

Vurdering af modeller for optælling af skuldere i so-besætninger

Erik Jørgensen
Det Jordbrugsvidenskabelige Fakultet
Aarhus Universitet

17. marts 2011

Resumé

Det Jordbrugsfaglige fakultet ved Aarhus Universitet er blevet anmodet af Fødevarestyrelsen om at udarbejde en vurdering af modeller for optælling af skuldere. Den ene (model 1) omfatter, at dyrlægen ved de månedlige besøg foretager stikprøvevis optælling af forekomsten af skuldre med *lette skuldere* blandt de diegivende søer. Dyrlægens optælling sammenholdes med landmandens egen, og hvis forskellen er for stor, foretager dyrlægen en totaloptælling, som erstatter landmandens egen. Den anden (model 2) omfatter alene totaloptællingen, som foretages 4 gange om året i forbindelse med de månedlige besøg, og hvor dyrlægen fastsætter tidspunkt for optælling tilfældigt, uden landmanden på forhånd ved, hvornår det foregår. Sammenligningen af de to modeller er foretaget ved at opsætte scenarier med 4 typer af landmænd udvalgt ud fra resultater i projektet, *En enkel og pålidelig skala for vurdering af skuldere på den levende so*, med udgangspunkt i en besætningsstørrelse på 300 årssøer. Beregningerne blev foretaget under en forudsætning om, at dyrlægens arbejdsindsats skulle være sammenlignelig ved de to modeller. Beregningerne viste, at den stikprøvevis optælling i model 1 ikke gav tilstrækkelig information til at kunne skelne mellem landmænd med samme præcision i vurderingerne, som dyrlægen, og landmænd med afvigende præcision. Den bedste skelnen kom ved at koncentrere indsatsen om totaloptællinger på tilfældige tidspunkter som i model 2. Muligheden for skelnen forbedres yderligere for model 2 ved kombination af observationer fra flere besøg og vurderes at give dyrlægen bedre muligheder for at planlægge sin arbejdstid. Desuden giver model 2 en dataindsamling, der om ønsket kan anvendes til validering af handlingsprogrammets effekt på skuldere udviklingen, og modellen sikrer fortsat fokus på besætningernes specifikke sundhedsproblemer.

Hvis forskellen mellem landmand og dyrlæge benyttes som udgangspunkt for iværksættelse af kalibrering af landmandens bedømmelserne bør omfanget vurderes nøjere. Beregningerne viste, at med den valgte besætningsstørrelse ville et kriterium med afvigelse på mere end 6% mellem de to optællinger medføre, at cirka hver femte besætning med tilfredsstillende registreringer årligt ville blive udpeget til efteruddannelsesaktiviteter.

1 Introduktion

DJF har fra fødevarestyrelsen modtaget følgende anmodning:

1.1 Spørgetema vedr. vurdering af modeller for optælling af skulderrsår i so-besætninger

I forbindelse med arbejdet med udarbejdelse af nye regler vedrørende skulderrsår (veterinærforligets projekt 4a), er der behov for en faglig vurdering af metoder for optælling af skulderrsår i sobesætninger. Foreløbig er der udarbejdet forslag til to konkrete modeller for, hvordan optællingen kan finde sted.

Det første forslag (*model 1*) er beskrevet i arbejdsgrupperapporten om skulderrsår hos søer (maj 2008). Forslaget indebærer, at landmanden månedligt fører egenkontrol med skulderrsår i sin besætning. Ved hvert rådgivningsbesøg udfører den praktiserende dyrlæge en stikprøvevis kontrol af søer med skulderrsår i farestalden, hvorved landmandens registreringer valideres. Er der uoverensstemmelse mellem landmandens optælling og dyrlægens stikprøve, skal dyrlægen undersøge alle søer i farestalden og vurdere, hvor mange der har lette skulderrsår.

Det andet forslag (*model 2*) blev præsenteret på mødet i projektets følgegruppe d. 1. november 2010. Dette forslag indebærer, at dyrlægen kvartalvist gennemgår alle søer i farestalden og registrerer søer med skulderrsår. Dyrlægens optælling af skulderrsår skal foregå uanmeldt. Landmanden skal fortsat optælle antallet af søer med skulderrsår forud for hvert rådgivningsbesøg.

