

Wageningen UR Livestock Research

Partner in livestock innovations



Rapport 536

Vereenvoudiging van het Welfare Quality® protocol voor vleeskalveren

Juli 2012

Colofon

Uitgever

Wageningen UR Livestock Research
Postbus 65, 8200 AB Lelystad
Telefoon 0320 - 238238
Fax 0320 - 238050
E-mail info.livestockresearch@wur.nl
Internet <http://www.livestockresearch.wur.nl>

Redactie

Communication Services

Copyright

© Wageningen UR Livestock Research, onderdeel
van Stichting Dienst Landbouwkundig Onderzoek,
2012

Overname van de inhoud is toegestaan,
mits met duidelijke bronvermelding.

Aansprakelijkheid

Wageningen UR Livestock Research aanvaardt
geen aansprakelijkheid voor eventuele schade
voortvloeiend uit het gebruik van de resultaten van
dit onderzoek of de toepassing van de adviezen.

Wageningen UR Livestock Research en Central
Veterinary Institute, beiden onderdeel van Stichting
Dienst Landbouwkundig Onderzoek vormen samen
met het Departement Dierwetenschappen van
Wageningen University de Animal Sciences Group
van Wageningen UR (University & Research
centre).

Losse nummers zijn te verkrijgen via de website.



De certificering volgens ISO 9001 door DNV
onderstreept ons kwaliteitsniveau. Op al onze
onderzoeksopdrachten zijn de Algemene
Voorwaarden van de Animal Sciences Group
van toepassing. Deze zijn gedeponeerd bij de
Arrondissementsrechtbank Zwolle.

Abstract

This report describes research into possibilities
to simplify the Welfare Quality® assessment
protocol for veal calves

Keywords

Veal calves, welfare assessment protocol

Referaat

ISSN 1570 - 8616

Auteur(s)

Kees van Reenen
Pieter Vereijken
Willem Buist
Jac Thissen
Bas Engel

Titel

Vereenvoudiging van het Welfare Quality®
protocol voor vleeskalveren

Rapport 536

Samenvatting

In dit rapport wordt de resultaten beschreven
van onderzoek naar mogelijkheden om het
Welfare Quality® protocol voor het beoordelen
van welzijn van vleeskalveren te
vereenvoudigen

Trefwoorden

Vleeskalveren, welzijnsprotocol

Rapport 536

Vereenvoudiging van het Welfare Quality® protocol voor vleeskalveren

Kees van Reenen
Pieter Vereijken
Willem Buist
Jac Thissen
Bas Engel

Juli 2012



Ministerie van Economische Zaken,
Landbouw en Innovatie

Dit onderzoek is uitgevoerd in opdracht van het Ministerie van Economische Zaken, Landbouw en Innovatie.

Samenvatting

In opdracht van het Ministerie van EL&I is onderzocht in hoeverre de meest uitgebreide versie van het Welfare Quality® (WQ) protocol voor de beoordeling van het welzijn van vleeskalveren kan worden vereenvoudigd.

Voor het onderzoek is gebruik gemaakt van een data-set met waarnemingen volgens het volledige WQ protocol op 224 vleeskalverbedrijven. Waarnemingen van elk bedrijf zijn volgens de in WQ ontwikkelde systematiek kwantitatief samengevat in scores (tussen 0 en 100) voor 12 criteria voor welzijn en 4 principes voor welzijn. Op basis van de scores voor de vier principes kreeg elk bedrijf een eindkwalificatie voor het welzijnsniveau, in termen van één van de volgende vier klassen: Excellent, Enhanced, Acceptable, en Not classified. Er zijn in totaal vier vereenvoudigingen onderzocht: (i) het bekorten van de waarnemingstijd voor gedrag, (ii) het voorspellen van het % kalveren dat abnormaal oraal gedrag laat zien uit de pensontwikkeling, (iii) het voorspellen van het % longen met matige tot ernstige tekenen van longontsteking uit klinische verschijnselen van respiratoire problemen waargenomen op het vleeskalverbedrijf op 13 weken, en (iv) het verkleinen van de steekproef voor klinisch onderzoek. Uitkomsten van het volledige protocol fungeerden als referentie (of gouden standaard). Uitkomsten van de vereenvoudigde versies van het protocol werden met deze gouden standaard vergeleken. Deze vergelijking vond plaats op drie niveau's: op het niveau van de eindkwalificatie, op het niveau van de principes, en op het niveau van de criteria. De kwaliteit van de vereenvoudigingen werd op basis van de vergelijking met de gouden standaard uitgedrukt in een aantal kwaliteitscriteria: (1) het % overeenstemming, (2) de sensitiviteit, (3) de specificiteit, en (4) de correlatie (alleen van toepassing op scores voor principes en criteria). Voor het voorspellen van de ene variabele uit de andere is gebruik gemaakt van logistische regressie-analyses.

Van de onderzochte vereenvoudigingen voldeden twee van de vier onderzochte strategieën, zij het niet voor alle criteria en principes. Reductie van de waarnemingstijd (van 90 naar 30 minuten per bedrijf) beïnvloedde de eindkwalificatie en scores van de 4 principes nauwelijks, had geen groot effect op scores voor criterium 10 (hoofdzakelijk bepaald door abnormaal oraal gedrag), maar werkte niet goed uit op scores van criterium 9 (bepaald door sociaal gedrag). Deze vereenvoudiging kan daardoor alleen worden gebruikt wanneer het accent wordt gelegd op abnormaal oraal gedrag, en minder aandacht wordt gegeven aan sociaal gedrag. Het schatten van abnormaal oraal gedrag uit de pensontwikkeling leek mogelijk op alle niveau's van het WQ integratiemodel (eindkwalificatie, scores van principe 4 en scores van criterium 10). De overige gedragskenmerken (zoals sociaal gedrag) hebben geen zinvolle relatie met de pensontwikkeling. Dit betekent dat deze vereenvoudiging alleen tot significante tijdwinst kan leiden wanneer zou worden afgezien van alle overige gedragswaarnemingen, en wanneer principe 4 alleen bepaald zou worden door criterium 10.

Summary

At the request of the Dutch Ministry of Economic Affairs, Agriculture and Innovation, a study was performed with the objective to examine possibilities to simplify the Welfare Quality® (WQ) assessment protocol for veal calves.

For this study a data-set with observations according to the complete WQ protocol on 224 farms was used. Observations on each farm were summarized into scores (between 0 and 100) for 12 criteria and 4 principles for animal welfare, according to the methodology that was developed within the WQ project. Based on the scores of the 4 principles, the overall level of welfare on each farm was classified according to one of four welfare classes: Excellent, Enhanced, Acceptable, and Not classified. Four strategies for simplifying the WQ protocol were examined: (i) reducing the time taken for behavioural observations, (ii) predicting the level of abnormal oral behaviour exhibited by the calves from the level of rumen development (observed post mortem), (iii) predicting the % of lungs with severe lesions indicative of pneumonia (observed post mortem) from prevalences of clinical signs of respiratory disease observed in live animals on-farm, and (iv) reducing the sample size (number of calves) for the clinical examination on-farm. The outcome of the complete protocol was used as the reference or gold standard. The outcomes of simplified versions of the protocol were compared with this gold standard. This comparison was done at three levels: at the level of the overall qualification of the level of welfare, at the level of principles, and at the level of criteria. Based on the comparisons with the gold standard, the quality of each simplification strategy was expressed in terms of a number of criteria: (1) the % of agreement, (2) the sensitivity, (3) the specificity, and (4) the correlation (the latter criterion only applies to scores for principles and criteria). Logistical regression analysis was used to predict one measure from the other.

Two of four simplification strategies that were examined gave acceptable results, although not for all criteria and principles. Reduction of the total time for behavioural observations (from 90 to 30 minutes per farm), hardly affected the overall qualification of the level of welfare or scores of principle 4, had a small effect on scores of criterion 10 (primarily determined by abnormal oral behaviour), but unfavourably affected scores of criterion 9 (determined by social behaviour). Therefore, this simplification strategy can only be used when the main focus would be on abnormal oral behaviour rather than social behaviour. Predicting abnormal oral behaviour from the level of rumen development seemed to be possible at all levels of comparison (overall qualification of the level of welfare, scores of principle 4, and scores of criteria). However, there were no meaningful associations between rumen development and any of the other behavioural measures (such as social behaviour). This means that this strategy for simplifying the WQ protocol for veal calves would only result in a significant reduction of observation time when principle 4 would be determined by criterion 10 only.

Inhoudsopgave

Samenvatting

Summary

1 Inleiding	10
2 Materiaal en Methoden	11
2.1 Data.....	11
2.2 Protocol	11
2.3 WQ integratiemodel	14
2.4 Vereenvoudigingsstrategieën	14
2.4.1 Uitgangspunt.....	14
2.4.2 Berekeningen.....	14
2.4.3 Beoordeling kwaliteit van de vereenvoudiging	15
2.4.4 De onderzochte vereenvoudigingen	17
3 Resultaten	19
3.1 De uitkomsten van het WQ integratiemodel	19
3.2 Resultaten verkorten waarnemingstijd gedrag	19
3.3 Het voorspellen van abnormaal oraal gedrag uit de pensontwikkeling	20
3.4 Het voorspellen van longpathologie uit klinische verschijnselen	21
3.5 Halveren van de steekproef voor klinisch onderzoek	21
4 Conclusies	22
Bijlagen.....	23

1 Inleiding

Vanaf 2006 wordt door Wageningen UR Livestock Research, in samenwerking met de verzamelde vleeskalversector in Nederland gewerkt aan de ontwikkeling van een welzijnsmonitor voor vleeskalveren. Het onderzoeksproject wordt gefinancierd door het Ministerie van EL&I en de Productschappen Diervoeder (PDV) en Vee Vlees en Eieren (PVE). Naast deze Nederlandse partijen zijn ook Franse en Italiaanse partijen bij het onderzoek betrokken. De facto was met ingang van 2007 sprake van één gezamenlijk Europees project waarin onderzoekers uit Nederland (WUR), Frankrijk (ISA-Lille) en Italië (Universiteit van Padova) nauw samenwerken met vertegenwoordigers van Nederlandse, Franse en Italiaanse vleeskalversector. Het streven is om te komen tot een Europese welzijnsmonitor voor vleeskalveren waarmee op een wetenschappelijk verantwoorde en praktisch haalbare manier de welzijnsstatus van vleeskalveren kan worden bepaald op vleeskalverbedrijven in de praktijk.

In overeenstemming met de benadering volgens Welfare Quality® (WQ) is een welzijnsmonitor vleeskalveren ontwikkeld waarbij een breed scala aan (dier)kenmerken wordt gemeten op elk individueel bedrijf. Het waarnemen van veel van deze kenmerken is arbeidsintensief waardoor het toepassen van het volledige protocol relatief veel tijd in beslag neemt. Dit kan een beperkende factor zijn bij de implementatie van de welzijnsmonitor in de praktijk.

Daarom is in opdracht van het Ministerie van EL&I een onderzoeksproject uitgevoerd waarin is gezocht naar mogelijkheden om de meest uitgebreide versie van de welzijnsmonitor vleeskalveren te vereenvoudigen. De resultaten van dit onderzoek staan in dit rapport.

2 Materiaal en Methoden

2.1 Data

Uitgangspunt voor het vereenvoudigsonderzoek was een data-set met waarnemingen aan vleeskalveren op in totaal 224 vleeskalverbedrijven verdeeld over Nederland, Frankrijk, en Italië. Onder de bezochte bedrijven bevonden zich zowel blankvlees- als rosébedrijven. Een onderverdeling van aantallen bedrijven naar land en type productie (blank versus rosé) staat in Tabel 1.

Tabel 1. Aantallen vleeskalverbedrijven waarop in het kader van het monitoringsonderzoek waarnemingen zijn gedaan, uitgesplitst naar land en type kalfsvleesproductie

Land	Type kalfsvleesproductie	
	Blank	Rosé
Nederland	100	50
Frankrijk	50	0
Italië	24	0

2.2 Protocol

Op alle bedrijven in Tabel 1 is hetzelfde welzijnsprotocol toegepast, bestaande uit een drietal deelprotocollen: (i) een gedragsprotocol, (ii) een protocol voor klinische gezondheid, en (iii) een protocol voor het scoren van pathologische afwijkingen aan organen op het slachthuis. Het welzijnsprotocol voor vleeskalveren is ontwikkeld volgens de systematiek van WQ. Volgens deze systematiek wordt het begrip dierenwelzijn betrekking op in totaal 12 zogenaamde Criteria, die op hun beurt kunnen worden samengevat in vier zogenaamde Principes. Deze Criteria en Principes vormen de basis voor de keuze van (dier)kenmerken, en daarmee voor de definitie van welzijnsprotocollen. Een overzicht van de 12 Criteria en 4 bijbehorende Principes staat in Tabel 3.

Tabel 3. Principes en Criteria die volgens de WQ benadering de basis vormen voor het begrip dierenwelzijn

Welfare Quality® Principes	Welfare Quality® Criteria
Goede voeding	1 Afwezigheid langdurige honger 2 Afwezigheid langdurige dorst
Goede huisvesting	3 Ligcomfort 4 Thermaal comfort (temperatuur) 5 Bewegingsvrijheid
Goede gezondheid	6 Afwezigheid van letsel 7 Afwezigheid van ziekte 8 Afwezigheid van pijn veroorzaakt door ingrepen
Normaal gedrag	9 Normaal sociaal gedrag 10 Normaal ander gedrag 11 Goede mens-dier relatie 12 Positieve emotionele toestand

Bij aanvang van het project Welzijnsmonitor Vleeskalveren is een groot aantal kenmerken gedefinieerd en opgenomen in deelprotocollen voor gedrag, klinische gezondheid en pathologie. In de loop van het project is een aantal oorspronkelijke variabelen komen te vervallen, en is van andere variabelen vastgesteld dat de manier waarop ze op dit moment zijn gedefinieerd anders zou moeten. Ook zijn in de loop van het project variabelen aan de deelprotocollen toegevoegd die daar in eerste instantie nog geen deel van uitmaakten.

Een beschrijving van de (dier)kenmerken die deel uitmaken van de laatste versie van het welzijnsprotocol voor vleeskalveren staat in Tabel 2. In Tabel 2 wordt ook aangegeven onder welk WQ Criterium elk kenmerk valt.

Tabel 2. Omschrijving van kenmerken die deel uitmaken van welzijnsprotocollen voor vleeskalveren, uitgesplitst naar de WQ Criteria

Welfare Quality® Criteria	Kenmerk	Toelichting
1 Afwezigheid langdurige honger	Achterblijvers	% kalveren met een lichaamsconditie meer dan 30% minder dan koppelgenoten
2 Afwezigheid langdurige dorst	Watervoorziening (hoeveelheid en timing); aantal en reinheid van drinkpunten	Beslisboom
3 Ligcomfort	Bursae	% kalveren met bursae ("dikke knieën")
4 Thermaal comfort	Reinheid van kalveren	% bevuilde kalveren ¹
5 Bewegingsvrijheid	Natte kalveren	% natte kalveren
6 Afwezigheid van letsel	Gladheid van de vloer ²	% kreukele kalveren
	Kreupelheid	% kalveren met wonden aan staart en/of oren
7 Afwezigheid van ziekte	Verwondingen	% kalveren dat hoest
	Hoesten	% kalveren met abnormale ademhaling
	Abnormale ademhaling	% kalveren met neusuitvloeiing
	Neusuitvloeiing	% aangetaste longen (verkleuringen)
	Longontsteking	% longen met pleuritis
	Pleuritis	Dunne mest op de vloer
	Dunne mest ³	% lebmagen met lesies (erosies, ulcera)
	Lesies in de lebmaag	% penzen met plaques (samengeklonterde pensvilli)
	Plaques in de pens	% dorre – doffe kalveren (geen glanzend haarkleed)
	Dorre kalveren	% overig duidelijk zieke kalveren
	Zieke kalveren	% kalveren met een Hemoglobinegehalte op 13 weken < 4.5 mmol/l
	Haemoglobinegehalte	% gestorven / geëuthanaseerde kalveren
	Uitval	
8 Afwezigheid van pijn veroorzaakt door ingrepen	Staart couperen	Niet relevant voor vleeskalveren
9 Normaal sociaal gedrag	Sociaal gedrag	Gemiddeld % kalveren dat sociaal likken laat zien
10 Normaal ander gedrag	Abnormaal oraal gedrag	Gemiddeld % kalveren dat abnormaal oraal gedrag laat zien
	Urinedrinken/bezuigen	% kalveren met gele bekken en/of nat/gezwollen preputium
	preputium soortgenoot	% kalveren dat kan worden aangeraakt tijdens de calf escape test
11 Goede mens-dier relatie	Angst voor mens	Qualitative Behaviour Assessment
12 Positieve emotionele toestand	QBA	

¹Het % bevuilde kalveren is nog niet gescoord. Inmiddels is er, als uitvloeisel van het onderzoek naar alternatieve vloeren, een protocol ontwikkeld voor het objectief scoren van de mate van bevuiling van vleeskalveren

²Voor het scoren van de gladheid – en andere belangrijke kwaliteiten – van de vloer is nog geen gevalideerd protocol beschikbaar.

³Dunne mest kan alleen afgelezen worden aan de mest op de vloer. Tot nu toe is dunne mest gescoord op hokniveau. De mate waarin dunne mest op een bedrijf voorkomt wordt dan uitgedrukt als het % hokken met dunne mest (i.c. het aantal hokken met dunne mest gedeeld door het totaal aantal gescoorde hokken, keer 100%). Wanneer de mate waarin dunne mest voorkomt op deze manier wordt uitgedrukt zijn bedrijven alleen onderling vergelijkbaar wanneer er sprake is van (ongeveer) gelijke groepsgroottes. Een alternatieve manier van scoren van dunne mest is nodig.

⁴QBA (Qualitative Behaviour Assessment) is tot op heden nog niet toegepast in het project Welzijnsmonitor Vleeskalveren, met uitzondering van de bedrijven in Italië.

In Tabel 4 staan dezelfde kenmerken vermeld als die in Tabel 2, maar nu uitgesplitst naar het deelprotocol waarvan ze deel uitmaken (d.w.z. het gedragsprotocol, het protocol klinisch gezondheid en het protocol voor pathologie). De waarnemingen behorend tot hetzelfde deelprotocol werden uitgevoerd door dezelfde waarnemer tijdens hetzelfde bezoek (niet op dezelfde dag). In het onderzoek tot nu toe traden als waarnemers op: (i) een dierenarts (voor de uitvoering van het klinisch protocol), (ii) een onderzoeker van Wageningen UR (voor het gedragsonderzoek), en (iii) een assistent veterinaire patholoog van het CVI (voor waarnemingen rond pathologie op het slachthuis). In de meest rechtse kolom van Tabel 4 staan de werkzaamheden voor elk deelprotocol kort omschreven, en staat tevens een indicatie van de benodigde tijd per bedrijf (koppel).

Tabel 4. Kenmerken uit de Welzijnsmonitor Vleeskalveren, uitgesplitst naar het deelprotocol waar ze deel van uitmaken

Deelprotocol	Kenmerken	Omschrijving en tijd (uren)
Klinische gezondheid	Achterblijvers Bursae Natte kalveren Kreupelheid Verwondingen Hoesten Hoesten Abnormale ademhaling Neusuitvloeiing Dunne mest Dorre kalveren Zieke kalveren Urinedrinken/bezuigen preputium soortgenoot	Beoordeling door dierenarts van steekproef van 300 kalveren Elk kalf wordt op alle kenmerken beoordeeld Nodig: 3 – 6 uur per bedrijf
Gedrag	Sociaal gedrag Abnormaal oraal gedrag Angst voor de mens QBA	Gedragswaarnemingen (36 hokken) en uitvoering van gedragstest (100 kalveren) Nodig: 8 uur
Pathologie	Longontsteking Pleuritis Lesies in de lebmaag Plaques in de pens	Beoordeling van organen op het slachthuis: 100 longen, 40 lebmagen, 40 penzen Nodig: 2 man x 4 = 8 uur
Overig ¹	Watervoorziening; aantal en reinheid van drinkpunten Gladheid van de vloer Reinheid van kalveren Haemoglobinegehalte Uitval	Onderdeel van 8 uur t.b.v. de realisatie van het gedragsprotocol

¹Overige waarnemingen maken geen deel uit van een specifiek deelprotocol. Gegeven met betrekking tot haemoglobinegehalte en uitval worden verstrekt door de kalverhouder c.q. de integratie. Waarnemingen aan reinheid van drinkpunten, reinheid van de kalveren en gladheid van de vloer zouden, bijvoorbeeld, gekoppeld kunnen worden aan het gedragsprotocol.

2.3 WQ integratiemodel

Eind 2009 is de eindrapportage opgeleverd van het Europese project WQ. Een onderdeel daarvan waren welzijnsprotocollen voor melkkoeien, vleesstieren en vleeskalveren (Welfare Quality®, Assessment Protocol for cattle – 7. Applied to veal calves). In het WQ Assessment Protocol voor vleeskalveren wordt ook een kwantitatieve methode beschreven om de waargenomen (dier)kenmerken voor ieder individueel bedrijf samen te vatten in scores (tussen 0 en 100) van de 12 Criteria, en om vervolgens de scores van de 12 Criteria samen te vatten tot scores (eveneens tussen 0 en 100) van de 4 Principles. De scores van de 4 Principles, tenslotte, vormen samen een profiel aan de hand waarvan, op grond van bepaalde beslissingsregels, een bedrijf een overall eindkwalificatie voor het welzijnsniveau krijgt toebedeeld in termen van één van de volgende vier klassen: (i) "Excellent" (zeer goed), (ii) "Enhanced" (verbeterd), (iii) "Acceptable" (aanvaardbaar), en "Not classified" (onvoldoende). De onderliggende rekenmodellen wordt gebruik gemaakt van wegingsfactoren en specifieke coëfficiënten die zijn verkregen op basis van een zogenaamde "multicriteria evaluatie" uitgevoerd door experts die in gecontroleerde sessies worden geconfronteerd met virtuele data-sets. In de uiteindelijke klassificatie van het welzijnsniveau op een bedrijf volgens de Welfare Quality® systematiek zit dus expliciet de weging en onderlinge waardering van (dier)kenmerken respectievelijk scores op de 12 Criteria besloten.

De inhoud van het in WQ opgeleverde Assessment Protocol voor vleeskalveren kwam op een aantal belangrijke punten niet overeen met de laatste versie van het protocol voor vleeskalveren zoals dat in het project Welzijnsmonitor Vleeskalveren (samenvverking tussen onderzoekers en sector uit Nederland, Frankrijk en Italië) was vastgesteld. Daarom voldeed ook het in WQ ontwikkelde integratiemodel voor vleeskalveren niet meer bij aanvang van het vereenvoudigsonderzoek. Daarom is, als eerste stap in het onderzoek naar mogelijkheden om het WQ protocol voor vleeskalveren te vereenvoudigen, het integratiemodel voor een groot deel vernieuwd, en helemaal in overeenstemming gebracht met het protocol zoals dat in het Nederlands-Frans-Italiaanse onderzoek is ontwikkeld. Daarbij is gebruik gemaakt van nieuwe informatie uit consultaties van experts, die al in december 2010 hadden plaatsgevonden.

