte.



Figur 2. Omriss av en fotografert laks med en numerisk verdi som angir graden av avvik fra en rett linje mellom ryggfinne og halerota.

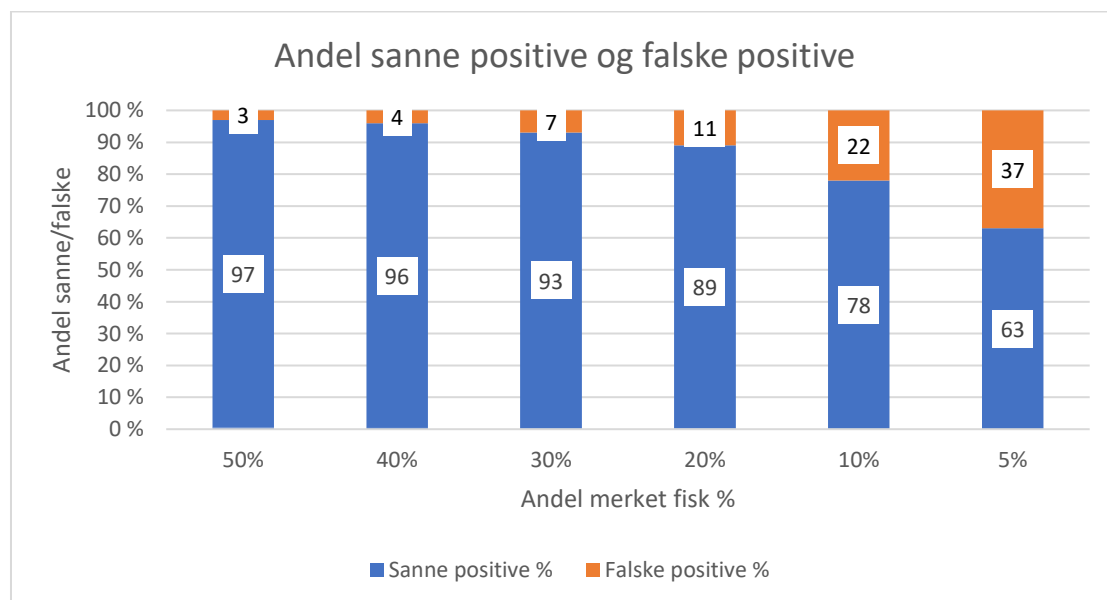
Gruppetilordning

På grunnlag av dette utvalget bilder ble det innledningsvis satt grenser for hvilke algoritme-verdier («fin-Value») som skulle klassifiseres som «merket» eller «umerket». Første gjennomkjøring av bildene viste at 3% av fiskene som automatisk ble klassifisert som «merket» i virkeligheten hadde intakt fettfinne, mens 3,5% av fiskene som var automatisk klassifisert som «umerket» i virkeligheten var fettfinneklippet. Dette kalles i statistisk språkbruk for henholdsvis «andel falske positive» og «andel falske negative». Vi spurte oss derfor: hva betyr slike andeler for en studie som vår?

Konsekvensen av feil gruppetilordning er avhengig av hvor stor andel merket fisk du bruker!

Dersom fordelingen av merket og umerket fisk i merda er 50-50, vil 3% feil tilordning gi samme andel “falske” registreringer for hver av gruppene. Dette endrer seg gradvis når andelen merket fisk blir mindre, ved at andel «falske positive» i større grad påvirker resultatene for den merkede gruppen slik vist i figur 3.

Det er viktig å tenke nøye over den store feilkilden som kan oppstå ved feilaktig tilordning når en planlegger et forsøk med liten andel i «merket» gruppe. Ethvert automatisk gjenkjenningssystem må derfor optimaliseres i forhold til studiens design til å gi et akseptabelt lavt nivå av andel «falske positive». I figur 3 ser man at så lenge fordelingen mellom merkede og umerkede individer i forsøket ligger i området 50/50 til 20/80 vil datagrunnlaget i merket gruppe ikke endres vesentlig. Med 10% merket fisk blir andelen «falske positive» hele 22% dersom 3% av fisken gis feil tilordning, og dette blir enda verre jo lavere andelen merket fisk er. I et oppsett med 10% merket fisk og 1 % feil tilordning utgjør andelen falske positive 8,5% i merket gruppe.