Anmodning

Det Jordbrugsvidenskabelige Fakultet (DJF), Aarhus Universitet, bedes foretage en faglig vurdering af hvert af de to forslag med særlig fokus på relevante fordele og ulemper, navnlig for så vidt angår optællingernes validitet og optællingsmetodernes praktiske konsekvenser. I den forbindelse bedes DJF give en vurdering af hyppigheden af uoverensstemmelser mellem dyrlægens og landmandens optællinger, herunder forslag til håndtering heraf, samt en vurdering af, hvor ofte model 1 vil give anledning til, at dyrlægen skal optælle samtlige søer i farestalden.

1.2 Tolkning

Vi opfatter problemstillingen som en redegørelse for en optimal strategi for dyrlægens kontrol af landmandens optælling af antal skuldre med graden *let skulderrsår* i besætningens farestald i forbindelse med de besøg, som dyrlægen foretager minimum 9 gange årligt.

Vi opfatter model 1 som indeholdende følgende tre stadier ved hvert månedligt dyrlægebesøg.

1. Landmandens egne registreringer umiddelbart før dyrlægebesøg (24-48 timer) (Obligatorisk)
2. En revisionslignende stikprøvekontrol foretaget af dyrlægen (Obligatorisk). (Stikprøvens omfang er endnu ikke fastlagt).

3. Hvis der er diskrepans mellem stikprøvekontrol og egenkontrol gennemfører dyrlægen fuld optælling af søer i farestalden. Kriteriet for diskrepans er ikke defineret endnu.
4. Dyrlægens fulde optælling er efter hidtidig plan tænkt at erstatte landmandsregistrering.

Hvor ofte den fulde optælling vil ske i model 1 vil være bestemt af stikprøvestørrelse og diskrepanskriterie, og beslutning om fuld optælling baseres alene på stikprøvens udfald og kan derfor først foretages på selve besøgsdagen. I tilfælde af diskrepans overvejes det tillige at sikre opfriskning af skalaens anvendelse.

Den foreslåede model 2 er i virkeligheden indeholdt i model 1. Hvis informationen i stikprøven er meget lille i model 1 svarer det til, at man trækker lod om at gennemføre den fulde optælling, hvor optælling af søer snarere end terningkast afgør udfaldet. I model 2 er terningkastet blot flyttet hen før, dyrlægen besøger besætningen. Det er nævnt, at udgangspunktet for model 2 er 4 årlige optællinger. Model 2 kan imidlertid implementeres med andre optællingshyppigheder uden problemer.

Model 1 indeholder i virkeligheden også det fulde dyrlægebaserede kontrolprogram. Diskrepanskriteriet kan sættes så lavt, at dyrlægen skal gennemføre den fulde optælling stort set hver gang.

Landmandens egne registreringer er i princippet ikke afhængig af valg af kontrolmodel. Dog kan der forventes en reduktion i kvaliteten af registreringerne, hvis hans registreringer for hyppigt bliver erstattet af dyrlægens fulde optælling. Der er en risiko for, at landmanden vil vurdere, at hans eller hans personales indsats i kontrolprogrammet er overflødig.

2 Scenario

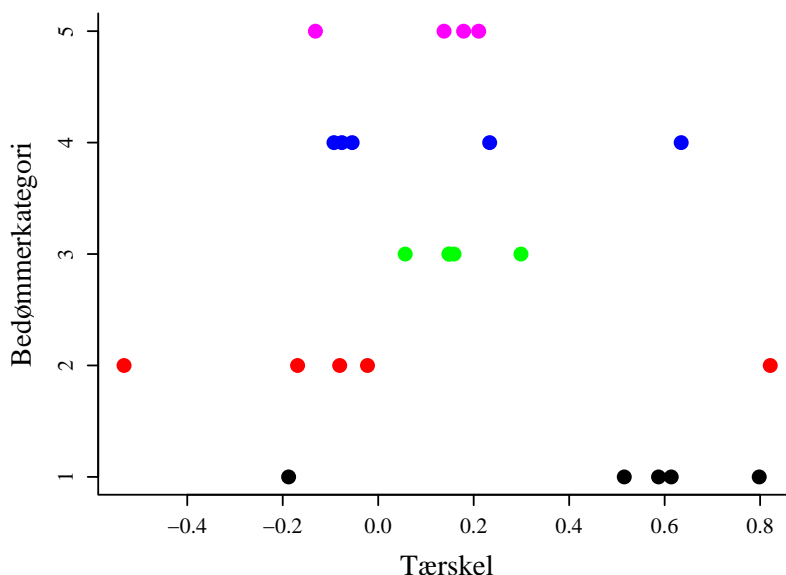
Vi har valgt at illustrere problemstillingen ud fra et scenarie for en besætning med 300 årssøer med ugedrift. Besætningen besøges af dyrlægen 9 gange om året i forbindelse med handlingsplan for skulderrsår. Umiddelbart før hvert besøg skal svineproducenten optælle forekomst af skuldre i besætningen med *lette skulderrsår*.