Het herziene integratiemodel staat beschreven in Annex 1 bij dit rapport, in hetzelfde format als de in WQ opgeleverde deliverable. Het is de bedoeling om dit document uiteindelijk als formele herziening van het integratiemodel voor vleeskalveren in de richting van het WQ Network te gebruiken.

2.4 Vereenvoudigingsstrategieën

2.4.1 Uitgangspunt

De zoektocht naar mogelijkheden voor vereenvoudiging van het WQ protocol was gebaseerd op de veronderstelling dat er betekenisvolle verbanden of correlaties tussen dierkenmerken zouden bestaan, bijvoorbeeld tussen enerzijds een dierkenmerk dat alleen met relatief veel tijd geregistreerd kan worden en, anderzijds, een dierkenmerk dat zonder veel moeite kan worden waargenomen. In dit geval zou het mogelijk zijn om met behulp van statistische technieken de waarde van het moeilijk te verkrijgen kenmerk te voorspellen uit de waarde van het gemakkelijk waar te nemen kenmerk. In een vereenvoudigde versie van het WQ protocol zou dan alleen het gemakkelijk waarneembare kenmerk daadwerkelijk worden gemeten en het moeilijk waar te nemen kenmerk zou uit het gemakkelijk waarneembare kenmerk via een rekenregel worden afgeleid.

Op deze manier zou het "bouwwerk" dat het WQ protocol feitelijk is, in tact blijven, en zouden ook op basis van een vereenvoudigd protocol scores voor de 12 Criteria en de vier Principles kunnen worden afgeleid. In een statistische analyse zouden verschillende varianten van meer of minder versimpelde versies van het WQ protocol kunnen worden bekeken, en als het ware "gecalibreerd" ten opzicht van de "gouden standaard" of referentie, i.c. de uitkomst van het WQ protocol gebaseerd op de volledige set oorspronkelijke waarnemingen.

2.4.2 Berekeningen

Als eerste stap is daarom gezocht naar verbanden en correlaties tussen dierkenmerken, bij voorkeur uit verschillende deelprotocollen (i.c. het gedragsprotocol, het klinisch gezondheidsprotocol en het protocol voor pathologie), omdat daarmee potentieel de meeste tijdwinst geboekt zou kunnen worden. Wanneer kenmerken gecorreleerd waren, dan is met behulp van (logistische) regressie het ene kenmerk uit het andere voorspeld. Vervolgens zijn zowel de oorspronkelijke als de voorspelde gegevens als invoer gebruikt voor het WQ integratiemodel (Annex 1). De uitkomsten van het

integratiemodel op basis van de oorspronkelijke kenmerken (referentie) werd vergeleken met de uitkomsten van het integratiemodel op basis van deels oorspronkelijke en deels voorspelde kenmerken (i.c., een vereenvoudigde versie van het protocol). Omdat het WQ integratiemodel op drie niveau's uitkomsten geeft (namelijk scores voor Criteria, scores voor Principles, en een eindkwalificatie) is de kwaliteit van de vereenvoudiging (d.w.z. de mate waarin een vereenvoudiging voldoet) ook op deze drie niveau's beoordeeld.

Naast het gebruik van voorspellers, is er nog een andere potentiële manier waarop het WQ protocol vereenvoudigd zou kunnen worden. Er zouden, bijvoorbeeld, minder dieren, hokken of organen onderzocht kunnen worden tijdens de beoordeling van de klinische gezondheid, het gedrag of afwijkingen aan organen op het slachthuis. In statistische termen staat een dergelijke vereenvoudigingsstrategie gelijk aan het verkleinen van de steekproef.

Sociaal en abnormaal oraal gedrag van vleeskalveren werd waargenomen tijdens drie periodes van elk 30 minuten, verspreid over de dag. Verkleining van de steekproef voor het gedragsonderzoek werd bewerkstelligd door de gebruikte waarnemingsperiode te verkleinen.

Om te bestuderen of verkleining van de steekproef aanvaardbaar zou zijn is met behulp van simulatiestudies de steekproef voor het klinische onderzoek op elk bedrijf gehalveerd. In totaal werd voor elk bedrijf 100 keer een a-selectie trekking gedaan zodanig dat in elk van de 100 simulaties de helft van de werkelijk bestudeerde steekproef werd verkregen.

2.4.3 Beoordeling kwaliteit van de vereenvoudiging

Voor het beoordelen van de kwaliteit van een vereenvoudiging zijn, afhankelijk van het niveau waarop de beoordeling plaatsvond (op het niveau van Criteria, Principles of eindkwalificatie), drie of vier kwaliteitscriteria gebruikt. De vier kwaliteitscriteria waren:

- (a) Het % overeenstemming
- (b) De sensitiviteit
- (c) De specificiteit
- (d) De Spearman rangcorrelatie

De betekenis van de eerste drie kwaliteitscriteria (% overeenstemming, sensitiviteit en specificiteit) laat zich het gemakkelijkst uitleggen aan de hand van een zogenaamde 2x2 tabel (zie Tabel 5):

Tabel 5. Voorbeeld van een 2x2 tabel voor de berekening van kwaliteitscriteria voor de beoordeling van een vereenvoudiging van het WQ protocol. Zie tekst voor uitleg.

		Vereenvoudiging		
		Positief	Negatief	
Referentie	Positief	A	B	A + B
	Negatief	C	D	C + D
		TOTAAL		N

Stel dat een welzijnsmonitor wordt gebruikt die onderscheid maakt tussen "goede" en "slechte" bedrijven. Voor het voorbeeld noemen we de "slechte" bedrijven "positief", en de "goede" bedrijven "negatief" (naar analogie met, bijvoorbeeld, een bloedtest die een bepaalde besmetting opspoort; een positieve test betekent dat de besmetting aanwezig is). Stel dat we deze welzijnsmonitor op een grote steekproef van N bedrijven toepassen. We maken daarbij gebruik van twee welzijnsmonitoren: een uitgebreide variant – de referentie, en een vereenvoudigde variant – de vereenvoudiging. Vervolgens kunnen we op grond van de uitkomsten van zowel de uitgebreide als de vereenvoudigde variant de bedrijven indelen in bovenstaande 2x2 tabel (zie Tabel 5). In elke "cel" van de tabel staat een aantal bedrijven. A zijn de bedrijven die zowel volgens de referentie als volgens de vereenvoudiging als "positief" werden gezien. B is het aantal bedrijven dat positief werd gezien door de referentie, maar negatief door de vereenvoudiging, etc. De uitgebreide monitor is ons ijkpunt (of de "werkelijkheid"), en we willen dus een uitspraak doen over de mate waarin de vereenvoudiging overeenkomt met de referentie. Daarvoor kunnen we gebruik maken van de hierboven genoemde kwaliteitscriteria.

- Het % overeenstemming in bovenstaand voorbeeld betreft de bedrijven waarop zowel de referentie als de vereenvoudiging dezelfde uitslag gaven. Als formule:

$$A + D / N \times 100\%$$

- Het aantal “werkelijk positieven” (dus volgens de referentie) is A+B. De sensitiviteit van de vereenvoudigde monitor heeft betrekking op het vermogen om werkelijk positieven op te sporen. Als formule: sensitiviteit = $A/A+B \times 100\%$.
- $100 - \text{de sensitiviteit} = \% \text{ vals negatief}$ (of $B/A+B \times 100\%$), dus bedrijven die ten onrechte als negatief werden beoordeeld door de vereenvoudigde monitor
- Het aantal “werkelijk negatieven” (dus volgens de referentie) is C+D. De specificiteit van de vereenvoudigde monitor heeft betrekking op het vermogen om werkelijk negatieven op te sporen. Als formule: specificiteit = $D/C+D \times 100\%$.
- $100 - \text{de specificiteit} = \% \text{ vals positief}$ (of $C/C+D \times 100\%$), dus bedrijven die ten onrechte als negatief werden beoordeeld door de vereenvoudigde monitor

Wanneer de vereenvoudigde monitor precies even goed zou zijn als de referentie, dan zouden alle bedrijven zich in de kwadranten A en D van Tabel 5 bevinden, en daarmee zouden het % overeenstemming, de sensitiviteit en de specificiteit allemaal 100% bedragen.

Om in staat te zijn om de bovenstaande drie kwaliteitscriteria ook te gebruiken voor het vergelijken van de uitkomsten van het WQ integratiemodel is het dus noodzakelijk dat deze uitkomsten worden uitgedrukt in twee klassen (vergelijk “positief” versus “negatief”), zodat bedrijven kunnen worden ondergebracht in een 2x2 tabel zoals weergegeven in Tabel 5.

In het vereenvoudigingsonderzoek zijn de volgende onderverdelingen in twee klassen gebruikt:

- (i) Voor wat betreft de eindkwalificatie (Excellent, Enhanced, Acceptable of Not classified) zijn de twee beste en de twee slechtse klassen samengenomen, dus Excellent + Enhanced versus Acceptable + Not classified.
- (ii) Voor wat betreft scores voor Principles of Criteria is een onderverdeling gemaakt tussen bedrijven met een score groter of kleiner dan een bepaalde grenswaarde. In totaal zijn drie grenswaarde in dit verband bekeken: 20, 55 en 80. Bij de grenswaarde van 20 is dus een onderscheid gemaakt tussen bedrijven met een score groter of kleiner dan 20; bij de grenswaarde van 55 is een onderscheid gemaakt tussen bedrijven met een score groter of kleiner dan 55, etc. De grenzen van 20, 55 en 80 zijn niet willekeurig gekozen, maar corresponderen met de grenzen die volgens de WQ benadering de overgangen markeren, op het niveau van individuele Criteria of Principes, tussen Not classified (< 20), Acceptable (tussen 20 en 55), Enhanced (tussen 55 en 80), en Excellent (> 80).

Ad (i): een voorbeeld van een 2x2 tabel zoals in Tabel 5 op basis van een onderverdeling van bedrijven op grond van de eindkwalificatie volgens het WQ protocol ziet er als volgt uit (NC = Not classified, A = acceptable, Enh = Enhanced en Exc = Excellent):

Vereenvoudiging			
Referentie	NC + A	Enh + Exc	
NC + A	A	B	A + B
Enh + Exc	C	D	C + D
		TOTAAL	N

Ad (ii): een voorbeeld van een 2x2 tabel zoals in Tabel 5 op basis van een onderverdeling van bedrijven op grond van de score van een Criterium of een Principe ziet er als volgt uit:

		Vereenvoudiging		
Referentie		Score < 20	Score ≥ 20	
Score < 20	A	B	A + B	
	C	D	C + D	
		TOTAAL	N	

Precies dezelfde tabellen kunnen gemaakt worden wanneer in plaats van 20 een grens van 55 of een grens van 80 wordt gehanteerd.

In het vereenvoudigsonderzoek bij vleeskalveren was N gelijk aan het totaal aantal bedrijven. Dit aantal was iets kleiner dan 224, omdat niet van alle bedrijven een volledige set waarnemingen beschikbaar was. De referentie had betrekking op het volledige WQ protocol, dus alle oorspronkelijke waarnemingen. De vereenvoudiging had betrekking op de uitkomsten van het WQ integratiemodel met als invoer deels voorspelde waarden, of uitkomsten op basis van een gereduceerde steekproef. Voor de beoordeling van de kwaliteit van een vereenvoudiging op het niveau van Criteria en Principes was nog een vierde criterium beschikbaar, namelijk Spearman rangcorrelatie. Deze correlatie werd berekend tussen de scores die verkregen werden volgens de referentie, en de scores die verkregen worden na vereenvoudiging. Deze correlatie geeft aan in hoeverre de rangschikking van bedrijven op grond van hun score voor een individueel Criterium of Principe (van laag naar hoog), verandert wanneer een vereenvoudigde monitor is gebruikt voor het berekenen van deze score. Hoe hoger de rangcorrelatie, hoe meer de rangschikking na vereenvoudiging lijkt op die volgens de referentie.

Bij elke waarde voor het % gelijk, de sensitiviteit en de specificiteit werd ook een schatting van het 90% betrouwbaarheidsinterval (90% confidence interval) bepaald (zie resultaten). Een groot betrouwbaarheidsinterval betekent dat er aanvullend onderzoek nodig is om de geschatte sensitiviteit, specificiteit of % gelijk uit dit onderzoek te bevestigen. Grote betrouwbaarheidsintervallen kunnen het gevolg zijn van gebrek aan voldoende bedrijven in één van de twee klassen die worden gebruikt om een 2x2 tabel op te zetten (zie hierboven).

2.4.4 De onderzochte vereenvoudigingen

Er zijn in totaal vier vereenvoudigingen onderzocht:

- (a) Het bekorten van de waarnemingsperiode voor gedrag. Zoals hierboven aangegeven werden gedragswaarnemingen gedaan tijdens drie tijdsintervallen van 30 minuten elk (ochtend, middag, namiddag). Vervolgens werden deze drie intervallen bij elkaar opgeteld, en werden de gedragskenmerken (gemiddelde % kalveren) berekend over de volle 90 minuten (de referentie). De vereenvoudiging bestond eruit dat gedragskenmerken werden berekend over één van de drie intervallen van 30 minuten. Daarmee bestond deze vereenvoudiging feitelijk uit drie sub-vereenvoudigingen: (i) alleen 30 minuten ochtend gebruiken, (ii) alleen 30 minuten middag gebruiken, en (iii) alleen 30 minuten namiddag gebruiken
- (b) Het voorspellen van het gemiddelde % kalveren dat abnormaal oraal gedrag laat zien uit de pensontwikkeling. Hoewel in het WQ protocol voor vleeskalveren als pensparameters alleen het % penzen met plaque is opgenomen, werd als extra variabele (omdat elke pens toch werd geopend en nauwkeurig bekeken) ook een score toegekend aan de mate van ontwikkeling (op een discrete schaal van 0 tot 4). Daarbij wordt gekeken naar de lengte en dichtheid van de villi. Als een maat voor het niveau van pensontwikkeling in een koppel kalveren is het % penzen met score 3 (goed) of score 4 (volledig ontwikkeld). Dit % was significant gecorreleerd met het gemiddelde kalveren dat abnormaal oraal gedrag liet zien. Het gemiddelde % kalveren dat abnormaal oraal gedrag liet zien werd voorspeld uit het % penzen met score 3 of 4 met behulp van logistische regressie (zie Annex 2 voor details).
- (c) Het voorspellen van het % longen met matige tot ernstige tekenen van longontsteking (% longen met score 3 of 4; longen werden gescoord op een discrete schaal van 0 = gaaf tot 4 = ernstige tekenen van longontsteking) uit klinische verschijnselen van respiratoire problemen waargenomen op het vleeskalverbedrijf op 13 weken (% hoestende kalveren, % kalveren met neusuitvloeiing, % kalveren met abnormale ademhaling). Ook aan deze vereenvoudiging lagen significante correlaties ten grondslag. Het % aangetaste longen werd voorspeld uit

percentages kalveren met klinische verschijnselen van respiratoire problemen met behulp van dubbele logistische regressie (zie Annex 2).

- (d) Het halveren van de steekproef in het klinisch onderzoek door middel van simulatie. Ook in deze simulatiestudie werd gebruik gemaakt van de hierboven genoemde kwaliteitscriteria, maar dan op een iets andere manier dan bij de andere drie vereenvoudigingen. De procedure was als volgt:

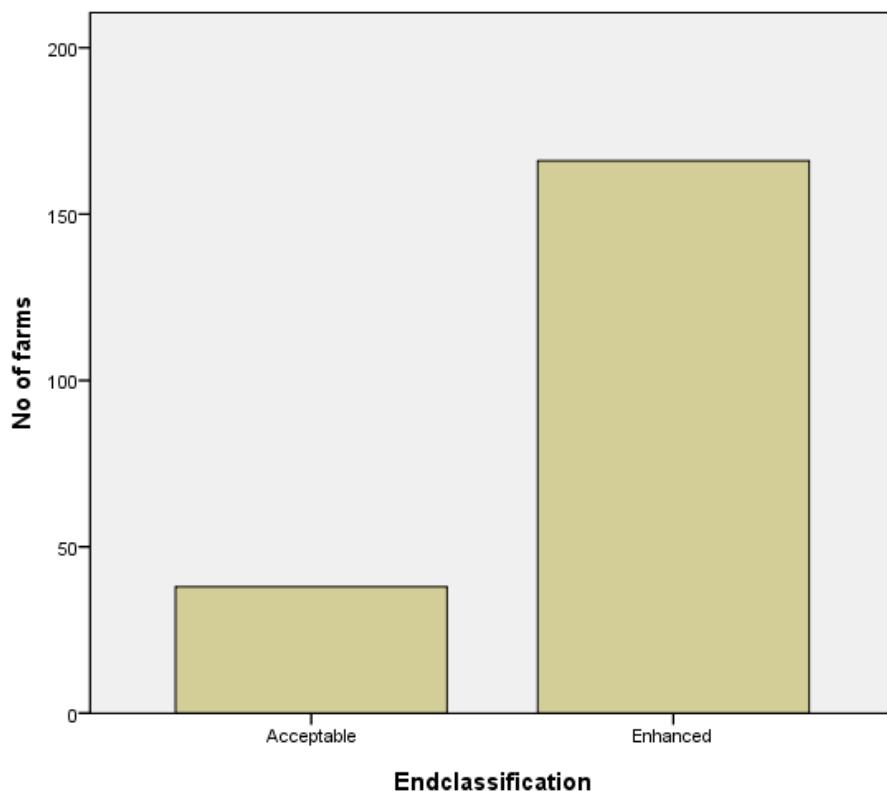
- 1) Er werd 100 x een aselecte trekking gedaan van de helft van de steekproef aan dieren
- 2) Voor elk van die 100 trekkingen werden de criteria bepaald die de kwaliteit van de vereenvoudiging weergegeven: % overeenkomst, sensitiviteit, en specificiteit
- 3) Van elk van deze kwaliteits-parameters werden de 100 gesimuleerde uitkomsten op volgorde gezet, van laag naar hoog
- 4) Als resultaat werden uit deze 100 waarden de 5e, de 50e, en de 95e waarde gerapporteerd (5, 50 en 95 percentiel). Naarmate deze drie waarden dichter bij elkaar, en dichter bij de 100% liggen, is halvering van de steekproef meer acceptabel.

Het effect van het halveren van de steekproef is alleen bestudeerd op het niveau van de eindkwalificatie en het niveau van Principes.

3 Resultaten

3.1 De uitkomsten van het WQ integratiemodel

In Figuur 1 wordt de verdeling weergegeven van de bedrijven over de respectievelijke klassen van de eindkwalificatie, na toepassing van het WQ integratiemodel.



Figuur 1. Verdeling van vleeskalverbedrijven over de klassen van de eindkwalificatie voor welzijn volgens het WQ integratiemodel.

Wat opvalt is dat de overgrote meerderheid van de bedrijven in dezelfde klasse valt, en dat dit de klasse Enhanced is. Hierbij moet worden opgemerkt dat in dit geval relatief soepele beslissingsregels zijn gehanteerd voor het omzetten van scores van de vier Principes naar eindkwalificatie (conform de WQ deliverable), en dat bovendien Criterium 8 (één van de drie Criteria die samen Principe 3, goed gezondheid, vormen) bij alle bedrijven de waarde 100 had (omdat in de vleeskalverhouderij in Europa staart couperen nooit wordt toegepast).

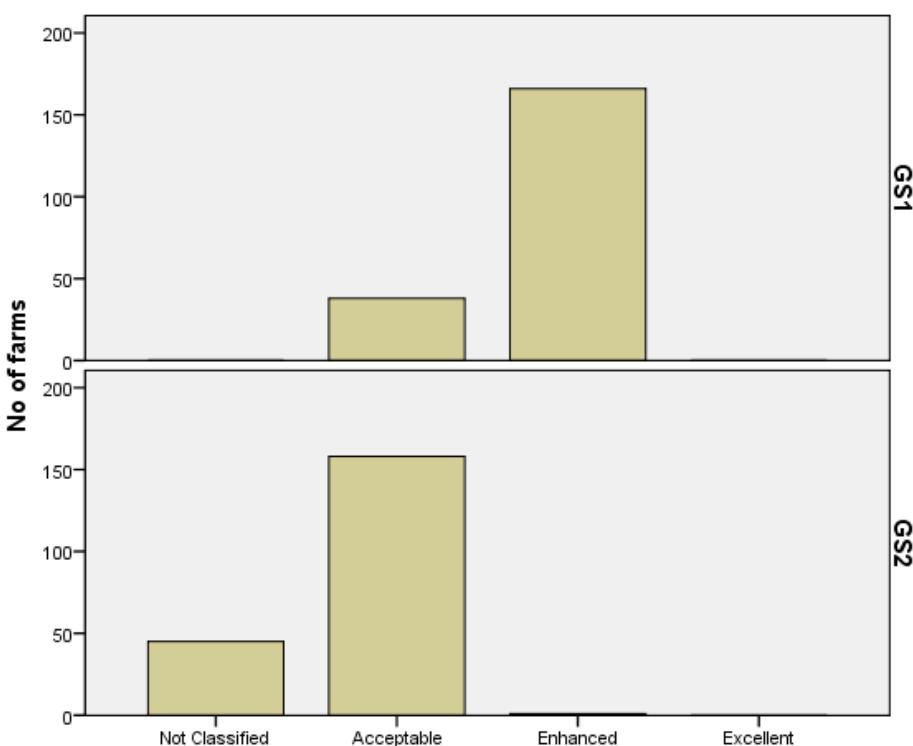
Het gebruik van de meest strenge beslissingsregels bij het bepalen van de eindkwalificatie (die aanvankelijk ook in WQ zouden worden toegepast) leidt ertoe dat de overgrote meerderheid van de vleeskalverbedrijven nu niet in de klasse Enhanced maar in de klasse Acceptable zit (zie Figuur 2, onderste plaatje met op de vertikale as 'GS2'). Door het hanteren van meer of minder stringente beslissingsregels (op basis van dezelfde waarnemingen) kan de verdeling van bedrijven over de klassen van de eindkwalificatie voor welzijn volgens het WQ integratiemodel in een bepaalde richting worden beïnvloed.

In tegenstelling tot de verdeling van bedrijven over de klassen voor de eindkwalificatie, lieten de verdelingen van bedrijven over klassen voor scores van Principles en Criteria wèl een behoorlijke variatie zien (zie Annex 3).

3.2 Resultaten verkorten waarnemingstijd gedrag

Het verkorten van de waarnemingstijd door maar één van de drie intervallen van 30 minuten te gebruiken voor het berekenen van de gedragsvariabelen geeft op het niveau van de eindkwalificatie een hoog % overeenkomst, een hoge sensitiviteit en een hoge specificiteit (> 90%). Dit geldt voor alle

drie de intervallen van 30 minuten (ochtend, middag, namiddag). In Annex 2 staan de uitkomsten in detail vermeld.



Figuur 2. Verdeling van vleeskalverbedrijven over de klassen van de eindkwalificatie voor welzijn volgens het WQ integratiemodel, wanneer verschillende beslissingsregels – variërend van soepel tot stringent – worden gebruikt voor het omzetten van scores voor de vier Principes naar de eindkwalificatie. De bovenste situatie ('GS1') is gelijk aan die in Figuur 1, en komt overeen met de in WQ gehanteerde aanpak. In het onderste plaatje ('GS2') is de laagste score van de vier scores voor Principes doorslaggevend voor de eindkwalificatie.