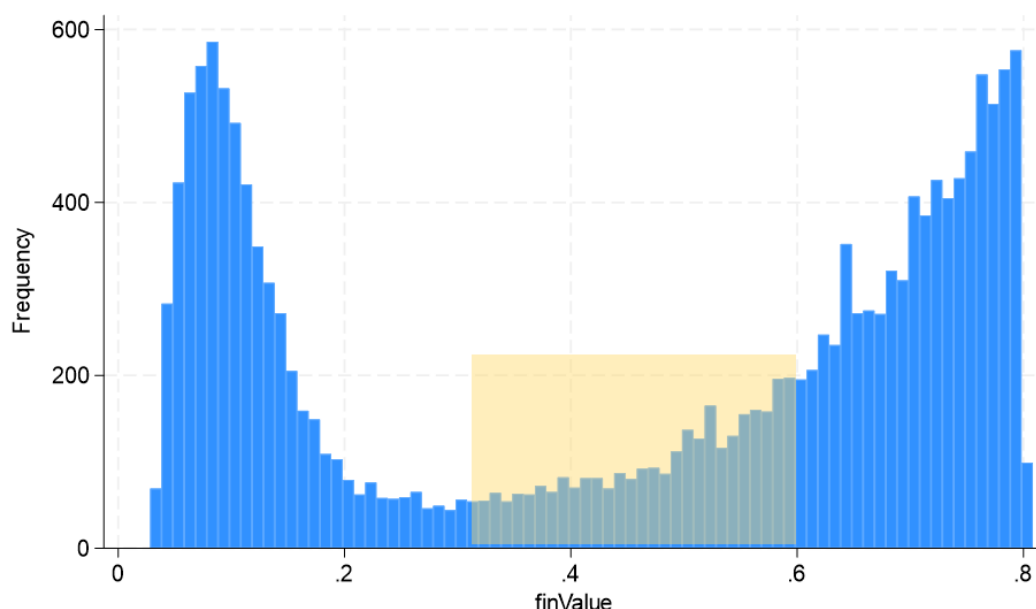


Figur 3. Illustrasjon som viser hvordan konsekvensene ved 3% feilaktig tilordning i merket gruppe er avhengig av andelen merket fisk.

For den umerkede (største) gruppen betyr en liten andel “falske negative” svært lite, ettersom det store flertallet registreringer uansett vil komme fra umerkede individer.

Forbedring av den automatiske tilordningen til forsøksgruppe

Arbeidet i dette prosjektet satte først fokus på optimalisering av kravene til de bildene som kan gi automatisk tilordning til henholdsvis «merket» og «umerket» gruppe, slik som bildevinkelen (en klar silhuett) og kontraster (lysforhold) som gir tilstrekkelig bildekvalitet. Dette førte til at flere av bildene ble forkastet som «ikke sikkert klassifiserbare». Man kan dessuten optimalisere tilordningen til gruppe ved å forkaste bilder som gir et usikkert algoritmeutfall, slik at kun bilder med entydig lave verdier tilordnes merket gruppe. Figur 4 viser et eksempel på hvordan fisk med usikker algoritmeverdi for fettfinne (0.3 – 0.6) kan forkastes for å redusere andel falske positive.



Figur 4. Illustrasjonseksempel som indikerer fordelingen av verdier for fettfinne blant 40 000 bilder fra ei merd. Verdier som er «usikre» (0.3 – 0.6) kan ekskluderes for å redusere feilklassifisering.

Når man har oppnådd et godt utgangspunkt for den automatiske tilordningen av data til begge grupper bør det gjennomføres en kontroll hvor et utvalg godkjente bilder bedømmes og klassifiseres av kyndig personell. Resultatet fra en manuell kontroll eller såkalt annotering kan videre brukes til «maskinlæring» med kunstig intelligens å ytterligere optimalisere tilordningen av merkede individer.

Det er også nødvendig at bildene som tas også er av en tilstrekkelig kvalitet når det gjelder registreringer av de utfallsvariablene studiene skal bruke. I denne studien måtte all fisk som skulle tilordnes gruppe også samtidig ha godkjent måling av vekt og sår.

Design av en studie

Design av en studie er grunnleggende for hva som er mulig å analysere og tolke i ettertid. Design av en studie må skje i forkant av studien og det er fornuftig å lage en «protokoll» som beskriver hva som skal gjennomføres.

- Formål – hva er målet med studien?
- Hvilke forskjeller skal kunne måles i studien med signifikant resultat?

- Hva er tilstrekkelig antall fisk må som registreres for å avdekke minste interessante forskjell?
- Kontroll på gruppetilordning, andelen falske positive.
- Hvordan randomisere for å sikre at fisken i behandlet gruppe og kontrollgruppa blir så like som mulig.
- Hva skal registreres, med beskrivelse av datalagring og datahåndtering.
- Analyse av datamaterialet med relevant statistisk metode.