Det betyder, at der ved hvert besøg vil være cirka 53 diegivende søer i farestalden. Vi antager, at besætningen har en sand forekomst af skulderrsår i den høje ende, det vil sige 30 % sand prævalens.

I forbindelse med del 2 i projektet, *En enkel og pålidelig skala for vurdering af skulderrsår på den levende so*, blev der foretaget en opgørelse af forskellige bedømmers tærskelværdi for at bedømme et sår som et skulderrsår efter den nye skala (det vil sige med en diameter større end 2 cm). Vi vil tage udgangspunkt i resultater fra dette forsøg, da det giver et godt data-grundlag til at belyse problemstillingen.

Et væsentligt resultat var, at det så ud til, at forekomsten af *lette skulderrsår* på soens to skuldre var stort set uafhængige. Vi vil derfor i det efterfølgende beregne præcisionen ud fra antal bedømte skuldre. Et andet resultat var en kortlægning af variabiliteten i bedømmelserne mellem forskellige grupper af bedømmere. Opgørelsen byggede på en såkaldt tærskelmodel, hvor forskellige bedømmere havde forskellig tærskel (grænse) for, hvornår de bedømte et sår som *ingen eller ubetydeligt skulderrsår* eller *let skulderrsår*. I Figur 1 er disse tærskler vist for de 5 bedømmergrupper, der indgik i undersøgelsen. Analysen viste, at bedømmergruppe 3 (praktiserede dyrlæger) var mest ensartede i deres opfattelse af skalaen, mens landmandsgruppen (gruppe 1) varierede mere og havde en højere tærskel. Dette skyldtes, at landmandsgruppen sjældnere fejlagtigt vurderede en skulder til at have et *let skulderrsår* og oftere fejlagtigt vurderede

en skulder til at have *ingen eller ubetydeligt skuldersår* end dyrlægerne. Der var derimod ikke noget, der tydede på, at de forskellige bedømmergrupper havde en større usikkerhed ved gentagne bedømmelser. En af konklusioner fra projektet var, at der var behov for et uddannelsesforløb, så der blev større overensstemmelse mellem de forskellige bedømmergrupper, men i dette notat vil vi anvende værdierne til at illustrere effekten af dyrlægens kontrolprogram.



Figur 1: Tærskel for skuldersårsbedømmelse i del 2 af projektet, *En enkel og pålidelig skala for vurdering af skuldersår på den levende so*. Bedømmergruppe 1 er landmænd, 3 er praktiserende dyrlæger.

Ved beregninger i dette notat har vi valgt at lade dyrlægens vurdering svare til en middel dyrlæge i Figur 1. Dertil har vi defineret 4 landmænd. A) Den problemfrie landmand, som bedømmer skuldrene med samme præcision som dyrlægen. B) Landmanden, der registrerer for mange falsk positive skuldersår, det vil sige har den laveste tærskel. C) Landmanden, der registrerer for få sandt positive skuldersår, det vil sige har den højeste tærskel, og endeligt D) En landmand, som registrerer lige så godt som dyrlægen, men efterfølgende ændrer fem *lette skuldersår* bedømmelser til *ingen eller ubetydeligt skuldersår*. Bedømmertyperne er opsummeret i tabel 1.