Ook op het niveau van Principe 4 (normaal gedrag), het Principe waarop deze variabelen van invloed zijn (zie Tabel 1 en Tabel 2), werkt de vereenvoudiging goed uit: % gelijk, sensitiviteit en specificiteit zijn hoog, en er zijn hoge rangcorrelaties tussen scores volgens het oorspronkelijke protocol en scores na vereenvoudiging (groter dan 0.90, zie Annex 2).

Op het niveau van Criteria (Criterion 9 en Criterion 10, zie Tabel 2) geven de vereenvoudigingen wisselende resultaten. Op Criterion 10, dat mede wordt bepaald door abnormaal oraal gedrag, werkt de vereenvoudiging het gunstigst uit: er is sprake van hoge % gelijk, een hoge sensitiviteit en een hoge specificiteit, en er zijn hoge rangcorrelaties tussen scores volgens het oorspronkelijke protocol en scores na vereenvoudiging (groter dan 0.90, zie Annex 2). Op Criterion 9, waarop sociaal gedrag van invloed is (zie Tabel 2) heeft de vereenvoudiging een minder gunstige invloed: relatief lage rangcorrelaties, en in een aantal gevallen een lage specificiteit (zie Annex 2).

3.3 Het voorspellen van abnormaal oraal gedrag uit de pensontwikkeling

Het voorspellen van abnormaal oraal gedrag uit de pensontwikkeling heeft een effect op de eindkwalificatie, op scores voor Principe 4 en op scores voor Criterion 10 (zie Tabel 1 en Tabel 2). Op alle niveau's (eindkwalificatie, scores voor Principes, en scores voor Criteria) leidt deze vereenvoudiging tot voldoende overeenstemming met de referentie: hoog % gelijk, hoge sensitiviteit en specificiteit, hoge rangcorrelaties tussen scores voor Principe 4 onder de referentie en na vereenvoudiging, en eveneens hoge rangcorrelaties tussen scores voor Criterion 10 onder de referentie en na vereenvoudiging (zie Annex 2 voor de details). Hierbij moet worden opgemerkt dat in een aantal gevallen grote betrouwbaarheidsintervallen zijn geschat voor de sensitiviteit en de specificiteit. Dit betekent dat aanvullend onderzoek noodzakelijk is.

3.4 Het voorspellen van longpathologie uit klinische verschijnselen

Het voorspellen van longpathologie uit klinische verschijnselen heeft een effect op de eindkwalificatie, scores voor Principe 3, en scores voor Criterium 7 (zie Tabel 1 en Tabel 2).

Op het niveau van de eindkwalificatie en scores voor Principe 3 is er voldoende overeenstemming tussen de referentie en deze vereenvoudiging, hoewel in een aantal gevallen grote betrouwbaarheidsintervallen zijn geschat (zie Annex 2 voor details). Op het niveau van scores voor Criterium 7 geeft het voorspellen van longpathologie uit klinische verschijnselen aanleiding tot een te lage specificiteit (rond de 50%).

3.5 Halveren van de steekproef voor klinisch onderzoek

Halveren van de steekproef had een effect op de eindkwalificatie, en op scores van alle vier de Principes (zie Tabel 1, Tabel 2 en Tabel 3).

Op het niveau van de eindkwalificatie leidde halvering van de steekproef tot een matig % overeenstemming (ca. 90%), en een matige specificiteit (< 80%), zie Annex 2 voor details.

Na vereenvoudiging schoten de sensitiviteit van Principes 2 en 4 (< 90%), en de specificiteit van Principe 3 tekort (zie Annex 2). Dat blijkt niet alleen uit hoge 50 percentielpunten, maar ook uit een grote verschillen tussen het 5 en 95 percentielpunt (zie Annex 2).

4 Conclusies

Gelet op de resultaten kunnen de volgende conclusies worden getrokken:

- (a) Algemeen gesteld, lijkt het erop dat hoe hoger het aggregatieniveau in het WQ integratiemodel (via Criteria en Principes naar het hoogste integratieniveau, de eindkwalificatie), hoe ongevoeliger het systeem lijkt te zijn voor vereenvoudigingen. Tegelijkertijd neemt ook de variatie sterk af naarmate gegevens op een hoger niveau zijn geaggregeerd: de variatie tussen bedrijven op het niveau van individuele dierkenmerken is aanzienlijk, die op het niveau van individuele Criteria en Principes behoorlijk, terwijl variatie op het niveau van de eindkwalificatie zeer gering is.
- (b) Reductie van de waarnemingstijd – van 90 minuten naar een interval van 30 minuten lijkt voor het schatten van het niveau aan abnormaal oraal gedrag goed mogelijk. Voor het schatten van sociaal gedrag werkt deze vereenvoudiging minder goed uit. Dit suggereert dat het uitvoeren van sociaal gedrag meer afhankelijk is van het moment op de dag. Door maar één van de drie intervallen van 30 minuten te gebruiken zou daardoor de schatting van het niveau van sociaal gedrag ten opzichte van de referentie onnauwkeuriger worden.
- (c) Schatten van abnormaal oraal gedrag uit de pensontwikkeling lijkt mogelijk, op alle niveau's van het WQ integratiemodel (eindkwalificatie, Principe 4 en Criterium 10). Hoge 90% betrouwbaarheidsintervallen geven tegelijkertijd aanleiding tot enige voorzichtigheid. De bruikbaarheid van deze vereenvoudiging behoeft nog een bevestiging met nieuwe data.
- (d) Het schatten van longpathologie uit klinische waarnemingen lijkt mogelijk op het niveau van de eindkwalificatie en Principes (Principe 3), maar niet op het niveau van criteria (Criterium 7)
- (e) Halvering van de steekproef voor klinisch onderzoek lijkt riskant, met name door een matige sensitiviteit op het niveau van Principes principles. Voor criteria is deze simulatie niet uitgevoerd, maar gelet op de consistente bevinding dat het WQ integratiemodel gevoeliger wordt voor vereenvoudigingen naarmate het aggregatieniveau afneemt, is de situatie voor Criteria na halvering van de steekproef naar verwachting niet anders (en waarschijnlijk ongunstiger).

Alles bij elkaar suggeren de resultaten van het vereenvoudigingsonderzoek dat er een beperkt aantal mogelijkheden bestaat om te komen tot "data gedreven" vereenvoudiging van het WQ protocol, in het bijzonder: (i) beperking van gedragsobservaties, en (ii) het voorspellen van abnormaal oraal gedrag uit pensontwikkeling

Ad (ii): dit leidt alleen tot significante tijdwinst wanneer vervolgens ook zou worden afgezien van alle overige gedragswaarnemingen, en wanneer Principe 4 (normaal gedrag) alleen bepaald zou worden door Criterium 10 (normaal ander gedrag).

ANNEX 1



Reports on the construction of welfare criteria for different livestock species

Part 3 – criteria construction for veal calves on farm

INRA, Institut de l'Elevage, LIP6
Department of Animal Science, University of Padova
Wageningen UR Livestock Research
Wageningen UR Biometris

Authors (Partners of WP2.3 and WP2.4):
M. Brscic, R. Botreau, C.G. van Reenen, J. Thissen and I. Veissier

November 29th 2011

Deliverable 2.8c, subtask 2.3.1.2

EU Food-CT-2004-506508



Contents

Executive summary	27
1 Criterion 1: Absence of prolonged hunger.....	29
1.1 Definition of the criterion	29
1.2 Measures to be used to assess the criterion	29
1.3 Analysis of the constraints	29
1.4 Type of mathematical construction proposed	29
1.5 Parameters to be defined.....	30
1.6 Adjustments of the mathematical construction	30
1.7 Calculation	33
2 Criterion 2: Absence of prolonged thirst.....	34
2.1 Definition of the criterion	34
2.2 Measures to be used to assess the criterion	34
2.3 Analysis of the constraints	35
2.4 Type of mathematical construction proposed	35
2.5 Parameters to be defined.....	36
2.6 Adjustments of the mathematical construction	36
2.7 Calculation	37
3 Criterion 3: Comfort around resting.....	38
3.1 Definition of the criterion	38
3.2 Measures to be used to assess the criterion	38
3.3 Analysis of the constraints	38
3.4 Type of mathematical construction proposed	39
3.5 Parameters to be defined.....	39
3.6 Adjustments of the mathematical construction	39
3.7 Calculation	46
4 Criterion 4: Thermal comfort	48
4.1 Definition of the criterion	48
4.2 Measures to be used to assess the criterion	48
4.3 Analysis of the constraints	48
4.4 Type of mathematical construction proposed	48
4.5 Parameters to be defined.....	48
4.6 Adjustments of the mathematical construction	49
4.7 Calculation	51

5 Criterion 5: Ease of movement.....	52
5.1 Definition of the criterion	52
5.2 Measures to be used to assess the criterion	52
5.3 Analysis of the constraints	53
5.4 Type of mathematical construction proposed	53
5.5 Parameters to be defined.....	53
5.6 Adjustments of the mathematical construction	53
5.7 Calculation	54
6 Criterion 6: Absence of injuries	55
6.1 Definition of the criterion	55
6.2 Measures to be used to assess the criterion	55
6.3 Analysis of the constraints	55
6.4 Type of mathematical construction proposed	56
6.5 Parameters to be defined.....	56
6.6 Adjustments of the mathematical construction	57
6.7 Calculation	65
7 Criterion 7: Absence of diseases	66
7.1 Definition of the criterion	66
7.2 Measures to be used to assess the criterion	66
7.3 Analysis of the constraints	66
7.4 Type of mathematical construction proposed	67
7.5 Parameters to be defined.....	69
7.6 Adjustments of the mathematical construction	69
7.7 Calculation	75
8 Criterion 8: Absence of pain induced by management procedures.....	77
8.1 Definition of the criterion	77
8.2 Measures to be used to assess the criterion	77
8.3 Analysis of the constraints	77
8.4 Type of mathematical construction proposed	77
8.5 Parameters to be defined.....	78
8.6 Adjustments of the mathematical construction	78
8.7 Calculation	79
9 Criterion 9: Expression of social behaviours.....	80
9.1 Definition of the criterion	80
9.2 Measures to be used to assess the criterion	80
9.3 Analysis of the constraints	80
9.4 Type of mathematical construction proposed	81

9.5	Parameters to be defined.....	81
9.6	Adjustments of the mathematical construction	81
9.7	Calculation	84
10	Criterion 10: Expression of other behaviours.....	85
10.1	Definition of the criterion	85
10.2	Measures to be used to assess the criterion	85
10.3	Analysis of the constraints	86
10.4	Type of mathematical construction proposed	86
10.5	Parameters to be defined.....	86
10.6	Adjustments of the mathematical construction	86
10.7	Calculation	95
11	Criterion 11: Good human-animal relationship.....	96
11.1	Definition of the criterion	96
11.2	Measures to be used to assess the criterion	96
11.3	Analysis of the constraints	96
11.4	Type of mathematical construction proposed	96
11.5	Parameters to be defined.....	96
11.6	Adjustments of the mathematical construction	97
11.7	Calculation	100
12	Criterion 12: Positive emotional state.....	101
12.1	Definition of the criterion	101
12.2	Measures to be used to assess the criterion	101
12.3	Analysis of the constraints	102
12.4	Type of mathematical construction proposed	102
12.5	Parameters to be defined.....	103
12.6	Adjustments of the mathematical construction	103
12.7	Calculation	105
Annex: List of the experts consulted to construct the criteria.....		107
Colophon.....		108

The Welfare Quality® Project is a European research project concerning the integration of animal welfare in the food quality chain: from public concern to improved welfare and transparent quality. Welfare Quality® is co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508.

The text of this report represents the authors' views and does not necessarily represent a position of the European Commission who will not be liable for the use made of such information.

Executive summary

Deliverable D2.8c contains the methods proposed to calculate criterion-scores and then principle-scores for the various animal types. The present document (D2.8c Part 2 - *Criteria construction for Veal calves on farm*) presents the construction of the criteria from measures taken on farm for veal calves. Each chapter of this deliverable details the construction of one criterion, following the common structure:

- *Definition of the criterion* (general definition, common to all animal types and period)
- *Measures used to construct the criterion* (measures are not detailed, see deliverables D2.10 to D2.13 for more information on the measures)
- *Analysis of the constraints* (elements having consequences on the construction and main description of the what the criterion assesses)
- *Type of mathematical construction proposed* (according to the measures and the identified constraints)
- *Parameters to be defined* (list of the parameters needed to assess the criterion according to the construction previously proposed)
- *Adjustments of the mathematical construction* (including the questions asked to experts, their answers and the exact adjustments made, during first and second consultation processes)
- *Calculation* (definition of the variables, from raw data or obtained by calculation, and algorithm summarising the construction)

The various methods used for the construction of the criteria in veal calves are listed in Table 0.1.

Table 0.1: Type of construction of welfare criteria in veal calves

Welfare criterion	No. of measures	Type of data	Construction
Absence of prolonged hunger	1	Cardinal measure at farm level	Calculation of an index based on the % of very lean calves (more than 30% behind the rest of the batch), transformed into a score using least-squares spline curve fitting
Absence of prolonged thirst	4	Qualitative or ordinal measures at group level	Lexicographic method leading to a given number of possible cases, scores directly attributed to each of the cases
Comfort around resting	2	Cardinal measures at farm level	Calculation of 2 indexes (1 for evidence of bursae & 1 for cleanliness) each based the % of animals with problems, transformed using least-squares spline curve fitting. These 2 sub-scores are synthesized thanks to a Choquet integral
Thermal comfort	1	Cardinal measure at farm level	Calculation of an index based on the % of wet animals, transformed into a score using least-squares spline curve fitting
Ease of Movement	1	Ordinal measure at group level	Resource-based measure describing slipperiness of the floor
Absence of injuries	2	Cardinal measures at animal level	Calculation of 2 indexes (1 for lameness & 1 for bitten tail/ear) each based on the % of animals with the problems, transformed using least-squares spline curve fitting. These 2 sub-scores are synthesized thanks to a Choquet integral
Absence of disease	13	Cardinal measures at farm level	Each measure is transformed onto an ordinal scale: no/moderate/severe problem ; Calculation of an index based on a linear combination of the proportions of moderate & severe problems and transformed using least-squares spline curve fitting
Absence of pain due to management procedures	0	Binary or qualitative measures at farm level	Not relevant for veal calves
Expression of social behaviours	1	Cardinal measures at farm level	Calculation of an index based on the frequency of social licking, transformed using least-squares spline curve fitting
Expression of other behaviours	2	Cardinal measures at farm level	Calculation of 2 indices (1 for abnormal oral behaviours & 1 for signs of urine drinking/cross-sucking): from behavioural observations the first one (% calves performing behaviours), and from clinical visit the second one(% of calves interested by problem); transformed using least-squares spline curve fitting. These 2 sub-scores are synthesized thanks to a Choquet integral
Good human-animal relationship	1	Cardinal measure at farm level	Calculation of an index based on the % calves that can be touched, transformed using least-squares spline curve fitting
Positive emotional state	1	Cardinal measure at farm level	20 qualitative terms are used. The appropriateness of each term to describe the farm is expressed on a 12.5 cm scale. A weighted sum is performed to turn the 20 data obtained into a single index which is turned into a score using least-squares spline curve fitting.

Key words: animal welfare, overall assessment, multicriteria evaluation, welfare criteria, veal calves, on farm

1 Criterion 1: Absence of prolonged hunger

1.1 Definition of the criterion

In this criterion, we aim at assessing whether the animals suffer or not from prolonged hunger or rather if they present or not signs of malnutrition. On farms, the focus is on chronic hunger and thus on the detection of very lean animals.

1.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13101	Calves' mean estimated weight and condition	Ordinal	Individual

Only **one** measure is foreseen for this criterion: **poor body condition** of each calf in comparison to the average of the batch. Then, two categories of calves are distinguished:

- normal condition - calves with a body condition above 85% of the average,
- very lean calves - calves with a lower body condition, approx. 30% less than average.

1.3 Analysis of the constraints

There is only 1 measure for this criterion

1.4 Type of mathematical construction proposed

At farm level, the considered measure will give the percentage of very lean calves:

Categories	Normal	Very lean
of observed animals	p_0	p

One score is calculated:

$$S = f(p) \quad \text{with } p, \text{ very lean calves}$$

1.5 Parameters to be defined

We need:

- to define the most appropriate function f to transform the index based on the of very lean calves into the criterion score.

1.6 Adjustments of the mathematical construction

▪ Questions asked to experts

A dataset corresponding to virtual farms (Table 1.1) was proposed to experts who were asked:

- to attribute a criterion score between 0 and 100 to each farm of the virtual dataset considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 1.1).
The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

Table 1.1: Dataset submitted to experts. They had to attribute criterion scores

	% of very lean calves	Score
Farm 1	0	
Farm 2	0.5	
Farm 3	1	
Farm 4	1.5	
Farm 5	2	
Farm 6	3	
Farm 7	5	
Farm 8	10	
Farm 9	20	
Farm 10	50	
Farm 11	100	

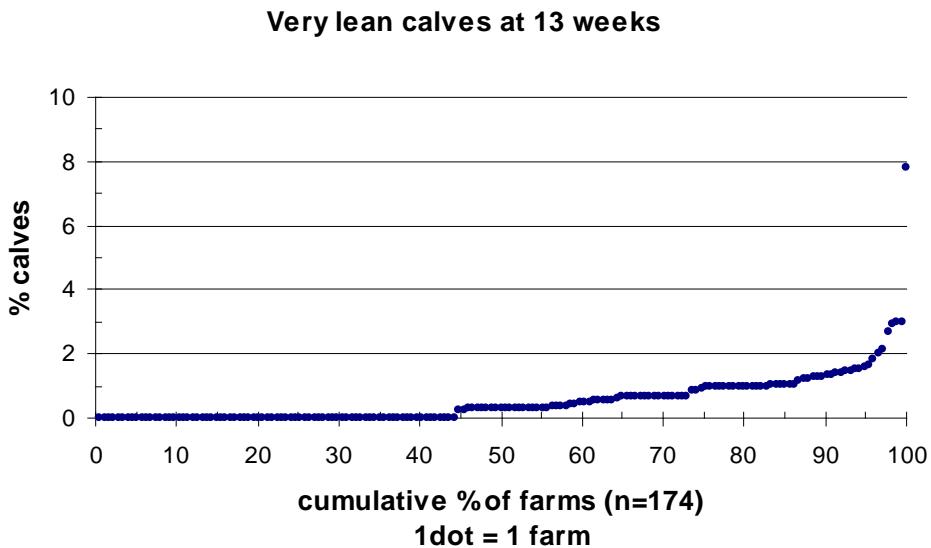


Figure 1.1: Distribution of % of very lean calves in 174 farms at 13 wks of fattening

- **Adjustment of the construction according to answers from experts**

The results of experts' evaluation of farms with the scores attributed by the 7 experts for the % of very lean calves are shown in Table 1.2 (Columns 3-9). Means of criterion-scores attributed by the 7 experts were calculated (Column 10) and farms were sorted according to the mean criterion-score.

The indexes resulting from the calculation based on the % of very lean calves (100-% of very lean calves) is shown in Table 1.2, Column 11.

Table 1.2: Results obtained from the experts asked

% very lean calves	Score for very lean calves							Mean	Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7		
0	100	100	100	100	100	100	100	100	100
0.5	90	100	75	80	75	80	80	83	99.5
1	80	80	60	60	50	40	70	63	99
1.5	70	70	50	50	45	20	50	51	98.5
2	60	50	40	35	40	17	30	39	98
3	50	40	30	25	20	15	20	29	97
5	40	30	20	15	15	10	10	20	95
10	30	20	10	10	10	5	5	13	90
20	20	20	0	0	5	0	0	6	80
50	0	0	0	0	0	0	0	0	50
100	0	0	0	0	0	0	0	0	0

Finally I-functions were defined so as to best match expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 1.2). The parameters of these spline functions are:

Coëfficiënten:

COEF	VALUE
a1	0.00000000000
b1	0.00000000000
c1	0.00000000000
d1	0.0000155696
a2	-3325.6689964213
b2	117.3765527561
c2	-1.3809006200
d2	0.0054308662
a3	-457074.7593237738
b3	14446.2947737734
c3	-152.2116143919
d3	0.5346614259

Knot(s): 2 interior knots set at 85 and 95.

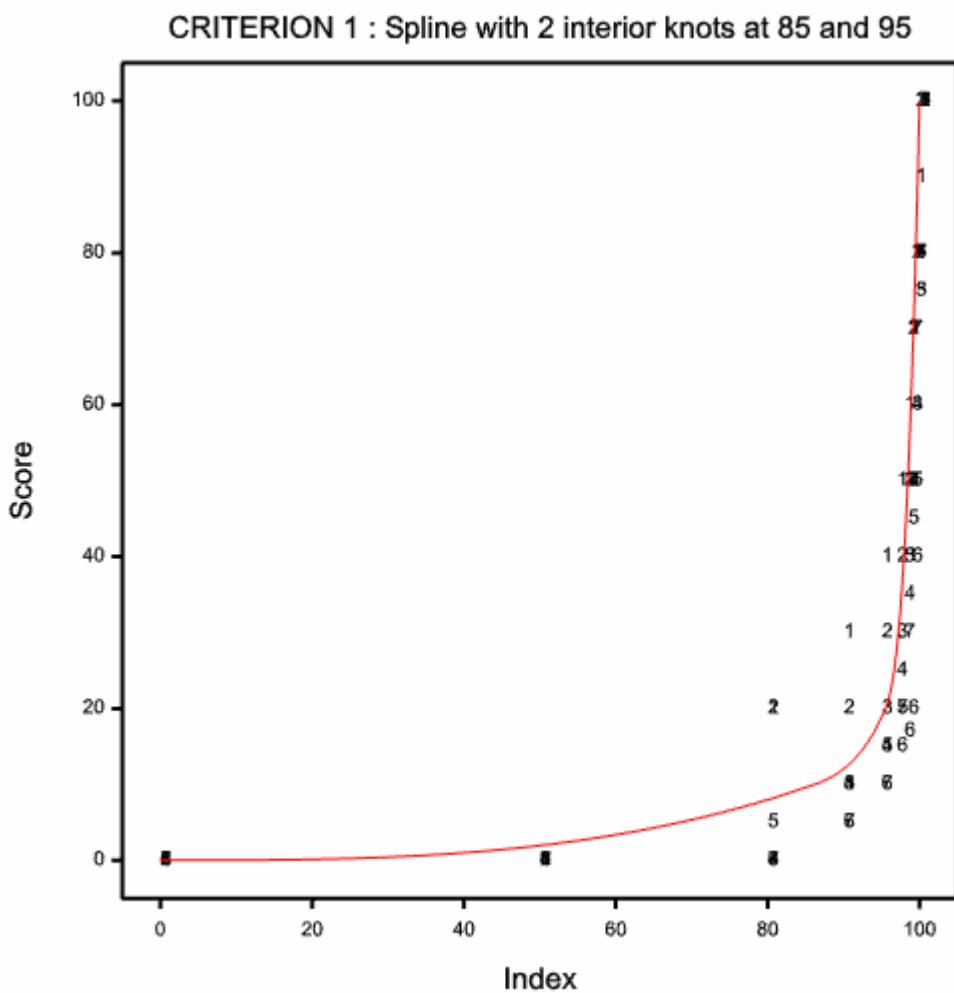


Figure 1.2: I-Spline function allowing to attribute a score for absence of hunger according to the proportion of very lean calves compared to the batch average

1.7 Calculation

The % of very lean calves is used to calculate an index:

$$\text{Index} = 100 - \% \text{ of very lean calves}$$

Then the index is transformed into a score with spline functions as follows:

$$\text{When } I \leq 85 \text{ then Score} = 0.0000155696 \times I^3$$

$$\text{When } 85 \leq I \leq 95 \text{ then Score} = -3325.669 + (117.37655 \times I) - (1.3809006 \times I^2) + (0.0054308662 \times I^3)$$

$$\text{When } I \geq 95 \text{ then Score} = -457074.76 + (14446.2948 \times I) - (152.211614 \times I^2) + (0.5346614259 \times I^3)$$

2 Criterion 2: Absence of prolonged thirst

2.1 Definition of the criterion

In this criterion, we aim at assessing whether the animals suffer or not from thirst. On farms, it is likely that thirst, when not enough water is available, does not lead to dehydration; hence the criterion will be assessed through water availability and not through signs of dehydration.