Gjennomføring av randomisering, behandling og merking

I prosjektet ble fisken badebehandlet med den probiotiske bakteriekulturen Stembiont Vital™ samt finneklippet, i etterkant av vaksinasjon. Ut fra vaksinasjonsmaskinen ble fisken sortert i størrelseskategoriene stor, medium og liten. Fisken som ble behandlet og finneklippet var "medium" sortering, ettersom størsteparten av fisken falt i denne gruppen og dette var eneste vektgruppe som kunne gi to fulle merder á ca. 160 000 fisk i sjø. Vi beregnet at det ville ta omtrent 2 lange arbeidsdager for å fylle et kar med ferdig vaksinert og behandlet medium-fisk i settefiskanlegget, og antallet fisk som ble tatt ut til behandling og finneklipping ble summert underveis for å sikre stor nok andel i løpet av to fulle dager.

Settefiskanlegget hadde to vaksineringsmaskiner gående samtidig, og et antall på rundt 50 fisk som kom ut fra hver vaksineringsmaskin ble behandlet og fettfinneklippet om gangen. Dette kalles «blokk-randomisering». Fettfinnen bør klippes bakfra og framover for å sikre at klippet starter ved finnebasis. Utløpet fra vaksinemaskinene var koblet med et felles utløpsrør, og fisken som skulle merkes ble hentet ut fra utløpsrøret via en Y-kobling som manuelt kunne åpnes og lukkes. Vannet med Stembiont Vital™ var tilsatt ca. 10 promille salt (Stembiont Vital™ inneholder levende bakterier som trenger salt for å holdes i live) og en liten dose vedlikeholds-bedøvelse. Antallet merkede fisker ble notert for hver pulje, og ca. hver 10ende pulje ble også veid.

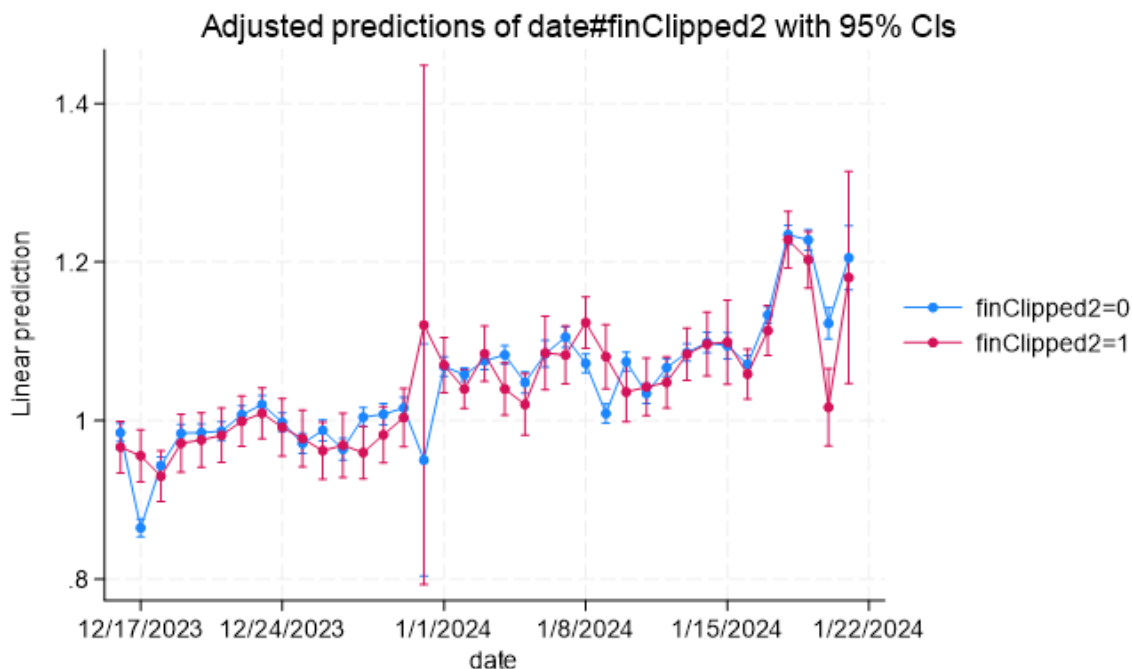
Erfaringene fra gjennomføringen viser at det er viktig å tilpasse prosedyrene for randomisering, behandling og merking til arbeidsflyten på oppdrettsanlegget. Da blir det også lettere å vekke interesse og sikre optimal gjennomføring og oppfølging på det stedet forsøksbehandlingen utføres.