Tabel 1: Bedømmertyper anvendt ved scenarierne, Se: sensitivitet, Sp: specificitet

Bedømmer	Tærskel	Se %	Sp %	Obs forekomst %	Kommentar
Dyrlæge	0.08	90	95	30	
Landmand A	0.08	90	95	30	Samme registrering som dyrlæge
Landmand B	-0.26	96	89	36	Mest sensitive
Landmand C	0.73	67	99	20	Mindst sensitive
Landmand D	0.08	90	95	-	Bevidst undervurdering

Et succesfuldt kontrolprogram vil kunne identificere landmand B, C, og D som problemer, men vil acceptere registreringerne fra landmand A.

3 De forskellige modeller

Hvis model 2 anvendes vil det være let for dyrlægen på forhånd at beregne arbejdsindsatsen. Så længe landmanden ikke på forhånd ved hvornår total optællinger sker, kan dyrlægen vælge de 4 årlige optællingsbesøg, så de passer bedst muligt ind i programmet. Tillige vil arbejdsindsatsen være den samme i alle besætninger uafhængig af diskrepans på besætningsniveau.

Arbejdsindsatsen til model 1 vil være vanskeligere at anslå. Der vil være et fast tidsforbrug til den stikprøvevise vurdering og derefter en risiko for at skulle lave en fuld optælling af søerne i farestalden afhængig af udfaldet af stikprøvetællingen. Hvis model 2 med de fire fulde optællinger skal svare i arbejdsomfang til model 1's stikprøveoptællingerne på de enkelte besøgsdage er den maksimale stikprøvestørrelse $4/9 \approx 44\%$ af søerne.

Men dertil kommer så, hvor sandsynligt det er, at der skal foretages fuld optælling efterfølgende. Igen vurderes det at være et krav, at optællingsomfanget i en problemløs besætning ikke må være større end ved model 2. Hvis der eksempelvis er en sandsynlighed for totaloptælling på 10 % hos en landmand som landmand A, vil der maksimalt kunne optælles 34 % af skuldrene ved de 9 årlige besøg.

Landmandens arbejdsindsats er i princippet den samme i de to modeller.

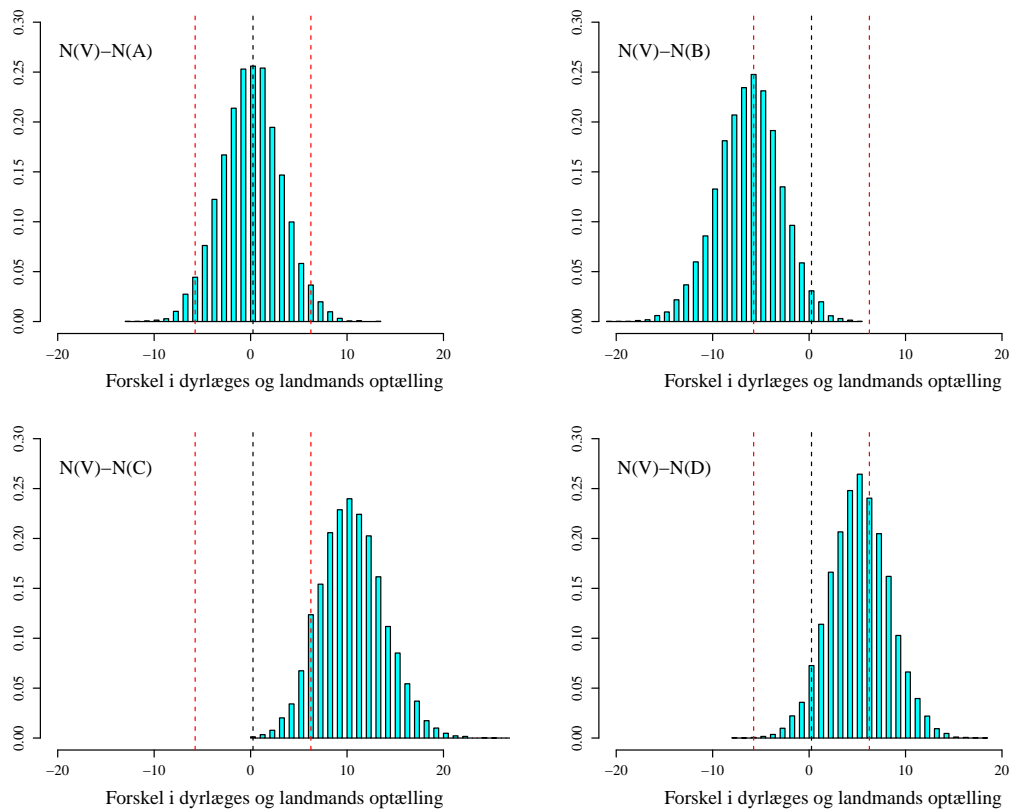
Totaloptællingen indgår som en del af begge modeller, derfor starter vi med model 2, som kun indeholder denne.