2.2 Measures to be used to assess the criterion

Id	Meas.	Measure	Nature	Level of observation
13102		Provision of water to calves	Binary	Group
13103		Timing of water supply to calves	Ordinal	Group
13104		Cleanliness of the drinkers	Binary	Group
13105		Number of water bowls per animal	Binary	Group

Because signs of dehydration on animals are observable only in extreme cases, it has been decided to assess absence of prolonged thirst through management-based measures. For each pen of calves, the observer shall answer the questions:

- Is water provided apart from milk replacer? (yes/no)

If yes:

- Are the drinkers clean?
- Is water available throughout the day?
- Are the drinkers in sufficient number according to recommendation?

After consultation with vets specialised in veal calves, it was decided that:

- When calves have access to water for less than 12 h a day, the access is considered not ad libitum and when water is available for 6 h or less and/or not available during daytime, the access is considered as restricted;
- At present, the recommendation from the Council of Europe for cattle (still under discussion) mentions that there should be one drinking place for 10 animals. However, because veal calves are provided with large amount of milk replacer, it is here considered that 1 water bowl for 20 calves is sufficient to cover the needs of calves for drinking water.

The field vets consulted were: T. Lorent (France), N Ferri (Italy), E. Canali (Italy).

2.3 Analysis of the constraints

Each pen of calves is described according to the 4 items listed above, which are evaluated separately. The four items bring 13 possible combinations (Figure 2.1).

The impact of each combination on calves' welfare can be evaluated. The criterion score may thus take only a finite number of values, that is one per combination.

Several pens may be observed on a given farm. This will require some integration of the information from groups to farm.

Since white veal receive milk replacer and, therefore, in contrast to pink veal are not dependent upon fresh water in order to obtain sufficient fluids, the minimum score assigned to white veal (situation y_1 , see Figure 2.1) is 30. The minimum score for pink veal, in contrast is 0.

2.4 Type of mathematical construction proposed

We propose

- to attribute a score to each of the 13 combinations according to the severity of the constraints they pose on animals,
- at farm level, to attribute a criterion-score equal to the worst situation observed on 10 of the pens

Hence the criterion-scores are discrete.

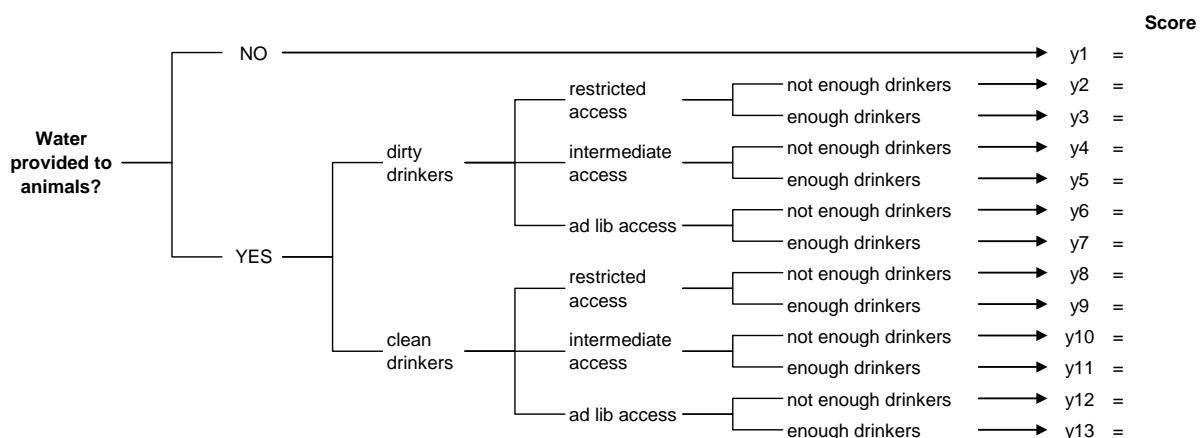


Figure 2.1: Combinations of conditions for water provision.

2.5 Parameters to be defined

We need to define the scores attributed to each of the 13 combinations (y1 to y11).

2.6 Adjustments of the mathematical construction

- **Questions asked to experts**

Figure 2.1 was shown to experts. They were asked to estimate scores y1 to y13.

The experts asked were B.J. Lensink, H. Leruste, K. Van Reenen, I. Veissier

- **Adjustment of the construction according to answers from experts**

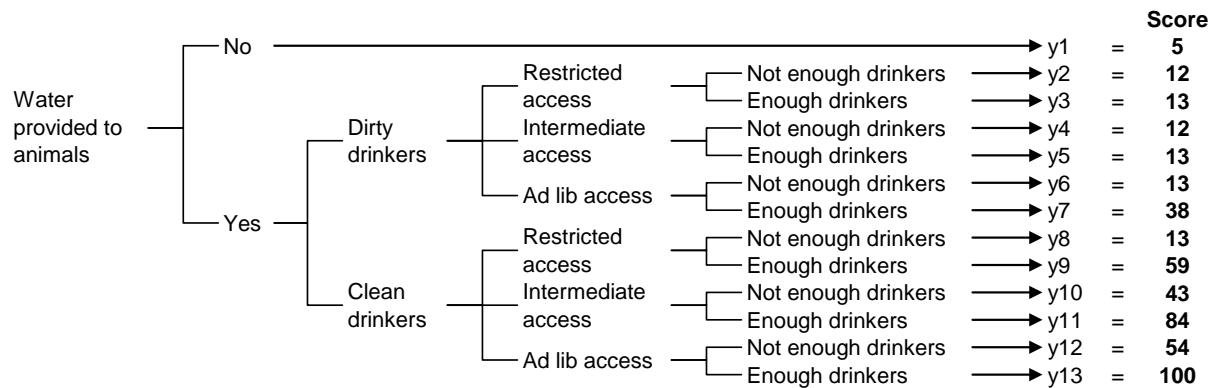
Table 2.1: Evaluation of the 9 combinations by experts

Situation	Criterion scores				
	<i>Expert 1</i>	<i>Expert 2</i>	<i>Expert 3</i>	<i>Expert 4</i>	<i>Mean</i>
y1	20	0	0	0	5
y2	22	5	15	5	12
y4	23	5	15	5	12
y3	25	5	15	5	13
y6	25	5	15	5	13
y5	26	5	15	5	13
y8	27	5	15	5	13
y7	50	40	45	15	38
y10	60	40	50	20	43
y12	70	40	75	30	54
y9	70	60	55	50	59
y11	80	80	95	80	84
y13	100	100	100	100	100

The mean of scores attributed by the experts to a given combination is considered the criterion-score to be attributed to a group of calves submitted to this combination.

2.7 Calculation

A decision tree is used to attribute a score to according to the combination of water supply, timing for water availability, cleanliness of drinkers, number of drinking points per animals:



The number of drinking points is considered sufficient if there is one point for 20 animals

The scores assigned to the 13 combinations should be repeated, and should be performed separately for white and pink veal. Scores assigned to white veal should range between 30 and 100, whereas scores for pink veal should range between 0 and 100.

3 Criterion 3: Comfort around resting

3.1 Definition of the criterion

This criterion assesses the comfort of the animals around resting, through the observation of presence of signs of bursae defined as swelling of the carpal bursae and the cleanliness of the calves.

3.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13106	Evidence of bursae	Binary	Individual
13107	Cleanliness of calves	Binary	Individual

The measures related to comfort around resting lead to only two variables that can be used to construct this criterion:

- evidence of bursae,
- dirty calves (calves with more than 25% of the surface covered by manure).

3.3 Analysis of the constraints

Evidence of bursae is assessed during the clinical/health visit and there is no behavioural (lying/resting) measure for comfort around resting. Evidence of bursae replaces the first proposed measure that regarded the lying position since experts were not confident with the measure "lying calves with no leg stretched" due to its frequency and validity at on-farm conditions (calves might be standing when the assessor is present). It has been suggested to include in Criterion "comfort around resting" also the measure regarding presence of hairless spots since, in veal calves, they are due to not proper housing facilities. However, there is no final decision on this measure.

Experts suggested to give less importance to the cleanliness of the calves, therefore, the new expert consultation regards the measure "evidence of bursae" and the combination of "evidence of bursae" and "cleanliness". Cleanliness alone was already submitted to experts in a previous consultation.

3.4 Type of mathematical construction proposed

The proposed way of reasoning in two stages will be maintained:

First, two elementary scores are attributed:

- a score for bursae from the % of calves showing signs of bursae:

$$S_a = f_a(p_a) \quad \text{with } p_a, \text{ calves with signs of bursae}$$

- a cleanliness score from the calves dirty:

$$S_b = f_b(p_b) \quad \text{with } p_b, \text{ dirty calves}$$

Second, these two elementary scores are bring together to form the criterion-score:

$$\text{Criterion score} = f(S_a, S_b)$$

3.5 Parameters to be defined

We need to define functions f_a , f_b , and f .

3.6 Adjustments of the mathematical construction

- **Questions asked to experts**

Two sets of data were shown to experts:

- one for bursae (Table 3.1a), considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 3.1).
- one for cleanliness score (Table 3.1b).
- one for the combination of the two elementary scores bursae and cleanliness (Table 3.2)

In regards to the cleanliness score, results from previous consultation will be maintained; experts involved were B.J. Lensink, H. Leruste, K. Van Reenen, I. Veissier.

In regards to bursae and the combination of bursae and cleanliness the experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

Table 3.1a: Dataset submitted to experts. They were asked to attribute bursae scores according to the % of calves with bursae

	% of calves with evidence of bursae	Score
Farm 1	0	
Farm 2	2	
Farm 3	5	
Farm 4	8	
Farm 5	10	
Farm 6	15	
Farm 7	20	
Farm 8	50	
Farm 9	70	
Farm 10	80	
Farm 11	100	

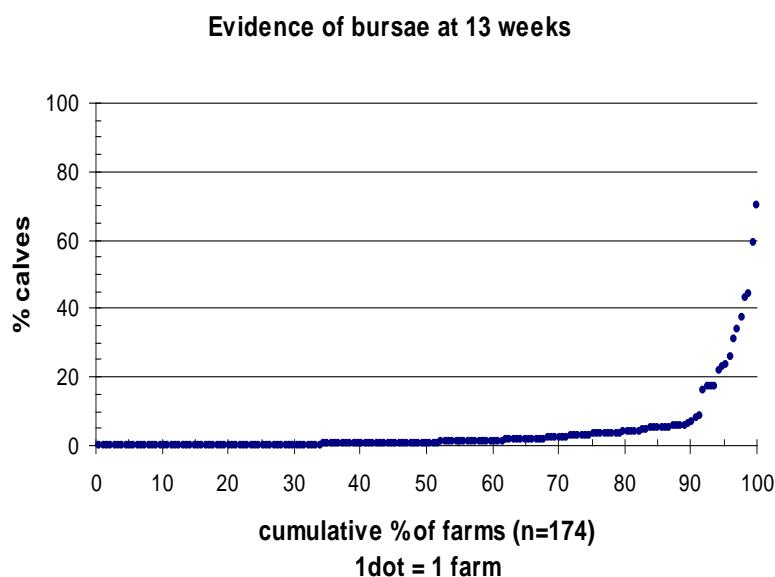


Figure 3.1: Distribution of % of calves with bursae in 174 farms at 13 wks of fattening

Table 3.1b: Dataset submitted to experts in the first consultation. They were asked to attribute cleanliness scores according to the % of calves that are dirty

	% calves dirty	Cleanliness score
Farm 1	0	
Farm 2	10	
Farm 3	20	
Farm 4	30	
Farm 5	40	
Farm 6	50	
Farm 7	60	
Farm 8	70	
Farm 9	80	
Farm 10	90	
Farm 11	100	

Table 3.2: Dataset submitted to experts. They were asked to attribute criterion scores for comfort around resting according to bursae score and cleanliness score

	Score for bursae	Score for cleanliness	Score
Farm 1	40	60	
Farm 2	50	50	
Farm 3	60	40	

- **Adjustment of the construction according to answers from experts**

Calculation of the bursae score

The results of experts' evaluation of % calves with bursae are shown in Table 3.3a.

First, the following index is calculated:

$$\text{Index}_a = 100 - (\% \text{ of calves with bursae})$$

The function f_a is based on I-Spline functions so as to transform this index into a score that matches best expert scores for bursae (i.e. minimise the sum of square errors with expert scores) (Figure 3.2a). The parameters of the spline functions are:

Knot(s): 2 interior knots at 80 and 85.

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	0.0000000000
c1	0.0000000000
d1	0.0000327033
a2	-8016.4679050037
b2	300.6175493252
c2	-3.7577194019
d2	0.0156898676
a3	2530.2182688138
b3	-71.6184300316
c3	0.6215273748
d3	-0.0014836490

CRITERION 3 Bursae score : Spline with 2 interior knots at 80 and 85

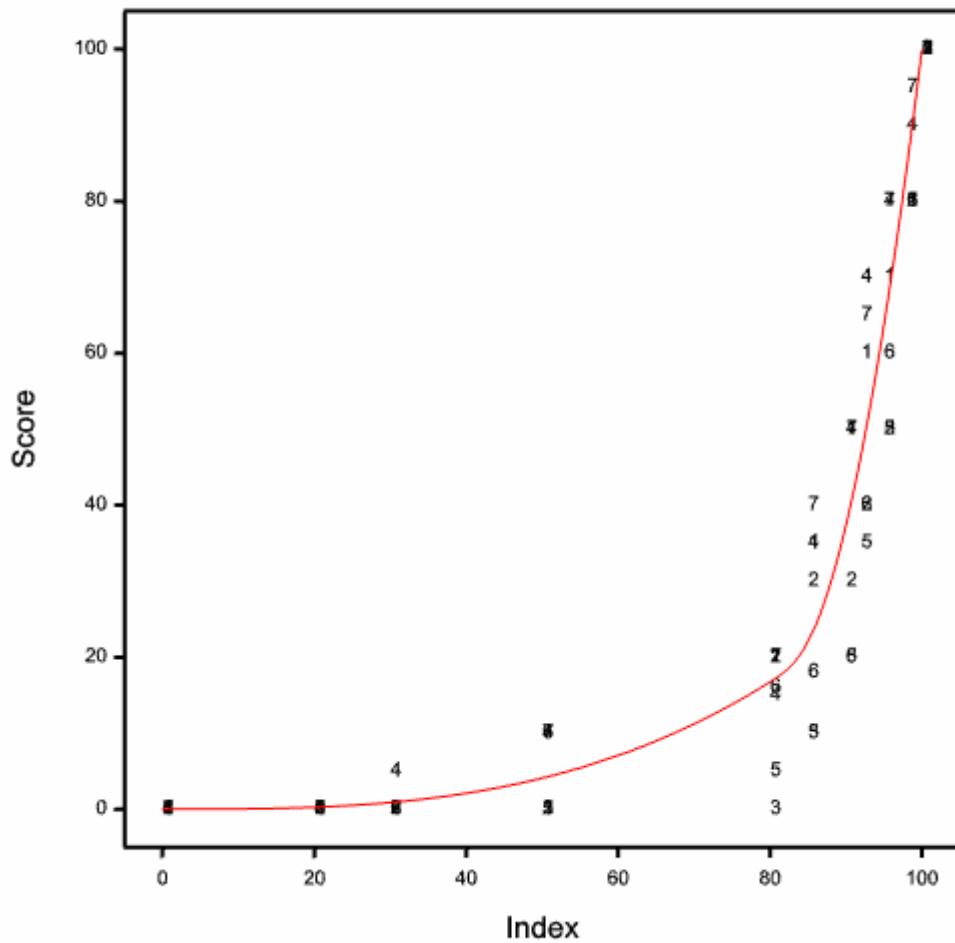


Figure 3.2a: I-Spline function allowing to attribute a score for calves with bursae

Table 3.3a: Results obtained from the experts asked to attribute bursae scores

% calves with bursae	Score for bursae							Index Mean
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
	1	2	3	4	5	6	7	
0	100	100	100	100	100	100	100	100 100
2	80	80	80	90	80	80	95	84 98
5	70	50	50	80	50	60	80	63 95
8	60	40	40	70	35	40	65	50 92
10	50	30	20	50	20	20	50	34 90
15	35	30	10	35	10	18	40	25 85
20	20	20	0	15	5	16	20	14 80
50	0	0	0	10	0	10	10	4 50
70	0	0	0	5	0	0	0	1 30
80	0	0	0	0	0	0	0	0 20
100	0	0	0	0	0	0	0	0 0

Calculation of the cleanliness score

The results of experts' evaluation of % calves dirty are shown in Table 3.3b.

First, the following index is calculated:

$$\text{Index}_b = 100 - (\% \text{ of dirty calves})$$

The function f_b is based on I-Spline functions so as to transform this index into a score that matches best expert scores for cleanliness (i.e. minimise the sum of square errors with expert scores) (Figure 3.2b). The parameters of these functions are:

Knot(s): 1 interior knot at 40

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	1.0782225212
c1	-0.0207495505
d1	0.0002554469
a2	16.5255333610
b2	-0.1611924809
c2	0.0102358245
d2	-0.0000027645

Index:

$$\text{When } l \leq 40 \quad \text{then Score} = (1.07822 \times l) - (0.0207496 \times l^2) - (0.0002554469 \times l^3)$$

$$\text{When } l \geq 40 \quad \text{then Score} = 16.526 - (0.161192 \times l) + (0.01023582 \times l^2) - (0.0000027645 \times l^3)$$

Table 3.3b: Results obtained from the experts asked to attribute cleanliness scores

% dirty calves	Score for cleanliness					Index
	Expert 1	Expert 2	Expert 3	Expert 4	Mean	
0	100	100	100	100	100	100
10	80	90	80	80	82.5	90
20	65	75	75	60	68.75	80
30	50	60	50	50	52.5	70
40	40	45	50	40	43.75	60
50	30	30	50	30	35	50
60	25	20	45	20	27.5	40
70	20	15	20	15	17.5	30
80	15	10	20	15	15	20
90	10	5	20	10	11.25	10
100	5	0	0	0	1.25	0

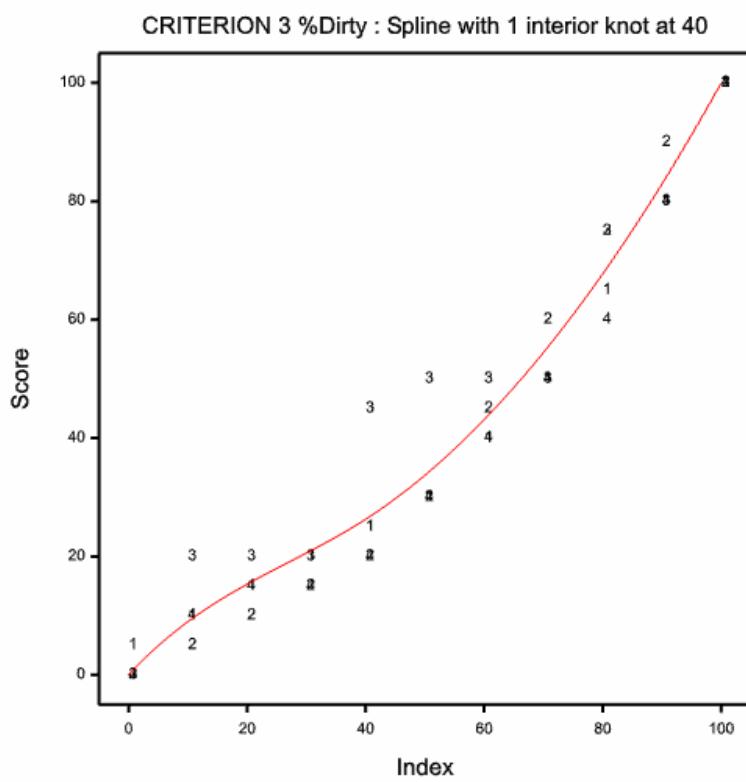


Figure 3.2b: I-Spline function allowing to attribute a cleanliness score for dirty calves

Combination of bursae score and cleanliness score into criterion score

A Choquet integral is calculated:

Criterion score (S)

$$\begin{array}{ll} \text{if } S_a < S_b & \text{then } S = S_a + (S_b - S_a) \times \mu_b \\ \text{else} & S = S_b + (S_a - S_b) \times \mu_a \end{array}$$

The parameters of the Choquet integral are determined thanks to a linear optimisation in order to best match the answers of experts (Table 3.4). This led to:

Capacity of bursae score $\mu_a = 0.66$

Capacity of the cleanliness score $\mu_b = 0.31$

Sum of capacities $= 0.97$

Indices of importance (Shapley values):

bursae score $\Phi(a) = \frac{1}{2} \times (1 + \mu_a - \mu_b) = 0.66$

cleanliness score $\Phi(b) = \frac{1}{2} \times (1 + \mu_b - \mu_a) = 0.31$

Interaction index $= 1 - \mu_a - \mu_b = 0.03$

The values show that:

- Bursae is considered twice as important as cleanliness;
- There is no interaction between the two scores. Hence the overall score results from a mere weighted sum.