Statistisk styrkeberegning etter optimalisering av automatisk tilordning

Når det er klart hvor strenge kriterier som må settes for at andelen «falske positive» skal bli lavere enn ønsket nivå, bør man estimere hvor mange merkede fisk som sannsynligvis vil bli registrert i løpet av ett døgn eller én uke. Hos hvor mange av den umerkede (kontroll-) gruppen med det utfallet («andel med sår») som skal registreres forekomme i samme tidsperiode? Disse verdiene bør brukes til en såkalt «styrkeberegning» av forsøket, som forteller hvor små gruppeforskjeller som kan avdekkes med statistisk sikkerhet. Med for få observasjoner (lav N) i en eller flere grupper er det fare for å forkaste mindre effekter selv om de er reelle, og man står i fare for at tilfeldige forskjeller feilaktig framstår som reelle. For sjeldne tilstander (og sår kan være én slik dersom det ikke blir et tydelig utbrudd) er dette relevant, noe som tilsier at en større andel av forsøksmerda bør merkes for å være sikker.

Figur 5 viser estimer for vekt per dag for merket og umerket gruppe prosjektet iFelt fra like før jul 2023 og ca. 5 uker framover. Det er «smale» konfidensintervall for umerket (blå) gruppe (som har

større antall registreringer), og noe videre intervall for merket gruppe som har ca. 11% av de daglige registreringene. Figuren viser at det er mulig å følge en dynamisk utvikling av vekt for behandlet gruppe og kontrollgruppe gjennom forsøket, og det samme gjelder også andre variabler som sår, lus mm. Antallet registreringer er stort nok for å kunne gjøre sikre statistiske analyser selv for spesifikke korte perioder.



Figur 5. Estimert vekt med 95% konfidensintervall for merket- og kontrollgruppe tidlig i forsøket for prosjekt iFelt. Registreringer er gjennomført med kamera med 1500 - 2000 målinger per dag, og med ca. 11% i merket gruppe.

Oppsummering

- Feltforsøk trengs for å kvantifisere effekt og varighet av vaksinasjon eller andre sykdomsforebyggende tiltak under reelle oppdrettsforhold.
- Forsøk med behandlet og ubehandlet fisk i samme merd gir samme betingelser, gruppene må være like og helst gjennomført med «blokkrandomisering».
- Innledende analyser viser at fjerning av fettfinnen godt lar seg gjenkjenne av programvare som benytter høyoppløselige bilder.
- Feilklassifisering av umerkede individer («falske positive») kan vri (gi bias) i datagrunnlaget for den merkede gruppen sterkt dersom andelen merket fisk er lav.
- Ønsker man å bruke mindre enn 20% merket fisk bør denne risikoen analyseres og systemet kontrolleres og optimaliseres før registreringene starter. Generelt mener vi at med 10% eller lavere andel merket fisk bør systemet gi tydelig mindre enn 1% «falske positive» tilordninger blant den umerkede fisken. 90% av fisken er da i umerket gruppe og dersom 1% av disse blir feilaktig tilordnet som merket vil det medføre at hele 8,5% av registreringene i merket gruppe blir falske positive.
- I løpet av et døgn bør man få et tilstrekkelig datagrunnlag for henholdsvis merket og umerket gruppe, slik at man i detalj kan følge utviklingen av sentrale variabler i hver gruppe over tid.

Litteraturhenvisning og vitenskapelige artikler om feltforsøk i norsk fiskeoppdrett:

Andrews M, Stormoen M, Schmidt-Posthaus H, Wahli T and Midtlyng PJ (2015). Rapid temperature-dependent wound closure following adipose fin clipping of Atlantic salmon *Salmo salar* L. *Journal of Fish Diseases* 38 (6), 523-531. DOI: 10.1111/jfd.12261.

Eksempler av studier med merkede fiskegrupper i samme merd:

Røsæg MV, Thorarinsson R and Aunsmo A (2021). Effect of vaccines against pancreas disease in farmed Atlantic salmon. *Journal of Fish Diseases* 2021:00;1-14. DOI:10.1111/jfd. 13505.

Aunsmo A, Guttvik A, Midtlyng PJ, Larssen RB, Evensen Ø and Skjerve E. (2008). Association of spinal deformity and vaccine induced abdominal lesions in harvest sized Atlantic salmon, *Salmo salar*. *Journal of Fish Diseases* 31, 515-524.

Aunsmo A, Larssen RB, Valle PS, Sandberg M, Evensen Ø, Midtlyng PJ, Østvik A and Skjerve E (2008). Improved field trial methodology for quantifying vaccination side-effects in farmed Atlantic salmon (*Salmo salar* L.). *Aquaculture* 284 (1-4), 19-24.

Midtlyng PJ and Lillehaug A (1998). Growth of Atlantic salmon *Salmo salar* after intraperitoneal administration of vaccines containing adjuvants. *Diseases of Aquatic Organisms* 32, 91-97.