3.1 Model 2: Stikprøvevis fuld optælling

Da den diagnostiske klassificering i *ingen eller ubetydeligt skuldersår* eller *let skuldersår* ikke er perfekt, vil dyrlægens og landmandens registreringer ikke være sammenfaldende. I Figur 2 er vist, hvordan udfaldet kan forventes at variere i de 4 scenarier. I figurens øverste venstre hjørne er vist fordeling mellem de to identiske bedømmere. Som det ses, varierer antallet mellem $\pm 10\%$ af de 106 skuldre. Som det fremgår, er fordelingen af udfaldene forskellige i de 4 scenarier, men det kan være vanskeligt at finde en grænse, så landmand A's optælling ikke også opfattes som en fejl.

Et diskrepans kriterium kunne være, at afvigelsen skulle være større end ± 6 , hvilket vil svare til, at der var en risiko på omkring 5 % for fejlklassificering af landmand A som et problem. Et andet kunne være ± 7 , der vil medføre en risiko på omkring 2.5 % for fejlklassificering. Da forskellene er heltallige kan disse værdier ikke defineres eksakt. Da der er 4 besøg pr. år, vil landmand A have en årlig risiko på henholdsvis omkring 20 % og omkring 10 % ved disse diskrepanskriterier.

Hvor godt problembesætningerne kan identificeres ved denne strategi er vist i tabel 2. Specielt fejlsce-
narie B og C identificeres med høj succesrate (45 % og 89 %), mens det er vanskeligere at konstatere den systematiske fejl på 5 sår per optælling (Scenarie D – 30 %). I tabellen er endvidere vist udfaldet ved kombination af to besøg til at identificere fejlbesætningerne. Som det ses, medfører denne mulighed en væsentlig bedre adskillelse af problembesætninger fra normale besætninger.



Figur 2: Forskel mellem dyrlæges og landmands optælling ved de 4 scenarier. $N(V)$ er antallet optalt af dyrlæge, $N(A, \dots, D)$ er antallet optalt af de 4 landmænd i scenarierne beskrevet i Tabel 1. Lodrette linjer angiver 0 og \pm grænser i diskrepanskræterium

Tabel 2: Model 2: Sandsynlighed (%) for at konstatere fejlregistrering ved kontroloptælling. Fuldstændig optælling 106 skuldre. Landmandstype er beskrevet i Tabel 1

Besøgs- hyppighed	\pm - Grænse for indgreb	Landmandstype			
		A	B	C	D
Et besøg	6	4	45	89	30
Et besøg	7	2	33	80	21
To besøg kombineret	8	5	80	99	62
To besøg kombineret	9	3	73	100	53

3.2 Model 1: Stikprøvevis optælling evt. med total optælling

Ved den stikprøvevis optælling i forbindelse med de 9 årlige besætningsbesøg skal der som nævnt regnes med, at kun en mindre del af de diegivende søer skal vurderes. Som nævnt vil en optælling hver gang af 44 % af søerne svare til den indsats, der skal til for at optælle samtlige søer 4 gange årligt. Hvis der udover dette skal være 'plads' til den efterfølgende totaloptælling, skal stikprøven yderligere begrænses, afhængigt af hvor hyppigt totaloptællingen foretages.

I Tabel 3 er vist beregnede værdier for forskellig stikprøvestørrelse. Eksemplerne omfatter stikprøvestørrelser på 20, 40 og 80.