Table 3.4: Estimation of criterion-scores by the 7 experts. Choquet integrals are used to estimate calculated score that best fit with the scores attributed by experts. The last column of the table show the error between the scores attributed by an expert and the calculated score

Expert	Bursae score	Cleanliness score	Criterion score attributed by expert	Calculated score	Error ²
	1	2			
Expert 1	40	60	60	46	192.02
	50	50	50	50	0.00
	60	40	40	53	172.73
Expert 2	40	60	40	46	37.73
	50	50	50	50	0.00
	60	40	60	53	47.02
Expert 3	40	60	48	46	3.45
	50	50	50	50	0.00
	60	40	52	53	1.31
Expert 4	40	60	45	46	1.31
	50	50	50	50	0.00
	60	40	50	53	9.88
Expert 5	40	60	40	46	37.73
	50	50	50	50	0.00
	60	40	60	53	47.02
Expert 6	40	60	45	46	1.31
	50	50	50	50	0.00
	60	40	55	53	3.45
Expert 7	40	60	45	46	1.31
	50	50	50	50	0.00
	60	40	55	53	3.45
					Σ Error ²
					560.00

3.7 Calculation

Two indexes are calculated, one for each measure:

$$I_b = 100 - (\% \text{ calves with bursae})$$

$$I_c = 100 - (\% \text{ dirty calves})$$

They are transformed into scores:

the index for bursae is computed into a score thanks to I-spline functions as follows:

When $I_b \leq 80$ then Score = $(2.696595 E^{-12} \times I_b) - (1.334287 E^{-13} \times I_b^2) + (0.0000401 \times I_b^3)$

When $I_b \geq 80$ then Score = $-14160.19 + (505.721 \times I_b) - (6.01969 \times I_b^2) + (0.023885 \times I_b^3)$
the index for cleanliness is computed into a score thanks to I-spline functions as follows:

When $I_s \leq 40$ then Score = $(2.55206816 \times I_s) - (0.0273235 \times I_s^2) + (0.0001061 \times I_s^3)$

When $I_s \geq 40$ then Score = $-3.5337 + (2.817 \times I_{sl}) - (0.033949 \times I_s^2) + (0.0001613 \times I_s^3)$

S_l and S_s are combined to produce a score for comfort around resting (S) thanks to a Choquet integral where:

$$\mu_l = 0.66$$

$$\mu_c = 0.31$$

4 Criterion 4: Thermal comfort

4.1 Definition of the criterion

In this criterion, we aim at assessing how the animal copes with its thermal environment, that is whether it suffers from drafts, extreme temperatures or humidity.

4.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13108	Wet calves	Cardinal	individual

Thermal comfort is assessed by the % of wet calves, this is supposed to be indicative of the quality of ventilation.

4.3 Analysis of the constraints

Only one measure is taken.

4.4 Type of mathematical construction proposed

One score is calculated:

$$S = f(p) \quad \text{with } p, \text{ wet calves}$$

4.5 Parameters to be defined

We need to define function f .

4.6 Adjustments of the mathematical construction

▪ Questions asked to experts

One set of data was shown to experts (Table 4.1). Experts were asked to attribute a criterion-score for thermal comfort considering the distribution of the % of wet calves on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 4.1).

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

Table 4.1: Dataset submitted to experts. They were asked to attribute a score for thermal comfort according to % of wet calves

	% wet calves	Score
Farm 1	0	
Farm 2	0.5	
Farm 3	1	
Farm 4	2	
Farm 5	5	
Farm 6	10	
Farm 7	20	
Farm 8	30	
Farm 9	60	
Farm 10	80	
Farm 11	100	

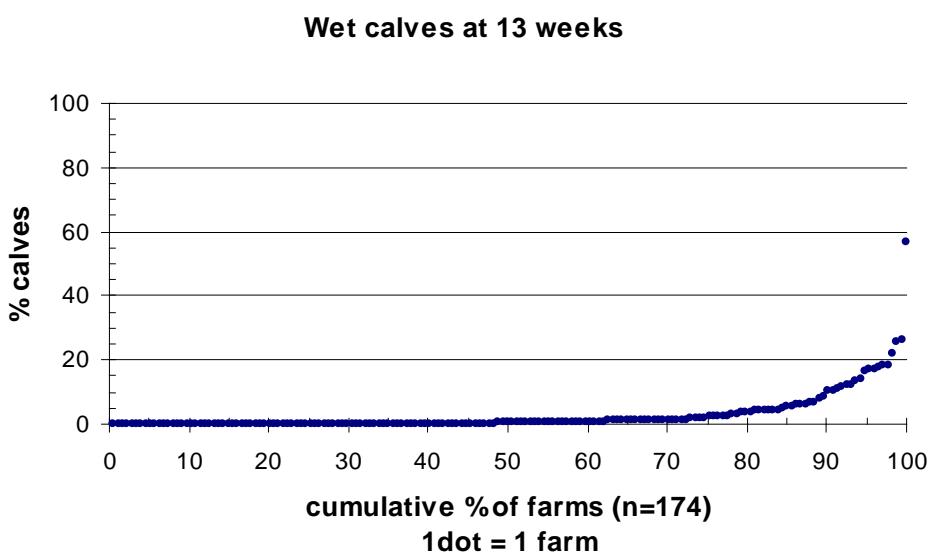


Figure 4.1: Distribution of the % of wet calves in 174 farms at 13 wks of fattening

- **Adjustment of the construction according to answers from experts**

Answers from experts are shown in table 4.2

Table 4.2: Scores attributed by experts to % of wet calves in relation to thermal comfort

% wet calves	Score for wet calves							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	100	100	100	100	100	100	100	100
0.5	85	90	90	95	80	80	95	88
1	75	80	80	90	70	70	80	78
2	65	60	60	80	60	50	60	62
5	55	50	40	50	30	40	50	45
10	45	40	20	30	20	20	40	31
20	35	20	0	20	10	10	30	18
30	20	20	0	15	5	0	20	11
60	0	0	0	10	0	0	5	2
80	0	0	0	5	0	0	0	1
100	0	0	0	0	0	0	0	0

First, the following index is calculated:

$$\text{Index} = 100 - (\% \text{ of wet calves})$$

The function f is based on I-Spline functions so as to transform this index into a score that matches best expert scores for wet calves (i.e. minimise the sum of square errors with expert scores) (Figure 4.2). The parameters of the spline functions are:

Knot(s): 2 interior knots at 60 and 96.

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	0.0000000000
c1	0.0032213345
d1	-0.0000206271
a2	-178.6409247424
b2	8.9320462368
c2	-0.1456461028
d2	0.0008064142
a3	-490697.2763798839
b3	15337.6395253617
c3	-159.8196835837
d3	0.5552301597

CRITERION 4 Wet calves : Spline with 2 interior knots at 60 and 96

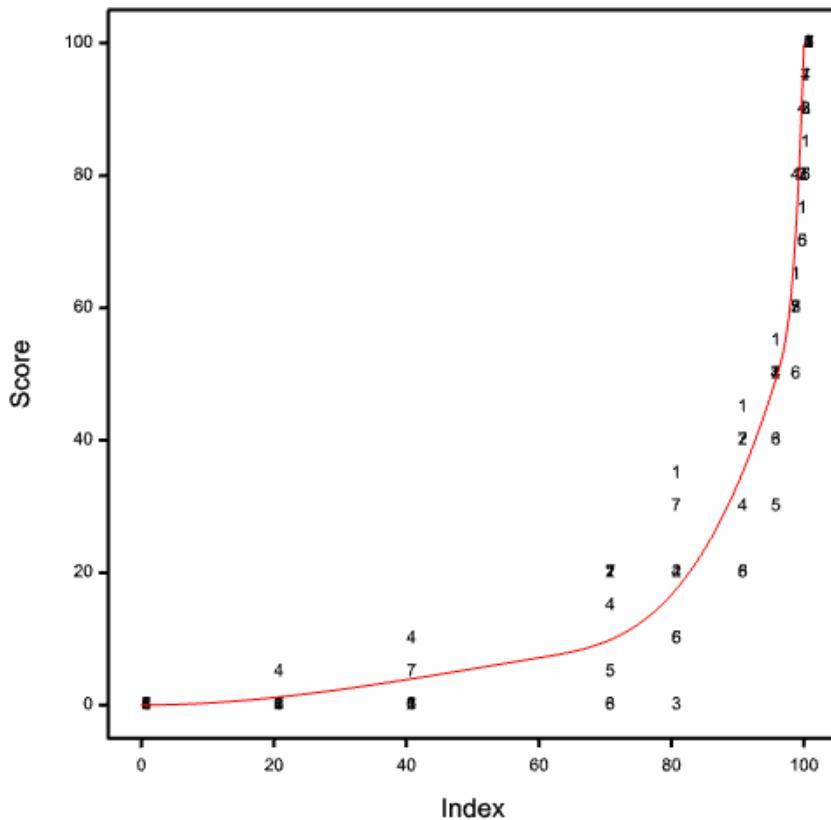


Figure 4.2: I-Spline function allowing to attribute a score for thermal comfort according to the % of wet calves

4.7 Calculation

Index $I = 100 - (\% \text{ wet calves})$

Index = I

$$\text{When } I \leq 60 \text{ then Score} = (0.00322133 \times I^2) - (0.0000206271 \times I^3)$$

$$\text{When } 60 \leq I \leq 96 \text{ then Score} = -178.641 + (8.93205 \times I) - (0.14564610 \times I^2) + (0.0008064142 \times I^3)$$

$$\text{When } I \geq 96 \text{ then Score} = -490697.28 + (15337.6395 \times I) - (159.8196836 \times I^2) + (0.5552301597 \times I^3)$$

5 Criterion 5: Ease of movement

5.1 Definition of the criterion

This criterion aims at assessing the locomotion of the animals, i.e. the opportunity to move and the ease of locomotion, linked closely to the quality of the environment and not due to the health of the animal (health is considered elsewhere, in Subcriteria 6, 7 and 8)

5.2 Measures to be used to assess the criterion

Id	Meas.	Measure	Nature	Level of observation
13109		Slipperiness of the floor	Ordinal	Group

Ease of movement is assessed in a pen by observing the floor, feeling how much grip there is on the floor and by taking a good look how the calves walk. Several categories of floors are distinguished:

- Dangerous: floors with openings in the floor that are big enough that a leg of calf can fall through, i.e. missing slats;
- Very slippery: it is difficult to stand on the floor, calves can't move easily and safely. This is observed i.e. on old hard wooden slats with a round surface instead of a flat one, which are wet due to urine and faeces, and with or without antiskid features;
- Slippery: eg hard wooden slats with flat surfaces wet of urine and faeces and with or without (aluminium) antiskid features;
- Moderately slippery: i.e. dry hard wooden slats with flat surfaces and with (aluminium) antiskid features; certain slats topped with rubber; wet concrete (slatted) floors;
- Slightly slippery: Slipperiness is limited to a large extent, observed with i.e. certain slats topped with rubber, dry concrete (slatted) floors, floors covered with a reasonable (i.e. 10 cm) layer of straw;
- Not at all slippery: Calves never slip, this could be obtained with a thick layer of straw (i.e. 40 cm), a floor covered with sand or outside at pasture.

5.3 Analysis of the constraints

There is only one measure, which is ordinal. The above classification needs improvement. This measure has not been applied in practice so far.

5.4 Type of mathematical construction proposed

A criterion-score is attributed to each category of slipperiness. This results in discrete scores. At farm level, we will attribute a criterion-score equal to the worst situation observed that concerns at least 15 of the calves.

5.5 Parameters to be defined

We need to define the scores attributed to each of category of slipperiness.

5.6 Adjustments of the mathematical construction

▪ Questions asked to experts

The description of the 6 categories were presented to experts and they have to attribute a score to each of them. The experts were first E. Bokker, M. Wolthuis-Fillerup and H. Spoolder who used that measure in WP2.4. They proposed scores (Table 5.1) and these were further discussed with I. Veissier , B.J. Lensink, H. Leruste, K., G. Cozzi

Table 5.1: Scores proposed by partners who used the measures in WP2.4

Category	Score
Dangerous	0
Very slippery	15
Slippery	25
Moderately slippery	50
Slightly slippery	75
Not at all slippery	100

▪ Adjustment of the construction according to answers from experts

The final results of the expert consultation are shown on Table 5.2. All experts agreed with the scores initially proposed.

Table 5.2: Final results of the expert consultation

Category	Initial proposed score	Expert 1	Expert 2	Expert 3
Dangerous	0	0	0	0
Very slippery	15	15	15	15
Slippery	25	25	25	25
Moderately slippery	50	50	50	50
Slightly slippery	75	75	75	75
Not at all slippery	100	100	100	100

5.7 Calculation

A score for ease of movement is attributed to each pen observed according to the slipperiness of the floor:

Slipperiness of the floor	Score
Dangerous	0
Very slippery	15
Slippery	25
Moderately slippery	50
Slightly slippery	75
Not at all slippery	100

Then the worse score obtained for at least 15 animals of the farm is attributed to that farm.

6 Criterion 6: Absence of injuries

6.1 Definition of the criterion

In this criterion, we aim at assessing external injuries due to physical causes and not to pathogens (which are included in Criterion 7). Voluntary interventions on animals (i.e. mutilations) are excluded and considered in a separate criterion (8). Lameness, whatever the cause, is also included in Criterion 6 as it is difficult to identify its causes and more likely due to a mechanical cause.

6.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13110	Lameness	Binary	Individual
13111	Bitten tail/ear	Binary	Individual

The measures are defined as follows:

- Lameness – the calf shows abnormality of movement caused by reduced ability to use one or more limbs in a normal manner
- Bitten tail/ear – the calf presents chewing wounds as damaged tail/ear with/without presence of fresh blood or scab

6.3 Analysis of the constraints

These two measures are scored independently one from the other without a scale for severity of the injury. As such we do not keep track if both injuries are on the same animal but we have the following data at farm level:

Regarding lameness:

Categories	Normal	Lame
% of observed calves	p_0	p_1

with $p_0 + p_1 = 100$

Regarding bitten tail/ear:

Categories	Normal	Bitten tail/ear
% of observed calves	q_0	q_1
with $q_0 + q_1 = 100$		

Lameness and chewing wounds (bitten tail/ear) cover separate areas related to absence of injuries. Experts fell more confident to take them into account separately.

6.4 Type of mathematical construction proposed

We propose to reason in two stages:

- first, two elementary scores are attributed : a score for lameness and a score for bitten tail/ear,
- second, these two elementary scores are bring together to form the criterion score.

The score for lameness will be directly derived from the % of lame calves:

$$\text{Score for lameness, } S_a = f_a(p_1)$$

The score for bitten tail/ear will be directly derived from the % of calves with bitten tail/ear:

$$\text{Score for bitten tail/ear, } S_b = f_b(q_1)$$

The criterion score is a function of

$$\text{Criterion score} = f(S_a, S_b)$$

6.5 Parameters to be defined

We need to define:

- for lameness: the most appropriate function f_a to transform the % of lame calves into a score for lameness,
- for bitten tail/ear: the most appropriate function f_b to transform the % of calves with bitten tail/ear into a score for bitten tail
- the most appropriate function f to merge the two elementary scores into the criterion-score for absence of injuries.

6.6 Adjustments of the mathematical construction

- **Questions asked to experts**

Three sets of data were shown to experts:

- one for lameness, with % of calves lame (Table 6.1a),
- one for bitten tail/ear, with % of calves with bitten tail/ear (Table 6.1b),
- one for the combination of the two elementary scores (Table 6.2).

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

They were asked:

- to attribute scores for lameness considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 6.1a).
- to attribute scores for bitten tail/ear considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 6.1b).
- to attribute a criterion-score for absence of injuries to the combinations of the two elementary scores.

Table 6.1a: Dataset submitted to experts regarding lameness. They were asked to attribute lameness scores

	% calves lame	Score for lameness
Farm 1	0	
Farm 2	1	
Farm 3	2	
Farm 4	3	
Farm 5	5	
Farm 6	10	
Farm 7	15	
Farm 8	20	
Farm 9	50	
Farm 10	80	
Farm 11	100	

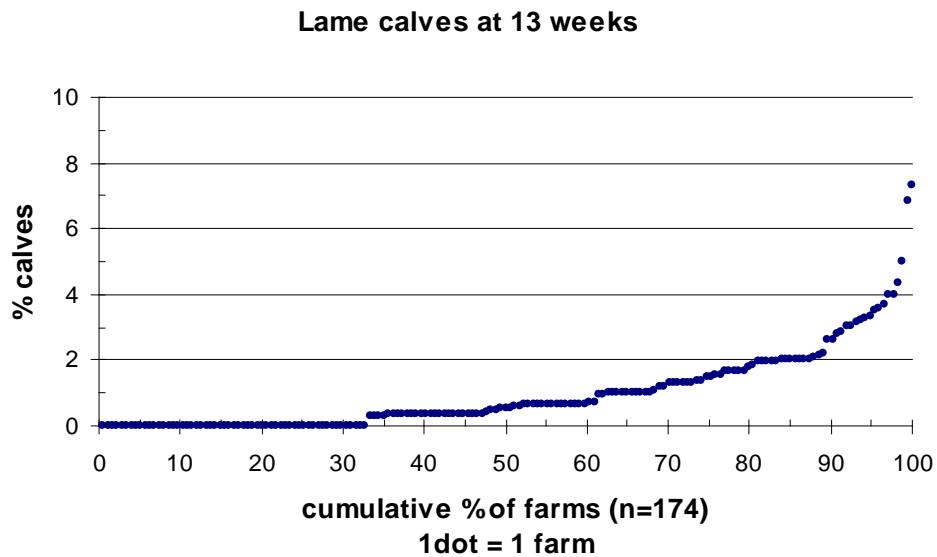


Figure 6.1a: Distribution of the % of lame calves in 174 farms at 13 wks of fattening

Table 6.1b: Dataset submitted to experts regarding bitten tail/ear. They were asked to attribute scores for bitten tail/ear

	% calves with bitten tail/ear	Score
Farm 1	0	
Farm 2	0.5	
Farm 3	1	
Farm 4	1.5	
Farm 5	2	
Farm 6	3	
Farm 7	5	
Farm 8	10	
Farm 9	20	
Farm 10	50	
Farm 11	100	

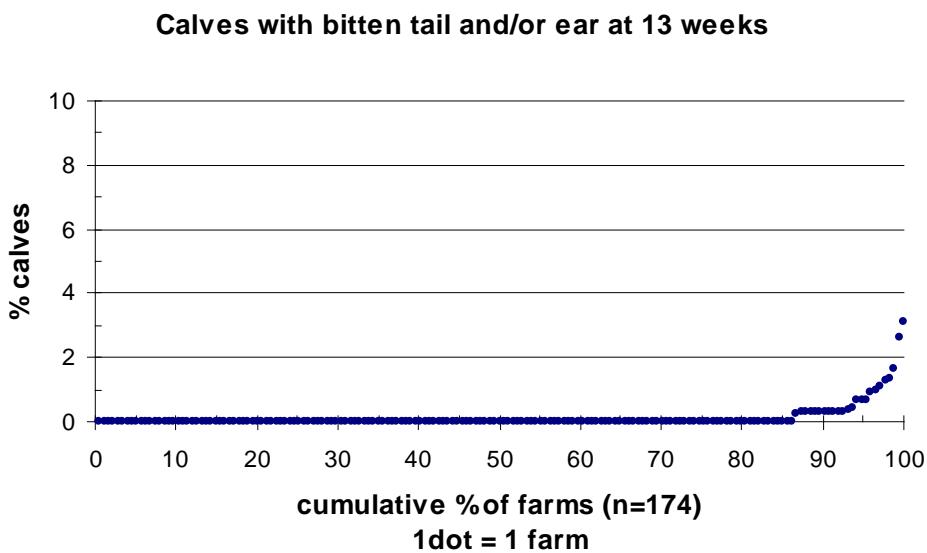


Figure 6.1b: Distribution of the % of calves with bitten tail/ear in 174 farms at 13 wks of fattening

Table 6.2: Dataset submitted to experts. They were asked to attribute criterion scores for absence of injuries according to scores for lameness and scores for bitten tail/ear.

	Score for lameness	Score for bitten tail/ear	Criterion score
Farm 1	40	60	
Farm 2	50	50	
Farm 3	60	40	

- **Adjustment of the construction according to answers from experts**

Calculation of the score for lameness

The results of experts' evaluation of farms with various % of lame calves are shown in Table 6.3a (Columns 2-8). We calculated the mean of the scores attributed by the 7 experts and the index:

$$\text{Index} = 100 - (\% \text{ calves lame}).$$

The function f_a was based on I-Spline functions so as to match best expert scores for lameness (i.e. minimise the sum of square errors with expert scores) (Figure 6.2a). The parameters of these functions are:

Knot(s): 2 interior knot at 80 and 95.

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	0.0000000000
c1	0.0000000000
d1	0.0000159256
a2	-2056.0839946325
b2	77.1031498064
c2	-0.9637893727
d2	0.0040317147
a3	-358368.3363710030
b3	11329.0686710084
c3	-119.4055281565
d3	0.4196167508

Index = I

When $I \leq 80$ then Score = $(0.0000159256 \times I^3)$

When $I \geq 80$ and $I \leq 95$ then Score = $-2056.084 + (77.10315 \times I) - (0.9637894 \times I^2) + (0.0040317147 \times I^3)$

When $I \geq 95$ then Score = $-358368.34 + (11329.06867 \times I) - (119.405528 \times I^2) + (0.4196167508 \times I^3)$

Table 6.3a: Results obtained from the experts asked to attribute scores for lameness

% calves lame	Score for lameness							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	100	100	100	100	100	100	100	100
1	80	90	60	65	70	50	60	68
2	70	80	50	45	50	10	50	51
3	55	50	30	30	35	19	40	37
5	45	40	20	20	20	17	30	27
10	35	30	0	15	10	15	20	18
15	20	20	0	10	0	10	10	10
20	15	20	0	5	0	0	5	6
50	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0

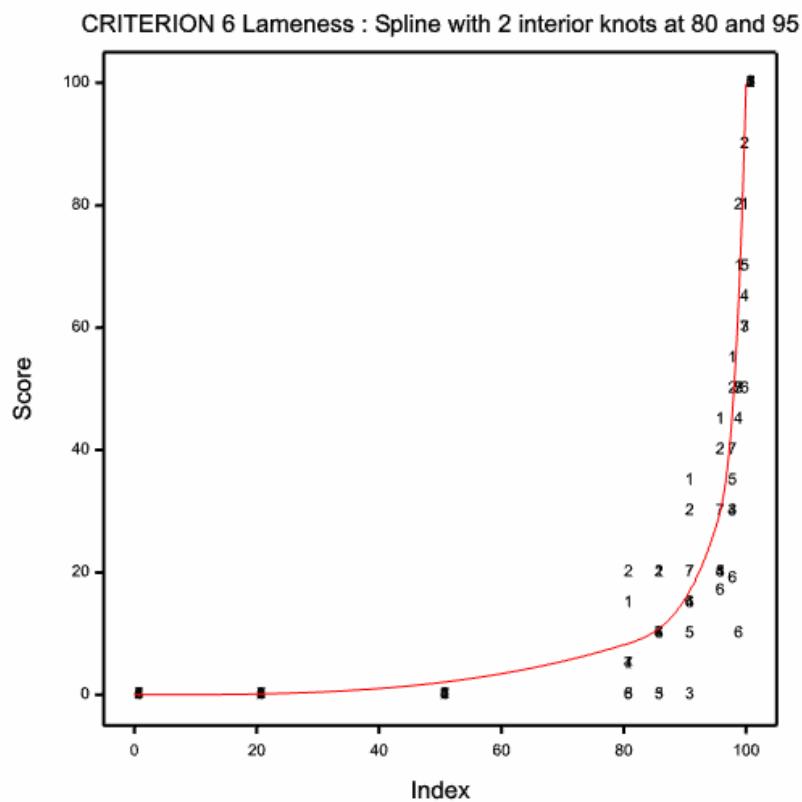


Figure 6.2a: I-Spline function allowing to attribute scores for lameness

Calculation of the score for bitten tail and/or ear

The results of experts' evaluation of % of calves with bitten tail/ear are shown in Table 6.3b.