Tabel 3: Model 1: Sandsynlighed (%) for at konstatere fejlregistrering ved kontroloptælling. Stikprøvevis optælling. Landmandstype er beskrevet i Tabel 1

Antal Skuldre	± %- Grænse for indgreb	Landmandstype			
		A	B	C	D
20	25	1	4	6	2
20	22	2	6	8	3
20	12	22	32	41	26
40	16	1	6	14	3
40	15	2	8	21	6
80	10	1	14	47	3
80	9	2	21	56	13

Grænserne for indgreb er valgt så landmand A har henholdsvis 1 og 2 procents risiko for at blive udvalgt til totaloptælling. Det giver samme årlige fejlrate på henholdsvis 10 % og 20 % som i model 2. Stikprøvestørrelsen på 80 medfører markant højere dyrlæge arbejdsindsats end model 2, mens stikprøvestørrelsen på 40 cirka svarer til model 2. En stikprøvestørrelse på 20 giver mere plads til totaloptællinger. Derfor er for stikprøvestørrelsen på 20 % valgt et snævrere diskrepanskriterie, som giver hyppigere totaloptællinger, men stadig færre end de 4 gange om året i model 2. Som det ses, er der kun små forskelle i sandsynligheden for at blive udvalgt til totaloptælling for problembesætningerne B, C og D og for ikke-problembesætningen A. Specielt er det vanskeligt at skelne A og D. De 5 skuldre, der er ændret fra *let skuldarsår* til *ingen eller ubetydeligt skuldarsår* skal jo indgå i dyrlægens stikprøve, før fejlregistreringen kan identificeres.

Det er klart, at beslutningen om at gennemføre kalibreringsaktiviteter ikke kan baseres på den stikprøvevis optælling. Her er man nødt til at anvende den opfølgende totaloptælling, hvis man skal begrænse uddannelsesindsatsen.

3.3 Stratificeret stikprøvetagning

En mulighed for at øge sikkerheden ved den stikprøvevis vurdering i model 1 er at udvælge stikprøven tilfældigt indenfor grupper (strata) i stedet for tilfældigt blandt samtlige søer i diegivningsstalden. I dette tilfælde vil den naturlige gruppering være at vælge to grupper bestående af skuldre med landmandens vurdering som *ingen eller ubetydeligt skuldarsår* og *let skuldarsår*. Eller snarere en gruppe af søer med

mindst en skulder med gradueringen *let skuldersår*, og en gruppe af søer uden denne graduering. Denne opdeling på soniveau vil forringe effekten af stratificeringen, men er nødvendig af effektivitetshensyn. En vurdering af stratificeringsmetoden er ikke medtaget i dette notat, primært fordi det vil være vanskeligt at implementere i praksis. Efter optællingen ville resultatet fra de to strata skulle sammenvejes ud fra størrelsen af de to strata, og den optimale stratificering ville afhænge af forventet prævalens af skuldersår og præcision i bedømmelsen. Dette ville efter vores opfattelse kræve udvikling af et programværktøj til håndtering af beregningerne. Selvom stratificering gennemføres forventes stadig, at model 2 vil være at foretrække.

3.4 Vurdering af enkelttyrsklassificeringer

Et argument for den stikprøvevise vurdering i model 1 har været, at dyrlæge og landmænd vil have mulighed for at diskutere forkerte gradtildelinger. Dette vil naturligvis også kunne finde sted i forbindelse med model 2, selv om det ikke er formaliseret. Som det fremgår af nedenstående er denne fejlretning imidlertid ikke uproblematisk, især ikke i forbindelse med model 1.

Som udgangspunkt bør vurderingen af landmandens evner til at bedømme ske på total opgørelsen. Da hverken Se eller Sp er lig 100 %, vil der forekomme fejlklassificeringer i både landmandens og dyrlægens bedømmelser. Den nye skala har ganske gode egenskaber med hensyn til sensitivitet og specificitet. Men der vil dog stadigvæk være omkring 5 % af skuldrene med *ingen eller ubetydeligt skuldersår*, som er fejlklassificeret, fordi de i virkeligheden burde bedømmes som *let skuldersår*. I gruppen, der bedømmes som *let skuldersår* af landmand A, vil der være omkring 10 % af skuldrene, der ikke burde indgå. Når dyrlægen checker registreringer, vil han jo også kunne tage fejl, men de to grupper med landmandens bedømmelse som *let skuldersår* henholdsvis *ingen eller ubetydeligt skuldersår* har markant anderledes sand prævalens, end når alle skuldre vurderes samlet. Det giver lidt overraskende udfald af fejlanalysen. Da det er en fejlretning, vil der tillige udelukkende være fokuseret på de tilfælde, hvor de to bedømmere ikke er enige.