We calculated the mean of the scores attributed by the 7 experts and the index:

$$\text{Index} = 100 - (\% \text{ calves with bitten tail/ear}).$$

The function f_b was based on I-Spline functions so as to transform the index into a score that matches best expert scores for chewing wounds (i.e. minimise the sum of square errors with expert scores) (Figure 6.2b). The parameters of these functions are:

Knot(s): 2 interior knots at 85 and 95.

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.0000000000
c1		0.0000000000
d1		0.0000070411
a2		-1417.2544574164
b2		50.0207455309
c2		-0.5884793589
d2		0.0023148033
a3		-585600.1816490864
b3		18497.9021184713
c3		-194.7766987807
d3		0.6836769576

Index = I

When $I \leq 85$ then Score = $(0.0000070411 \times I^3)$

When $I \geq 85$ and $I \leq 95$ then Score = $-1417.254 + (50.02075 \times I) - (0.5884794 \times I^2) + (0.0023148033 \times I^3)$

When $I \geq 95$ then Score = $-585600.18 + (18497.9021 \times I) - (194.776699 \times I^2) + (0.6836769576 \times I^3)$

Table 6.3b: Results obtained from the experts asked to attribute scores for bitten tail/ear

% calves with bitten tail/ear	Score for bitten tail/ear							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	100	100	100	100	100	100	100	100.0
0.5	60	90	80	90	60	20	75	68 99.5
1	50	80	60	80	40	15	50	54 99.0
1.5	45	50	40	70	20	10	40	39 98.5
2	40	40	20	60	15	5	35	31 98.0
3	30	30	10	45	10	3	20	21 97.0
5	20	20	0	25	0	0	10	11 95.0
10	15	0	0	15	0	0	0	4 90.0
20	0	0	0	10	0	0	0	1 80.0
50	0	0	0	5	0	0	0	1 50.0
100	0	0	0	0	0	0	0	0 0.0

CRITERION 6 Bitten tail/ear : Spline with 2 interior knots at 85 and 95

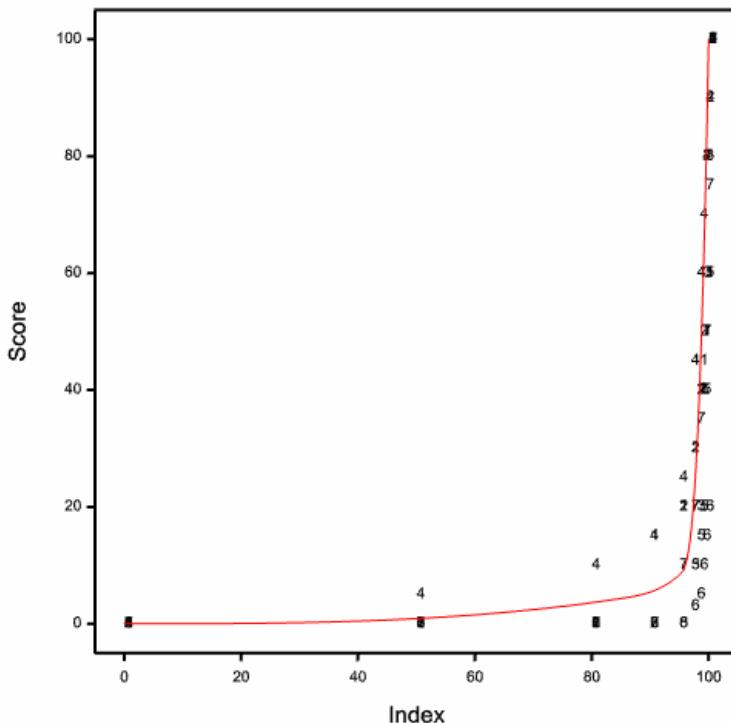


Figure 6.2b: -Spline function allowing to attribute a score for calves with bitten tail/ear

Combination of lameness score and chewing wounds score into a criterion score

A Choquet integral is calculated:

Criterion score (S):

$$\begin{aligned} \text{if } S_a < S_b &\quad \text{then} \quad S = S_a + (S_b - S_a) \times \mu_b \\ \text{else} &\quad \qquad \qquad \qquad S = S_b + (S_a - S_b) \times \mu_a \end{aligned}$$

The parameters of the Choquet integral are determined thanks to a linear optimisation in order to best match the answers of experts (Table 6.4). This led to:

Capacity of the lameness score $\mu_a = 0.35$

Capacity of the score for bitten tail/ear $\mu_b = 0.29$

Sum of capacities $= 0.64$

Indices of importance (Shapley values):

lameness score $\Phi(a) = \frac{1}{2} \times (1 + \mu_a - \mu_b) = 0.35$

score for bitten tail/ear $\Phi(b) = \frac{1}{2} \times (1 + \mu_b - \mu_s) = 0.29$

Interaction index = $1 - \mu_a - \mu_b = 0.36$

The values show that:

- Lameness is slightly more important than bitten tail/ear,
- The interaction between the two scores is positive but limited.

Table 6.4: Estimation of criterion-scores by the 7 experts. Choquet integral is used to calculate scores that best fit with the scores attributed by experts. The last column of the table shows the error between the scores attributed by experts and the calculated score

Expert	Score for	Score for bitten	Criterion score	Calculated	Error ²
	Ilameness	tail/ear	attributed by expert	score	
	1	2			
Expert 1	40	60	60	46	200.02
	50	50	50	50	0.00
	60	40	40	47	49.00
Expert 2	40	60	30	46	251.45
	50	50	40	50	100.00
	60	40	50	47	9.00
Expert 3	40	60	49	46	9.88
	50	50	50	50	0.00
	60	40	52	47	25.00
Expert 4	40	60	40	46	34.31
	50	50	50	50	0.00
	60	40	55	47	64.00
Expert 5	40	60	50	46	17.16
	50	50	50	50	0.00
	60	40	50	47	9.00
Expert 6	40	60	40	46	34.31
	50	50	50	50	0.00
	60	40	40	47	49.00
Expert 7	40	60	52	46	37.73
	50	50	48	50	4.00
	60	40	42	47	25.00
					$\Sigma Error^2$
					919

6.7 Calculation

Two partial scores are calculated, one for lameness, and one for bitten tail/ear, before being combined into a criterion score.

Partial score for lameness: The % of lame calves is used to calculate the index for lameness:

Index = $100 - (\% \text{ calves lame})$.

This index is computed into a score thanks to I-spline functions as follows:

When $I_s \leq 84$ then Score = $(-1.8995 E^{-12} \times I_s) + (0.000794 \times I_s^2) - (0.000005799 \times I_s^3)$

When $I_s \geq 84$ then Score = $-14160.19 + (505.721 \times I_s) - (6.01969 \times I_s^2) + (0.023885 \times I_s^3)$

Partial score for bitten tail/ear: The % of calves with bitten tail/ear is used to calculate an index for bitten tail/ear:

Index = $100 - (\% \text{ calves with bitten tail/ear})$.

This index is computed into a score thanks to I-spline functions as follows:

When $I_l \leq 81$ then Score = $(2.96294 E^{-11} \times I_l) - (1.173575 E^{-12} \times I_l^2) + (1.06108 E^{-14} \times I_l^3)$

When $I_l \geq 81$ then Score = $-624.82 + (26.802 \times I_l) - (0.38288 \times I_l^2) + (0.0018735 \times I_l^3)$

Score for absence of injuries: The two partial scores are combined using a Choquet integral.

The parameters of the Choquet integral are:

$\mu_s=0.35$ and $\mu_l=0.29$

7 Criterion 7: Absence of diseases

7.1 Definition of the criterion

This criterion aims at assessing the health of the animals through the occurrence of diseases (either metabolic diseases or diseases produced by pathogens (virus, microbes, parasites...)).

7.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13112	Coughing	Binary	Individual
13113	Abnormal breathing	Binary	Individual
13114	Nasal discharge	Binary	Individual
13115	Pneumonia	Binary	Individual
13116	Pleuritis	Binary	Individual
13117	Liquid manure	Binary	Individual
13118	Bloated rumen	Binary	Individual
13119	Abomasal lesions	Binary	Individual
13120	Rumen with plaques	Binary	Individual
13121	Haemoglobin	Cardinal	Herd
13122	Dull calves	Binary	Individual
13123	Obviously sick calves	Binary	Individual
13124	Mortality	Cardinal	Herd

7.3 Analysis of the constraints

We see three major constraints:

- 1- Several symptoms can be related to the same disease (i.e. coughing and abnormal breathing).
- 2- The symptoms are not observed with the same time scale (i.e. haemoglobin is taken from farm data of the regular Hb checks while bloated rumen is directly observed during the visit). In addition, the diseases may be experienced more or less aversively by the animals.
- 3- In regards to liquid manure, the measure was at first recorded at pen level (WP 2.4), but as such it showed validity problems due to differences in small and large groups. It has been agreed, therefore, that the measure needs to be at calf level regardless

of the calves being housed in small or large groups. In practice liquid manure will be checked on the pen floor and if present then at individual level.

Alternatively, manure may be checked on the floor only. On farms with small groups, the level of liquid manure is expressed as a % calculated as: no of pens with liquid manure/the total number of pens inspected * 100%. This percentage will then be transformed in a qualitative score at farm level of either 0 (no liquid manure), 1 (moderate liquid manure), and 2 (serious problem with liquid manure). On farms with large groups, based on the surface covered with liquid manure, a qualitative score is assigned at farm level of either 0, 1 or 2. A qualitative score of 1 refers to a Warning (see below), and a qualitative score of 2 refers to an Alarm (see below).

7.4 Type of mathematical construction proposed

To take into account the first two constraints, we decided to:

1- group the symptoms that can be related to the same disease. This brought 4 categories:

- a) Coughing, Abnormal breathing, Nasal discharge, Pneumonia, Pleuritis (respiratory disease)
- b) Liquid manure, Bloated rumen, Abomasal lesions, Rumen plaques (digestive disorders)
- c) Haemoglobin (anemia)
- d) Dull calves, Obviously sick calves, Mortality (general)

Note that dull calves and mortality are taken apart because numerous causes are possible. Haemoglobin will be taken from regular on-farm haemoglobin check data as % of calves with Hb below 4.5 mmol/L at 13 weeks of age. Pneumonia, pleuritis, abomasal and rumen lesions are detected at slaughter but related to the farm.

2- define for each symptom an 'alarm' threshold above which it is considered that there is a serious problem on the farm (i.e. the prevalence of the disorder is such that a vet would recommend that a remedial measure is taken at farm level). Since such a threshold would not give variable results (a farm would probably already obtain a low welfare score when the alarm threshold is over passed once), we decided to use another threshold called 'warning' threshold equivalent to half of the alarm threshold.

We decided not to consider the severity of a symptom for an animal because this would have been too subjective.

The construction proposed is the following:

- for each of the 4 categories of symptoms (a, b, c, d as defined above), we consider the worst situation brought by the symptoms in that category:
 if the frequency of at least one symptom is above the alarm threshold, we consider that there is an alarm for the category
 if there is no alarm but at least one warning, then a warning is attributed to the category.
- we calculate the total number of warnings (N_1) and of alarms (N_2) (with $(N_1 + N_2)$ ranging between 0 and 4)
- we calculate a weighted sum attributing a higher weight to p_2 than to p_1 . The following formula allows the score to range between 0 and 100 with the worst situation set at 0:

$$I = \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^2 w_k N_k}{w_2} \right) \right)$$

= w_1 , weight attributed to N_1 and w_2 , weight attributed to N_2
 N_{tot} = maximum number of warnings or alarms (here 4)

- experts are generally more severe at the highest portion of the scale. Hence some transformation of this index may be necessary to better match to expert opinion:

$$\text{Criterion score} = f \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^2 w_k N_k}{w_2} \right) \right)$$

7.5 Parameters to be defined

03.

We need to define:

- the alarm threshold for each symptom,
- the weights w_1 and w_2 to be used in the weighted sum,
- the most appropriate function f to transform the index derived from the weighted sum into a criterion score.

7.6 Adjustments of the mathematical construction

▪ Questions asked to experts

▪

Questions to define thresholds for each symptom

Field vets were asked the limit beyond which they consider a remedial measure shall be taken on the farm for each disorder. The vets involved during 2007-2008 were two from Italy, one from France and a group of Dutch field vets who had a specific meeting during which they exchanged their views. Moreover during a meeting held in Paris on the 24th November 2010, 3 more field vets were consulted (one per each country).

The alarm threshold for a given symptom was then calculated as the weighted mean of the experts' answers. Vets from each country counted for one in the weighted sum, while the group of Dutch vets counted for 2.

The warning threshold was set at half the alarm threshold.

Measure ¹	Alarm threshold
Coughing	
Abnormal breathing	
Nasal discharge	
Pneumonia	
Pleuritis	
Liquid manure	
Bloated rumen	
Abomasal lesions	
Rumen with plaques	
Haemoglobin	
Dull calves	
Obviously sick calves	
Mortality	

¹% of calves affected

In regards to the 'alarm' threshold for anemia it has been proposed to consider the initial haemoglobin level of the calves (first Hb check) as follows:

Initial (i.e. 3 weeks after arrival)	During visit (i.e. 13 weeks after arrival)
% of calves with Hb below 4.5mmol/L	% of calves with Hb below 4.5mmol/L
< 10	alarm at 9
10-15	alarm at 11
15-20	alarm at 13
> 20	alarm at 15

and if not controlled then the most severe case is applicable.

As far as abomasal lesions are concerned, for now the % of abomasal with one or more lesions is considered. This will be revised in that the lesion score (which takes into consideration both the number and size of abomasal lesions, and is, therefore, indicative of the severity) will be used in the final monitoring system. Prior to inclusion of the lesion score into the system, the following steps have to be taken: (i) Similar to for example Hb, a threshold has to be defined which determines, at the level of the individual calf, whether or not the level of lesions should be considered severe. (ii) Using this threshold, the % of calves in each batch with a lesion score higher than the threshold is determined. (iii) using expert consultation, Alarm and Warning thresholds have to be defined for the % of calves in a batch with a lesions score higher than the overall threshold.

Questions to define the weights w_1 and w_2 and function f

Combinations of number of warnings and alarms (Table 7.1) were proposed to experts who were asked

- to sort the farm from best to worst,
- then to attribute a criterion score between 0 and 100.

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen. The frequency of warnings (Figure 7.1a) and alarms (Figure 7.1b) and the distribution of the farms according to the number of warnings and alarms (Figure 7.2) in the dataset gathered during WP 2.4 ($n = 50$ farms for which the dataset was complete including haemoglobin data) were also shown to the experts.

Table 7.1: Dataset submitted to experts. They had to sort the farm from best to worst then to attribute criterion scores

		Moderate problem (Warning)	Serious problem (Alarm)	Rank	Score
	Normal				
Farm 1	4	0	0		
Farm 2	0	4	0		
Farm 3	0	0	4		
Farm 4	2	2	0		
Farm 5	2	0	2		
Farm 6	0	2	2		
Farm 7	3	1	0		
Farm 8	3	0	1		
Farm 9	0	3	1		
Farm 10	0	1	3		
Farm 11	1	0	3		
Farm 12	1	3	0		

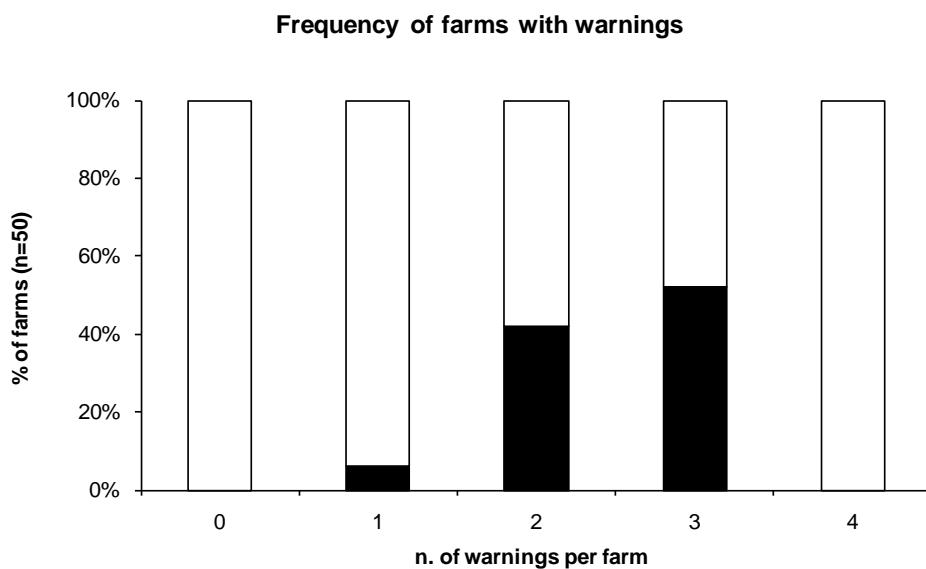


Figure 7.1a: Frequency of farms according to the number of warnings at 13 wks of fattening

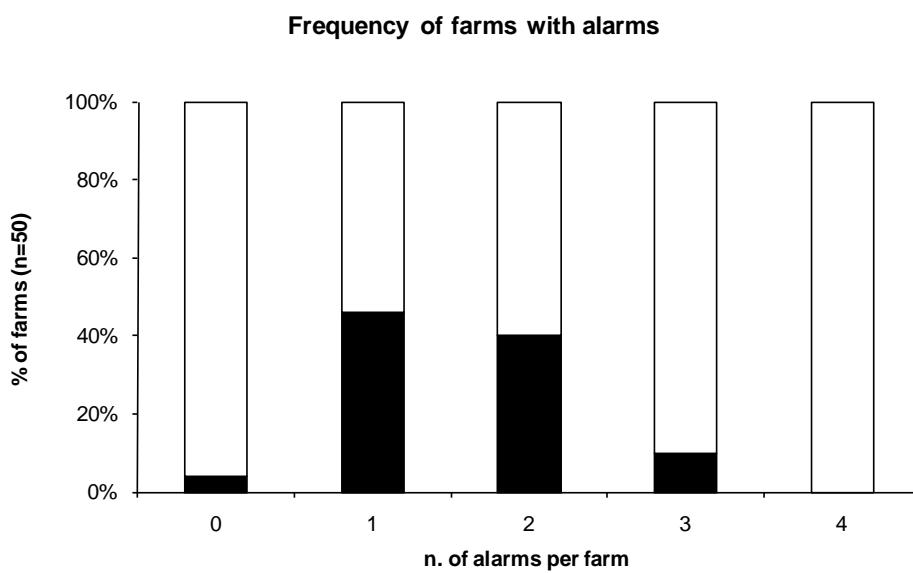


Figure 7.1b: Frequency of farms according to the number of alarms at 13 wks of fattening

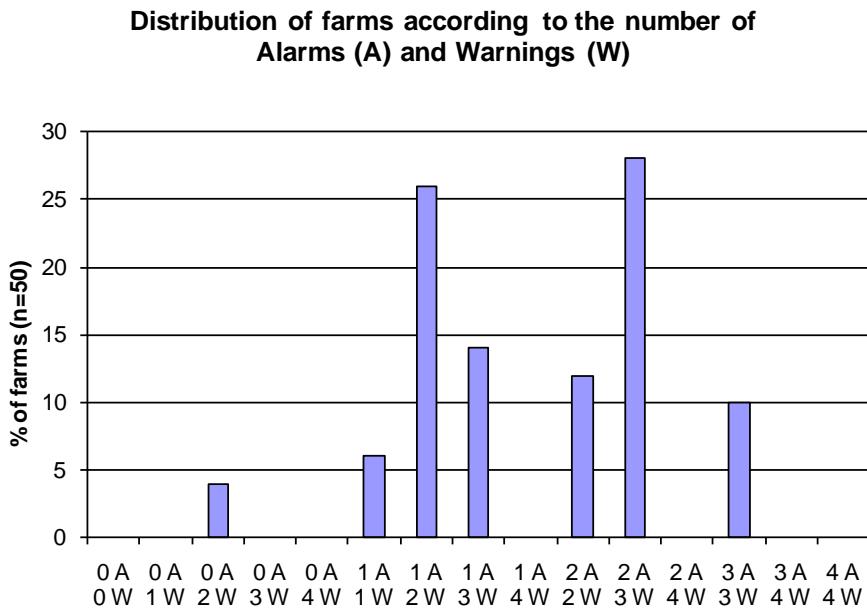


Figure 7.2: Distribution of farms according to the number of alarms (A) and warnings (W) at 13 wks of fattening

- **Adjustment of the construction according to answers from experts**

Warning and alarm thresholds

The field veterinarians' answers and the thresholds decided for each symptom are shown in Table 7.2.

Table 7.2: Thresholds proposed by field vets and final alarm and warning thresholds decided.

Measure ¹	consultation (24/11/2010)			previous consultation (2007, 2008)				Alarm	Warning
	Vet	Vet	Vet	Vet	Vet	Vet	Vets		
	NL	IT	FR	IT	IT	FR	NL		
Coughing	7	12	10	5	5	10	20	11.1	5.6
Abnormal breathing	3	7	2	8	5	5	10	6.3	3.1
Pneumonia	20	25	10				25	21.0	10.5
Pleuritis	40	40	3				10	20.6	10.3
Nasal discharge	10	13	12	5	5	7.5	25	12.8	6.4
Liquid manure	15	20	5				10	12.0	6.0
Abomasal lesions	90	80	10				25	46.0	23.0
Plaque in rumen	70	50	10				25	36.0	18.0
Bloated rumen	5	10	4	3	3	7.5	20	9.1	4.5
Obviously sick calves	1	10	4				10	7.0	3.5
Dull calves	3	6	4	3	5	5	10	5.8	2.9
Mortality	4	5	7	6	5	2.5	5	4.9	2.5
Haemoglobin < 4.5 mmol/l	5	10	4					6.3	3.2

¹ some measures were added between the two consultations of field vets so experts did not answer all questions.

Weights and function f

The results of experts' evaluation for the combinations of warnings and alarms are shown in Table 7.3 (Columns 5-10). We calculated the mean of criterion-scores attributed by the 6 experts (one expert did not fill the table due to misunderstanding of the question and to disagreement with the three options for normal, moderate and serious problem); then the farms were sorted according to the mean criterion-score (Table 7.3, Columns 11,12).

The weights W_1 and W_2 were determined by linear optimisation so as to match the ranking of farms according to the mean score attributed by experts:

$$W_1 = 1$$

$$W_2 = 3$$

The results of the index resulting from the weighted sum obtained by each combination are shown in Table 7.3, Column 13.

Finally I-Spline functions were defined so as to match best expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 7.3). The parameters of these functions are:

Knot (s): 1 inferior knot set at 25

Values of coefficients:

COEF	VALUE
a_from_knot	-2.0799416386728766248381817
b1_before_knot	0.1636973944008163761587582
b2_before_knot	0.0153189453829875536078342
b3_before_knot	-0.0001257176175149784852405
b1_from_knot	0.4132903910438945893623952
b2_from_knot	0.0053352255172212760883643
b3_from_knot	0.0000073986473620672038587

Table 7.3: Results obtained from the experts in regards to the number of normal situations (N), warnings (W), and alarms (A) present in the virtual farms

Farm	Problems on farm			Score						Rank	Index	
	N	W	A	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6			
Farm 3	0	0	4	0	0	5	0	0	0	0.8	12	0.0
Farm 10	0	1	3	20	0	10	0	0	10	6.7	11	16.7
Farm 11	1	0	3	30	0	20	0	0	10	10	10	25.0
Farm 6	0	2	2	40	20	25	0	20	15	20	9	33.3
Farm 9	0	3	1	30	50	35	30	30	20	32.5	8	50.0
Farm 5	2	0	2	50	40	40	20	20	30	33.3	7	50.0
Farm 2	0	4	0	50	60	45	60	40	45	50	6	66.7
Farm 8	3	0	1	80	55	75	60	35	50	59.2	5	75.0
Farm 12	1	3	0	70	70	55	70	45	70	63.3	4	75.0
Farm 4	2	2	0	80	80	80	80	50	85	75.8	3	83.3
Farm 7	3	1	0	90	90	90	90	70	95	87.5	2	91.7
Farm 1	4	0	0	100	100	100	100	100	100	100	1	100.0

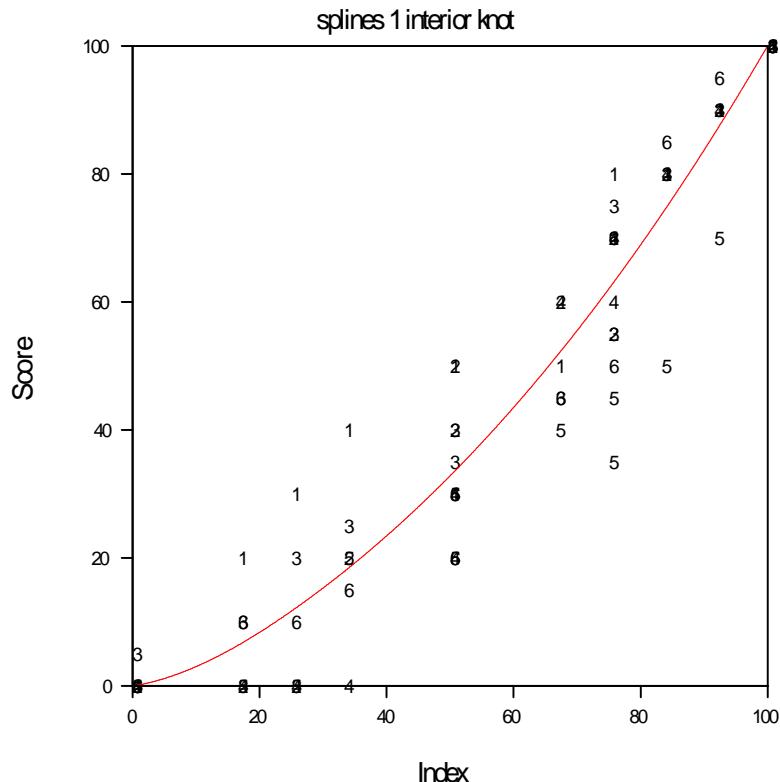


Figure 7.3: I-Spline function allowing to attribute a health score in regards to the number of warnings and alarms for the various symptoms recorded.
X axis, index derived from the weighted sum; Y axis, score for the absence of diseases

7.7 Calculation

The frequency of symptoms (expressed as % of calves affected) is compared to warning and alarm thresholds as follows:

Measure ¹	Alarm threshold	Warning threshold
Coughing	11.1	5.6
Abnormal breathing	6.3	3.1
Pneumonia	21.0	10.5
Pleuritis	20.6	10.3
Nasal discharge	12.8	6.4
Liquid manure	12.0	6.0
Abomasal lesions	46.0	23.0
Plaque in rumen	36.0	18.0
Bloated rumen	9.1	4.5
Obviously sick calves	7.0	3.5
Dull calves	5.8	2.9
Mortality	4.9	2.5
Haemoglobin < 4.5 mmol/l	6.3	3.2

¹ % of calves affected

Symptoms are grouped areas

- Coughing, Abnormal breathing, Nasal discharge, Pneumonia, Pleuritis (respiratory disease)
- Liquid manure, Bloated rumen, Abomasal lesions, Rumen plaques (digestive disorders)
- Haemoglobin (anemia)
- Dull calves, Obviously sick calves, Mortality (general)

total number of area = 4

The severity of problems is estimated per area:

- if in an area the frequency of one symptom is above the warning threshold and the others are below the alarm threshold, a warning is attributed to the area
- if in an area the frequency of one symptom is above the alarm threshold, then an alarm is attributed to the area
- else no problem is recorded

An index is calculated as

$$I = \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^2 w_k N_k}{w_2} \right) \right)$$

with N_{tot} = number of areas = 4
 N_k number of problems according to intensity
 $k=0$ no problems
 $k=1$ warnings
 $k=2$ alarms
 w_k weight of problems
 $w_0=0$
 $w_1=1$
 $w_2=3$

Then the index is transformed into a score according to I-spline functions as follows :

COEF	VALUE
a_from_25	-2.0799416386728766248381817
b1_before_25	0.1636973944008163761587582
b2_before_25	0.0153189453829875536078342
b3_before_25	-0.0001257176175149784852405
b1_from_25	0.4132903910438945893623952
b2_from_25	0.0053352255172212760883643
b3_from_25	0.0000073986473620672038587

When $I \leq 25$ then Score = $(0.163697 \times I) + (0.0153189 \times I^2) - (0.0001257 \times I^3)$

When $I \geq 25$ then Score = $-2.07994 + (0.41329 \times I) + (0.005335 \times I^2) + (0.0000073986 \times I^3)$

8 Criterion 8: Absence of pain induced by management procedures

8.1 Definition of the criterion

This criterion aims at assessing the pain due to management procedures like mutilations or stunning. The pain which is here assessed is not due to an injury or a disease, but to human intervention.

8.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13125	Routine mutilations (tail docking)	Qualitative	Herd

Veal calves are neither dehorned nor castrated. They can have their tail docked but this is rare. The construction of this criterion will be similar to that of dairy cows regarding tail docking.

8.3 Analysis of the constraints

See deliverable D2.8b Part 2 (dairy cows)

8.4 Type of mathematical construction proposed

If no mutilation: score = 100

See deliverable D2.8b Part 2 (dairy cows) for scores attributed to tail docking procedures. A decision tree will be followed (Figure 8.1)



Figure 8.1: Combinations of conditions for tail docking

8.5 Parameters to be defined

We need to attribute a score to each situation from $y_{2,1}$ to $y_{2,9}$

8.6 Adjustments of the mathematical construction

- **Questions asked to experts**

See deliverable D2.8b Part 2 (dairy cows)

- **Adjustment of the construction according to answers from experts**

See deliverable D2.8b Part 2 (dairy cows)

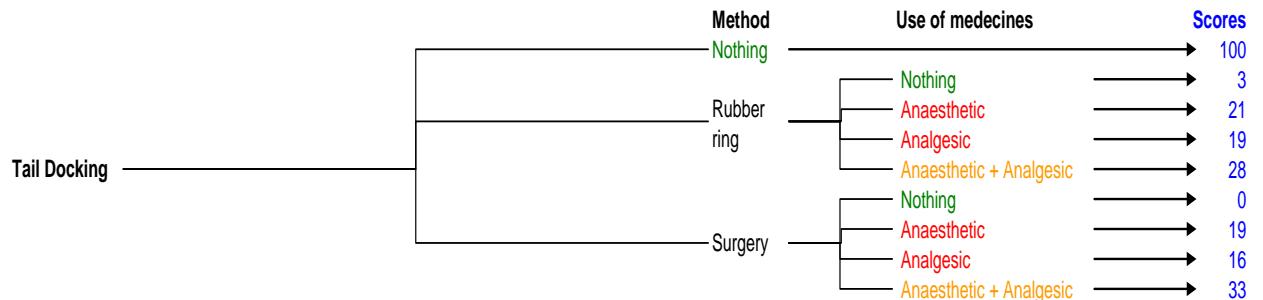
The scores attributed to $y_{2,1}$ to $y_{2,9}$ are given in Table 8.1.

Table 8.1. Average scores attributed by experts

Situation	Score
$y_{2,1}$	100
$y_{2,2}$	3
$y_{2,3}$	21
$y_{2,4}$	19
$y_{2,5}$	28
$y_{2,6}$	0
$y_{2,7}$	19
$y_{2,8}$	16
$y_{2,9}$	33

8.7 Calculation

The score for absence of pain due to management procedures is attributed according to whether and how animals are tail docked. A decision tree is used to attribute scores:



9 Criterion 9: Expression of social behaviours

9.1 Definition of the criterion

In this criterion, we aim at assessing the expression of social behaviours.

9.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13126	Social licking	Cardinal	Group

The calves are observed for 3 times 30 min by scan sampling, with 2 min between scans (total 15 scans per group per each observation). The data correspond to the frequency of a given behaviour divided by the number of calves observed and scans performed on the same session. The average of the three sessions is taken and data are expressed as percentage of calves performing social licking (= % of observations in which a calf is observed performing this behaviour).

9.3 Analysis of the constraints

There is only one measure in this criterion. Social liking is indicative of preferential relationships, however the frequency of this behaviour among calves should not be either too high (indicative of frustration) nor too low (indicative of lack of interests towards pen-mates) for a good functioning of the group of calves.

9.4 Type of mathematical construction proposed

First a calculation of an index is needed based on the percentage of observations when social liking is observed:

$$\text{Index} = 100 - (\% \text{ calves performing social licking}).$$

Then to use a function f to match expert scores:

$$\text{Criterion score} = f((\mathbf{w}_1 p_1 + \mathbf{w}_2 p_2) / (\max[\mathbf{w}_1, \mathbf{w}_2]))$$

9.5 Parameters to be defined

We need:

- to define an appropriate function f to transform the index into the criterion score.

9.6 Adjustments of the mathematical construction

▪ Questions asked to experts

A dataset of farms with various % of social licking (Table 9.1) was shown to experts

They were asked:

- to attribute a criterion score between 0 and 100 to each farm considering the distribution of the frequency of social licking on real farms (distribution, average and maximum) found from WP2.4 partners in 157 farms at about 13 weeks of age of the calves (Figure 9.1).

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

Table 9.1: Dataset submitted to experts. They had to attribute criterion scores

	frequency of social licking (% of calves)	Score
Farm 1	0	
Farm 2	1	
Farm 3	1.5	
Farm 4	2	
Farm 5	2.5	
Farm 6	3	
Farm 7	5	
Farm 8	7	
Farm 9	10	
Farm 10	20	
Farm 11	100	

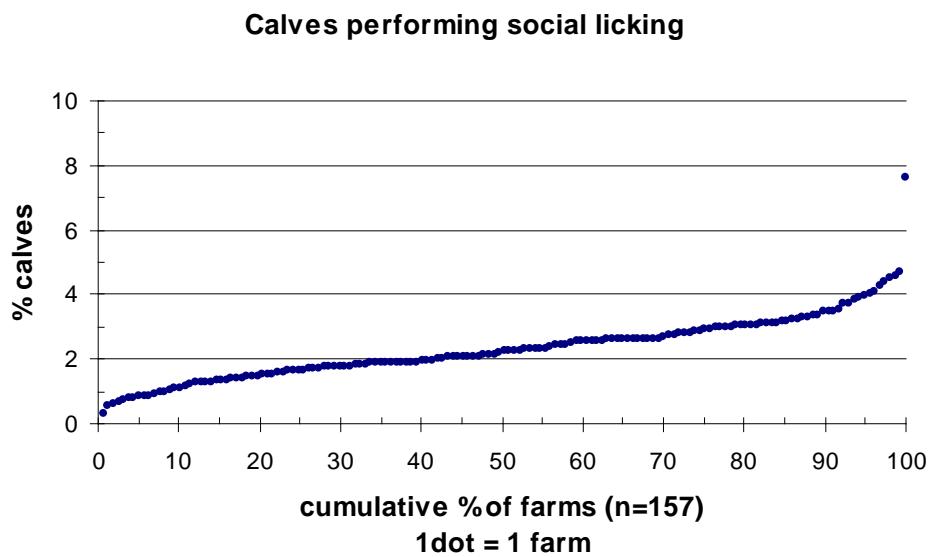


Figure 9.1: Distribution of % of calves performing social licking in 157 farms at 13 wks of age of the calves

- **Adjustment of the construction according to answers from experts**

The results of experts' evaluation of farms with various % of social licking are shown in Table 9.2 (Columns 2-8). We calculated the mean of criterion-scores attributed by the 7 experts Table 9.2 (Columns 9).

Finally B-spline functions were defined so as to match best expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 9.2). The parameters of these spline functions are:

Knot(s): 2 interior knots set at 80 and 95.

Coefficients

COEF	VALUE
a1	0.0000000000
b1	191.1188389194
c1	-4.6354179487
d1	0.0281265790
a2	53999.6689546353
b2	-1833.8687464888
c2	20.6769268643
d2	-0.0773415243
a3	318310.3448266463
b3	-10180.5217097207
c3	108.5364321630
d3	-0.3856204918

Table 9.2: Results obtained from the experts asked

frequency of social licking (% calves)	Score for social licking							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	0	0	0	10	0	20	0	4 100
1	50	0	20	50	5	50	20	28 99
1.5	75	0	40	75	10	50	40	41 99
2	100	20	50	95	15	50	60	56 98
2.5	100	50	60	100	20	50	80	66 98
3	100	60	80	95	30	50	100	74 97
5	75	80	100	80	50	100	80	81 95
7	50	90	100	60	65	50	50	66 93
10	40	100	80	25	100	20	20	55 90
20	20	50	20	15	60	0	0	24 80
100	0	0	0	0	40	0	0	6 0

In the case of Expert 5, the score of 40 in the last row was changed into 0.

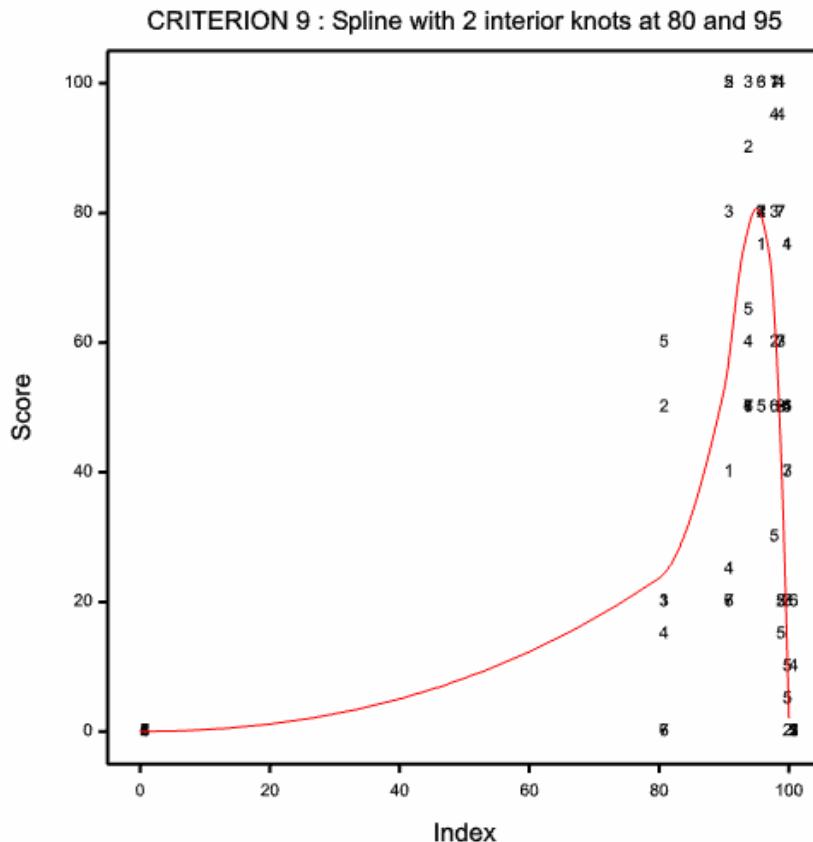


Figure 9.2: B-spline functions allowing to attribute a score for the expression of social behaviour according to the frequency of social licking

9.7 Calculation

The % of calves performing social licking is used to calculate an index:

Index I = (100 - % of calves performing social licking)

Index I is then transformed into a score thanks to B-spline functions:

When $I \leq 80$ then Score = $(191.1188 \times I) - (4.635418 \times I^2) + (0.0281265790 \times I^3)$
 When $I \geq 80$ and $I \leq 95$ then Score = $53999.67 - (1833.86875 \times I) + (20.6769269 \times I^2) - (0.0773415243 \times I^3)$
 When $I \geq 95$ then Score = $318310.34 - (10180.52171 \times I) + (108.5364322 \times I^2) - (0.3856204918 \times I^3)$

10 Criterion 10: Expression of other behaviours

10.1 Definition of the criterion

In this criterion, we aim to assess the expression of abnormal oral behaviours and cross sucking/urine drinking signs.

10.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13127	Abnormal oral behaviours	Cardinal	Group
13128	Urine drinking signs	Binary	Individual

The measures are defined as:

Abnormal oral behaviours defined as frequency of tongue playing/rolling and oral manipulation of a substrate:

- Tongue playing/rolling = calf makes repeated movements with the tongue in or outside the mouth
- Oral manipulating substrate = Calf licks, nibbles, sucks, and bites at the fence, wall, bucket, trough, floor or object not specific in a stereotypical way. An object is each attribute which is added in the pens without any use (toys).

Prevalence of calves with signs of urine drinking (suckers and victims):

- Urine sucker = calf with wet and/or yellow snout.
- Urine suckers' victim = calf with swollen, hairless prepuce with red skin, wet belly button.

The assessment is performed in two ways:

Abnormal oral behaviours are assessed through behavioural observations where calves are observed for 3 times 30 min by scan sampling, with one scan every 2 min (total 15 scans per group per each observation). The data correspond to the frequency of the behaviours divided by the number of calves observed and scans performed on the same session. The average of the three sessions is taken and data are expressed as proportion of calves performing abnormal oral behaviours (tongue playing/rolling and/or oral manipulating substrate).

Prevalence of calves with signs of urine drinking (suckers and victims) is assessed during the clinical/health visit. There is no behavioural measure since it would assess only suckers and

therefore underestimate the problem. The data correspond to the % of calves showing signs of urine drinking/cross-sucking.

10.3 Analysis of the constraints

The two measures included in this criterion have both a negative meaning and there is no measure reflecting positive non-social behaviours such as i.e. play behaviours. Moreover, they are assessed in different ways.

10.4 Type of mathematical construction proposed

We propose to reason in two stages

First, two elementary scores are attributed:

- a score for % of calves performing abnormal oral behaviours:

$$S_a = f_a(p_a) \quad \text{with } p_a, \% \text{ abnormal oral behaviours}$$

- a score for % of calves with urine drinking signs:

$$S_b = f_b(p_b) \quad \text{with } p_b, \% \text{ calves with urine drinking signs}$$

Second, these two elementary scores are brought together to form the criterion-score for expression of other behaviours:

$$\text{Criterion score} = f(S_a, S_b)$$

10.5 Parameters to be defined

We need to define functions f_a , f_b , and f .

10.6 Adjustments of the mathematical construction

- **Questions asked to experts**

Three sets of data were shown to experts:

- one for abnormal oral behaviours, with % of calves performing abnormal oral behaviours (Table 10.1a),
- one for urine drinking, with % of calves showing signs of urine drinking (Table 10.1b),
- one for the combination of the two elementary scores (Table 10.2).

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

They were asked:

- to attribute scores for abnormal oral behaviours considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 157 farms at about 13 weeks of age of the calves (Figure 10.1a).
- to attribute scores for urine drinking considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at 13 weeks of fattening (Figure 10.1b).
- to attribute a criterion-score for expression of other behaviours to the combinations of the two elementary scores.

Table 10.1a: Dataset submitted to experts. They were asked to attribute scores according to the % of calves performing abnormal oral behaviours.

	% calves performing abnormal oral behaviours	Score
Farm 1	0	100
Farm 2	5	30
Farm 3	10	20
Farm 4	15	0
Farm 5	20	0
Farm 6	30	0
Farm 7	40	0
Farm 8	50	0
Farm 9	70	0
Farm 10	85	0
Farm 11	100	0

**Calves performing abnormal oral behaviours:
tongue playing/rolling and manipulating a substrate**

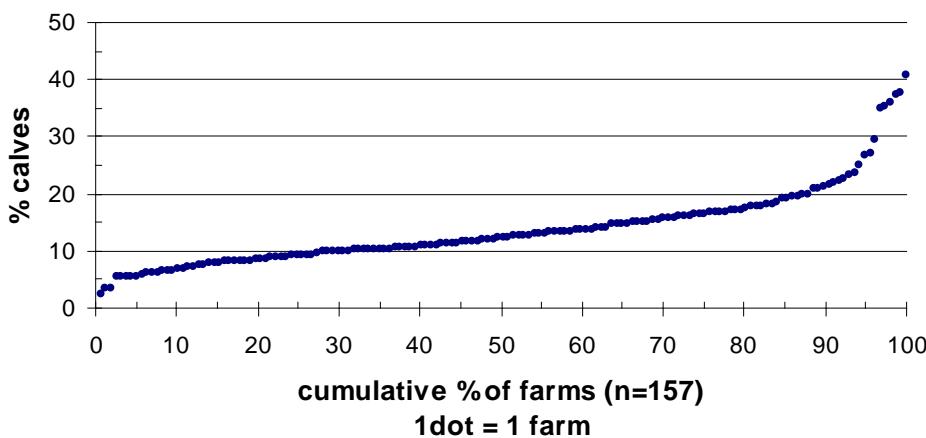


Figure 10.1a: Distribution of % of calves performing abnormal oral behaviours in 157 farms at 13 wks of age of the calves

Table 10.1b: Dataset submitted to experts. They were asked to attribute scores for urine drinking according to the % of calves showing signs of urine drinking/cross-sucking.