Hvis de to bedømmere er lige gode, som landmand A og dyrlægen i vores scenarie, vil der blandt de skuldre, hvor de er uenige, være lige så mange, hvor landmanden har ret, som hvor dyrlægen har ret. I tilfældet landmand C med den høje specificitet vil landmanden have uret med hensyn til *ingen eller ubetydeligt skuldersår* i omkring 70 % af tilfældene, men han vil have ret i 76 % af tilfældene med hensyn til *let skuldersår*. Dette skyldes, at de bedømmelser, der er uenighed om, ligger på vippen.

Hvis dyrlægen gerne vil benytte besætningsgennemgangen til at diskutere brugen af skalaen, er det derfor vigtigt, at dyrlægen ikke i sin gennemgang tager udgangspunkt i grænsetilfælde, men derimod koncentrerer sig om de grelle fejlregistreringer. Afstemningen af grænsetilfælde bør ske ud fra det undervisningsmateriale med fotomateriale, som er under udarbejdelse.

4 Konklusion

Beregningerne blev foretaget under en forudsætning om, at dyrlægens arbejdsindsats skulle være sammenlignelig ved de to modeller. Med denne forudsætning viste beregningerne, at den stikprøvevise optælling i model 1 ikke gav tilstrækkelig information til at kunne skelne mellem landmænd med samme præcision i vurderingerne som dyrlægen, og landmænd med afvigende præcision. På grund af den lille

stikprøvestørrelse var det ikke muligt at skelne mellem problemløse besætninger og problembesætninger. Udfaldet af stikprøvetællingen fungerede nærmest som en generator af tilfældige tal med henblik på et tilfældigt valg af tidspunkt for totaloptælling. Dette tilfældige valg kan ligeså godt foretages af dyrlægen inden besætningsbesøget. Den bedste skelnen kom derfor ved at koncentrere indsatsen om totaloptællinger på tilfældige tidspunkter, som i model 2. Muligheden for skelnen mellem problembesætninger og problemløse besætninger kan yderligere forbedres for model 2 ved kombination af observationer fra flere besøg, og vurderes at give dyrlægen bedre muligheder for at planlægge sin arbejdstid. Desuden giver model 2 en dataindsamling, der om ønsket kan anvendes til validering af handlingsprogrammets effekt på skulderyrsudviklingen. Endelig vil model 2 indebære mindre risiko end model 1 for, at handlingsplanen for skulderyrs som en bivirkning reducerer fokus på andre specifikke sundhedsproblemer i besætningerne, da ikke alle besøg har fokus på skulderyrs.

Vi forstår, at dyrlægens kontrol af landmandens optællinger skal benyttes som grundlag for at vurdere, om der er behov for kalibrering af landmandens bedømmelser. Hvis dette er tilfældet, bør omfanget nøjere vurderes. Beregningerne i dette notat viste, at med den valgte besætningsstørrelse vil et kriterium med afvigelse på mere end 6% mellem de to optællinger, medføre at cirka hver femte besætning med tilfredsstillende registreringer årligt bliver udpeget til kalibrering. Problembesætninger vil have en langt større sandsynlighed for at blive udpeget. Såfremt model 1 vælges, er det vigtigt at behovet for efteruddannelsesaktiviteter besluttet på baggrund af totaloptællingerne og ikke på baggrund af stikprøvetællingerne.

Vi forventer, at dyrlægen fortsat vil gennemføre en uformel fejlretning af landmandens registreringer, selv hvis model 2 gennemføres, for eksempel i forbindelse med, om der findes søer i farestalden, som burde være i en sygesti. Det vil være udmærket, at diskutere registreringer, der efter dyrlægens opfattelse er oplagt forkerte som del i en erfaringsudveksling. Derimod bør grænsetilfælde ikke anvendes i erfaringsudvekslingen.