	% calves with signs of urine drinking	Score
Farm 1	0	100
Farm 2	0.5	80
Farm 3	1	70
Farm 4	1.5	60
Farm 5	2	50
Farm 6	3	40
Farm 7	4	30
Farm 8	5	20
Farm 9	10	0
Farm 10	50	0
Farm 11	100	0

**Calves with signs of urine drinking at 13 weeks:
drinkers and victims**

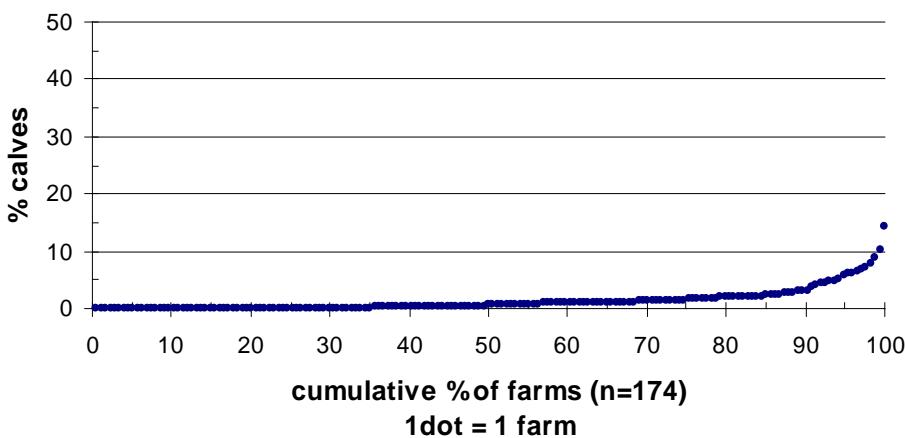


Figure 10.1b: Distribution of % of calves with signs of urine drinking in 174 farms at 13 wks of fattening

Table 10.2: Dataset submitted to experts. They were asked to attribute criterion scores for the expression of other behaviours

	Score for abnormal oral behaviours	Score for urine drinking	Score
Farm 1	40	60	
Farm 2	50	50	
Farm 3	60	40	

- **Adjustment of the construction according to answers by experts**

Calculation of the score for abnormal oral behaviours

The results of experts' evaluation of farms with various % of calves performing abnormal oral behaviours are shown in Table 10.3a.

First the following index is calculated:

$$\text{Index}_a = 100 - (\% \text{ of calves performing abnormal oral behaviours})$$

The function f_a was based on I-Spline functions so as to transform the index into a score that best matches expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 10.2a). The parameters of the spline functions were:

Knot(s): 1 interior knot at 60

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	0.0000000000
c1	0.0000000000
d1	0.0000220495
a2	-263.0830665233
b2	13.1541533260
c2	-0.2192358888
d2	0.0012400266

Index = I

When $I \leq 60$ then Score = $0.0000220495 \times I^3$

When $I \geq 60$ then Score = $-263.083 + (13.15415 \times I) - (0.2192359 \times I^2) + (0.0012400266 \times I^3)$

Table 10.3a: Results obtained from the experts asked to attribute scores for abnormal oral behaviours

% calves performing abnormal oral behaviours	Score for abnormal oral behaviours							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	100	100	100	100	100	100	100	100
5	30	90	75	80	75	50	80	69
10	20	70	50	70	50	40	50	50
15	0	50	25	60	35	20	45	34
20	0	40	10	25	20	15	40	21
30	0	20	0	15	0	10	20	9
40	0	0	0	10	0	5	10	4
50	0	0	0	5	0	0	0	1
70	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0

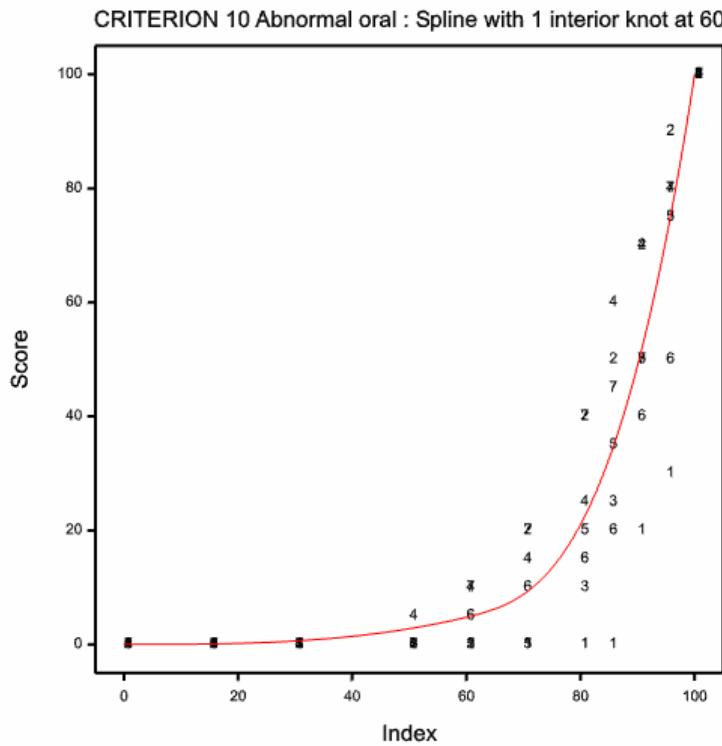


Figure 10.2a: I-Spline function allowing to attribute a score for abnormal oral behaviours

Calculation of the score for urine drinking

The results of experts' evaluation of farms with various % of calves with signs of urine drinking are shown in Table 10.3b.

First the following index is calculated:

$$\text{Index}_b = 100 - (\% \text{ of calves with signs of urine drinking})$$

The function f_b was based on I-Spline functions so as to match best expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 10.2b). The parameters of these functions are:

Knot(s): 2 interior knots at 90 and 95.

Coefficients:

COEF	VALUE
a1	0.0000000000
b1	0.0000000000
c1	0.0000000000
d1	0.0000125445
a2	-45640.8557121555
b2	1521.3618547246
c2	-16.9040205824
d2	0.0626200281
a3	-216073.3635558901
b3	6903.4409516242
c3	-73.5574837462
d3	0.2614041059

Index = I

When $I \leq 90$ then Score = $0.0000125445 \times I^3$

When $I \geq 90$ and $I \leq 95$ then Score = $-45640.86 + (1521.36185 \times I) - (16.9040206 \times I^2) + (0.0626200281 \times I^3)$

When $I \geq 95$ then Score = $-216073.36 + (6903.44095 \times I) - (73.5574837 \times I^2) + (0.2614041059 \times I^3)$

Table 10.3b: Results obtained from the experts asked to attribute scores for urine drinking

% calves with signs of urine drinking	Score for urine drinking							Mean	Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7		
0	100	100	100	100	100	100	100	100	100
0.5	80	90	90	80	80	70	90	83	99.5
1	70	70	80	75	60	50	80	69	99
1.5	60	50	60	60	40	45	70	55	98.5
2	50	40	50	55	35	40	60	47	98
3	40	30	40	45	30	30	50	38	97
4	30	30	30	35	25	25	30	29	96
5	20	20	20	25	20	20	20	21	95
10	0	0	0	15	10	5	0	4	90
50	0	0	0	0	0	0	0	0	50
100	0	0	0	0	0	0	0	0	0

CRITERION 10 Urine drinking : Spline with 2 interior knots at 90 and 95

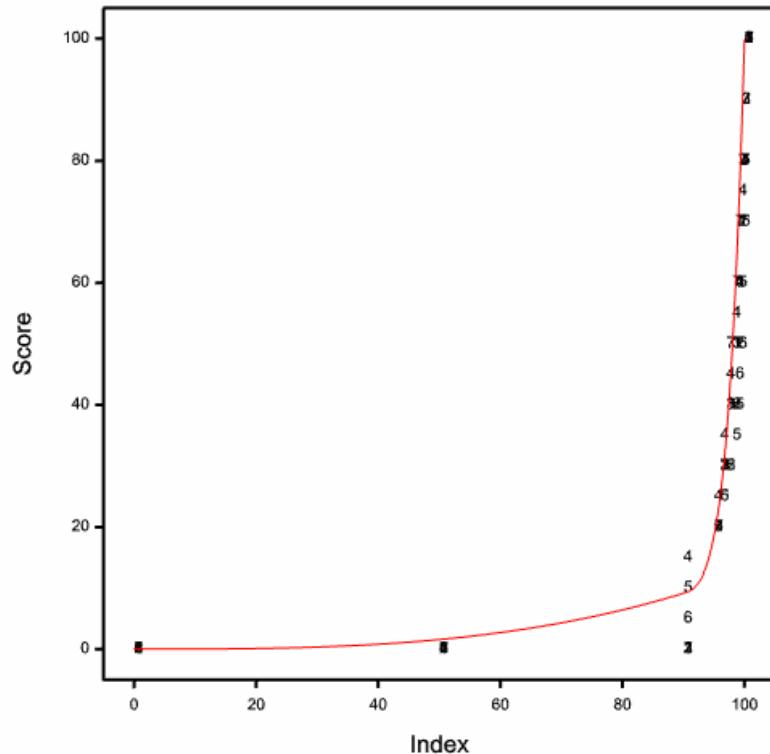


Figure 10.2b: I-Spline function allowing to attribute a score for urine drinking

Combination of the scores for abnormal oral behaviours and urine drinking into criterion score for expression of other behaviours

A Choquet integral is calculated:

Criterion score (S)

$$\begin{array}{ll} \text{if } S_a < S_b & \text{then } S = S_a + (S_b - S_a) \times \mu_b \\ \text{else} & S = S_b + (S_a - S_b) \times \mu_a \end{array}$$

The parameters of the Choquet integral are determined thanks to a linear optimisation in order to best match the answers of experts (Table 10.4). This led to:

Capacity of the score for abnormal oral behaviour $\mu_a = 0.11$

Capacity of the score for urine drinking $\mu_b = 0.61$

Sum of capacities $= 0.77$

Indices of importance (Shapley values):

$$\text{score for abnormal oral behaviour} \quad \Phi(a) = \frac{1}{2} \times (1 + \mu_a - \mu_b) = 0.11$$

$$\text{score for urine drinking} \quad \Phi(b) = \frac{1}{2} \times (1 + \mu_b - \mu_a) = 0.61$$

$$\text{Interaction index} = 1 - \mu_a - \mu_b = 0.23$$

The values show that:

- Urine drinking is more than five times as important as abnormal oral behaviour,
- The interaction between the two scores is positive but limited.

Table 10.4: Estimation of criterion-scores by the 7 experts. Choquet integral is used to calculate scores that best fit with the scores attributed by experts. The last column of the table shows the error between the scores attributed by an expert and the calculated score

Expert	Score for abnormal oral behaviours	Score for urine drinkinr	Criterion score attributed by expert	Calculated score	Error ²
	1	2			
Expert 1	40	60	50	52	4.59
	50	50	40	50	100.00
	60	40	30	42	147.45
Expert 2	40	60	60	52	61.73
	50	50	50	50	0.00
	60	40	40	42	4.59
Expert 3	40	60	50	52	4.59
	50	50	50	50	0.00
	60	40	50	42	61.73
Expert 4	40	60	55	52	8.16
	50	50	50	50	0.00
	60	40	45	42	8.16
Expert 5	40	60	55	52	8.16
	50	50	50	50	0.00
	60	40	45	42	8.16
Expert 6	40	60	40	52	147.45
	50	50	50	50	0.00
	60	40	40	42	4.59
Expert 7	40	60	55	52	8.16
	50	50	50	50	0.00
	60	40	45	42	8.16
					ΣError^2
					586

10.7 Calculation

Two partial scores are calculated, one for abnormal oral behaviour and one for urine drinking, before being combined into a criterion score.

Partial score for abnormal oral behaviour

An index is calculated from the % of calves performing abnormal oral behaviours:

$$\text{Index } I_p = 100 - (\% \text{ calves performing abnormal oral behaviours})$$

I_p is transformed into a score S_p thanks to I-spline functions:

$$\text{When } I_p \leq 70 \quad \text{then } S_p = (-1003813 E^{12} \times I_p) + (468784 E^{14} \times I_p^2) + (0.00003797 \times I_p^3)$$

$$\text{When } I_p \geq 25 \quad \text{then } S_p = -788.0149 + (33.772 \times I_p) - (0.482458 \times I_p^2) + (0.002335 \times I_p^3)$$

Partial score for urine drinking

An index is calculated from the % of calves with signs of urine drinking

$$\text{Index } I_a = 100 - (\% \text{ calves with signs of urine drinking})$$

I_a is transformed into a score S_a thanks to spline functions:

$$\text{When } I_a \leq 87 \quad \text{then } S_a = (4.00864 E^{11} \times I_a) - (1.5905 E^{12} \times I_a^2) + (1.39688 E^{14} \times I_a^3)$$

$$\text{When } I_a \geq 87 \quad \text{then } S_a = -29972.95 + (1033.5498 \times I_a) - (11.8798798 \times I_a^2) + (0.0455168 \times I_a^3)$$

Score for expression of other behaviour:

S_p and S_a are combined into a score thanks to a Choquet integral. The parameters of the Choquet integral are:

$$\mu_p = 0.11 \text{ and } \mu_a = 0.61$$

11 Criterion 11: Good human-animal relationship

11.1 Definition of the criterion

In this criterion, we aim at evaluating the level of animals' fear of humans or, on the contrary, the positive reactions of animals towards humans (search for the contact with humans).

11.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13129	Calf escape test	Ordinal	Individual

The test consists of targeting one calf, approaching this calf and looking at how close it can be approached. Even though the test was developed on a 5 point scale, it has been decided to keep the best category of calves, those that can be touched.

11.3 Analysis of the constraints

Only one measure will be considered for this criterion. The main constraint was linked to the option of choosing 1) only the best category of calves (% of calves targeted by the experimenter than accept being touched) or 2) the frequency of calves in the first and second best categories (% of calves that can be approached with 2 steps and % of calves that can be touched). The final decision fell on the best category of calves.

11.4 Type of mathematical construction proposed

A function g will be defined to compute proportion of calves accepting being touched so that

$$\text{Score for human-animal relation} = g(p) \quad \text{with } p, \% \text{ of calves being touched}$$

11.5 Parameters to be defined

We need to define an appropriate function g to convert the proportion of calves accepting being touched to a criterion-score.

11.6 Adjustments of the mathematical construction

▪ Questions asked to experts

A dataset corresponding to virtual farms was proposed to experts who were asked to attribute a criterion score between 0 and 100 to each virtual farm (Table 11.1) considering the distribution of the parameter on real farms (distribution, average and maximum) found from WP2.4 partners in 174 farms at about 13 weeks of age of calves (Figure 11.1).

The experts asked were E. Bokkers, M. Brscic, G. Cozzi, F. Gottardo, B.J. Lensink, H. Leruste, K. Van Reenen.

Table 11.1: Dataset submitted to experts. They were asked to attribute scores for human-animal relation according to the % of calves accepting being touched

	frequency of calves that can be touched (% of calves)	Score
Farm 1	0	
Farm 2	5	
Farm 3	10	
Farm 4	15	
Farm 5	20	
Farm 6	30	
Farm 7	40	
Farm 8	50	
Farm 9	60	
Farm 10	80	
Farm 11	100	

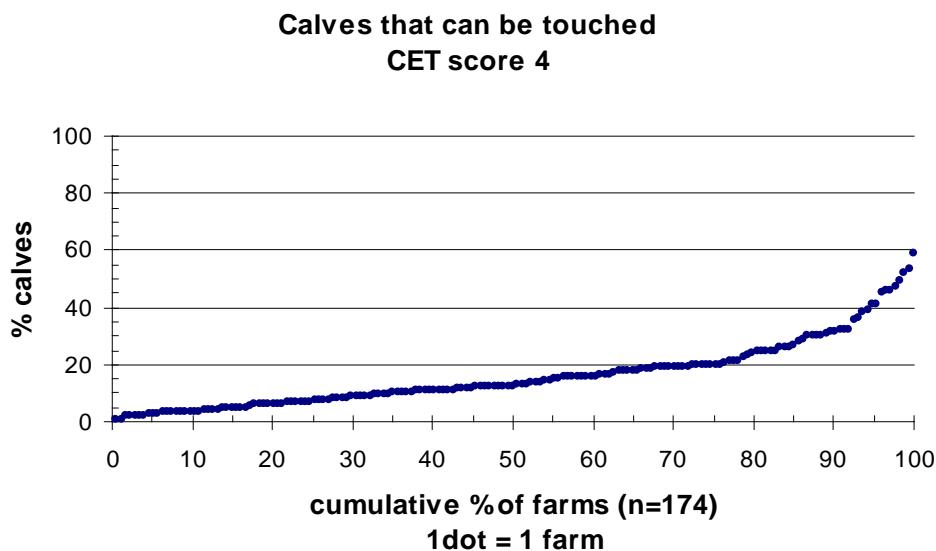


Figure 11.1: Distribution of % of calves accepting being touched 174 farms at 13 wks of age of the calves

- **Adjustment of the construction according to answers from experts**

The results of experts' evaluation of % calves accepting being touched are shown in Table 11.2.

The function g is based on I-Spline function so as to match best expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 11.2):

Knot(s): 1 interior knot at 40

Coefficients:

COEF	VALUE
a_from_knot	14.3179763621012554608569189
b1_before_knot	3.4064390093375997992097837
b2_before_knot	-0.0523318046715347678921049
b3_before_knot	0.0003309973159920349765860
b1_from_knot	2.3325907822848472861210212
b2_from_knot	-0.0254855989969006331563595
b3_from_knot	0.0001072789353783111548262

Table 11.2: Results obtained from the experts asked to attribute scores for human-animal relation according to % of calves accepting being touched.

% calves that can be touched	Score for calves accepting being touched							Index
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	
0	30	0	0	0	0	0	0	4
5	50	0	20	20	10	10	10	17
10	60	20	40	30	20	20	20	30
15	60	30	50	50	25	30	30	39
20	60	50	60	60	30	50	50	51
30	60	60	70	70	40	70	60	61
40	60	70	80	80	50	100	75	74
50	60	80	90	90	60	100	90	81
60	60	90	100	95	70	100	95	87
80	60	100	100	100	80	100	100	91
100	60	100	100	100	100	100	100	94

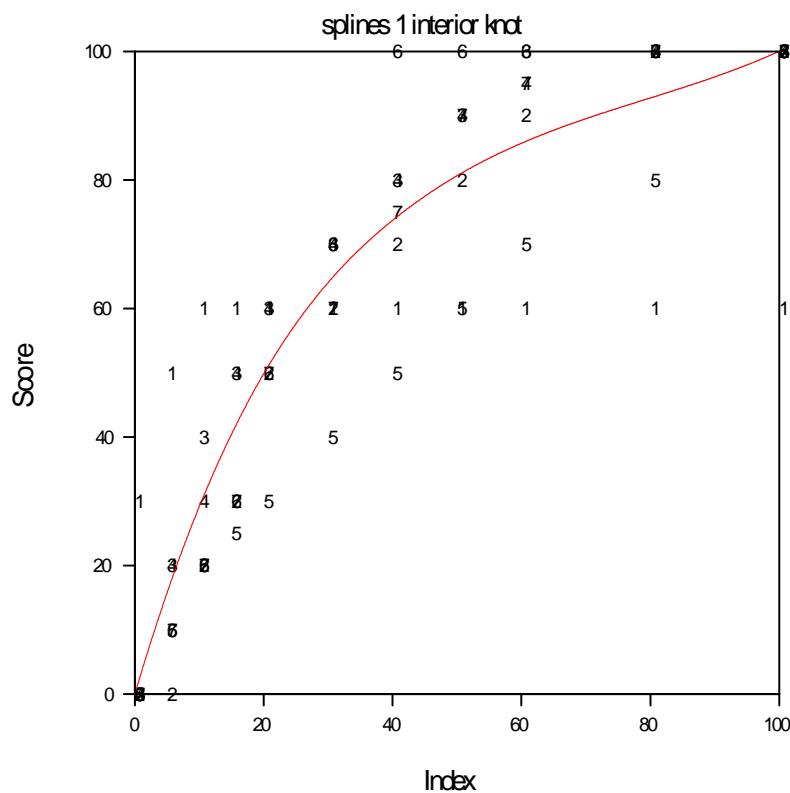


Figure 11.2: I-Spline function allowing to attribute a score for human-animal relation according to % of calves accepting being touched

11.7 Calculation

The % calves that can be touched (I) is transformed into a score thanks to I-spline functions:

$$\text{When } I \leq 40 \text{ then } S = (3.406439 \times I) - (0.05233 \times I^2) + (0.00033 \times I^3)$$

$$\text{When } I \geq 40 \text{ then } S = 14.317976 + (2.33259 \times I) - (0.025485 \times I^2) + (0.000107 \times I^3)$$

12 Criterion 12: Positive emotional state

12.1 Definition of the criterion

This criterion aims at assessing the positive emotional state of the animals. This excludes fear of humans as this point is addressed in Criterion 11.

12.2 Measures to be used to assess the criterion

Id_Meas.	Measure	Nature	Level of observation
13130	Qualitative Behaviour Assessment	Cardinal	Individual

20 qualitative terms are used. For each term, its appropriateness to describe the animals observed is expressed on a 12.5 cm scale. A Principal Component Analysis has been carried out on the 24 veal farms observed in WP2.4. The first axis expresses the general mood of the animal (happy or calm vs. frustrated or apathetic). The coefficients obtained by the various terms on this axis are used to turn the 20 data obtained (from terms) into a single score. This is done by a weighted sum (weights shown in Table 12.1) plus a constant equal to 0.445439.

Table 12.1: qualitative terms used and weights to calculate the “general mood” score obtained by the farm:

Terms	Weights
Uneasy	-0.01543
Frustrated	-0.01482
Apathetic	-0.01383
Tense	-0.01357
Agitated	-0.01237
Depressed	-0.01081
Distressed	-0.01016
Fearful	-0.0093
Bored	-0.00508
Active	-0.00376
Boisterous	-0.00131
Inquisitive	-0.001
Playful	0.003339
Lively	0.00476
Indifferent	0.006489
Calm	0.008234
Sociable	0.01162
Friendly	0.012387
Happy	0.012412
Relaxed	0.015592

12.3 Analysis of the constraints

Only one measure will be used to construct the criterion: a continuous data at farm level. This index corresponds to a weighted sum derived from a PCA. This index is thus rather abstract to experts.

12.4 Type of mathematical construction proposed

A function shall be used to transform the index a farm obtained to a value-score.

12.5 Parameters to be defined

The function f to transform the index a farm obtained to a value-score need be defined.

12.6 Adjustments of the mathematical construction

- **Questions asked to experts**

We decided to consult only the WelfareQuality® partners that were involved in Qualitative Behaviour Assessment (namely Francoise Wemelsfelder who coordinated Qualitative Behaviour Assessment, in turn she contacted the persons who did the assessment on the farms).

We asked them to decide which values of the index shall bring scores 0, 20, 50, 80, 100.

- **Adjustment of the construction according to answers from experts**

The results of experts' evaluation are shown in Table 12.2. The experts considered that an index of 0 corresponds to an animal mood which is both not positive and not negative, hence they attributed a score of 50 (which means not good not bad). The experts then used the maximum and the minimum values for the index observed in WP2.4, whatever the animal type observed. These were +6 and -8. They considered that these should correspond to the end of the value scale, i.e. 100 and 0. Scores 20 and 80 were given to index values in between 0 and minimum or maximum. The correspondence between index and score values is summarised in Table 12.2.

Table 12.2: Correspondence between some index and score values according to experts consulted

Index	Score
-8	0
-4	20
0	50
3	80
6	100

The function f is based on I-Spline function so as to match best expert scores (i.e. minimise the sum of square errors with expert scores) (Figure 12.1). The parameters of these functions are:

Knot(s): 1 interior knot at 0

Coefficients:

	COEF	VALUE
a_from_0		50
b1_before_0		-10
b2_before_0		-1.25000
b1_from_0		11.66667
b2_from_0		-0.55556

with

When Index < 0 then Score = b1_before*Index + b2_before*Index*Index

When Index >= 0 then Score = a + b1_from*Index + b2_from*Index*Index

In addition the score can vary between 0 and 100 only hence

- if a calculation brings a values below 0 then Score = 0
- if a calculation brings a values above 100 then Score = 100

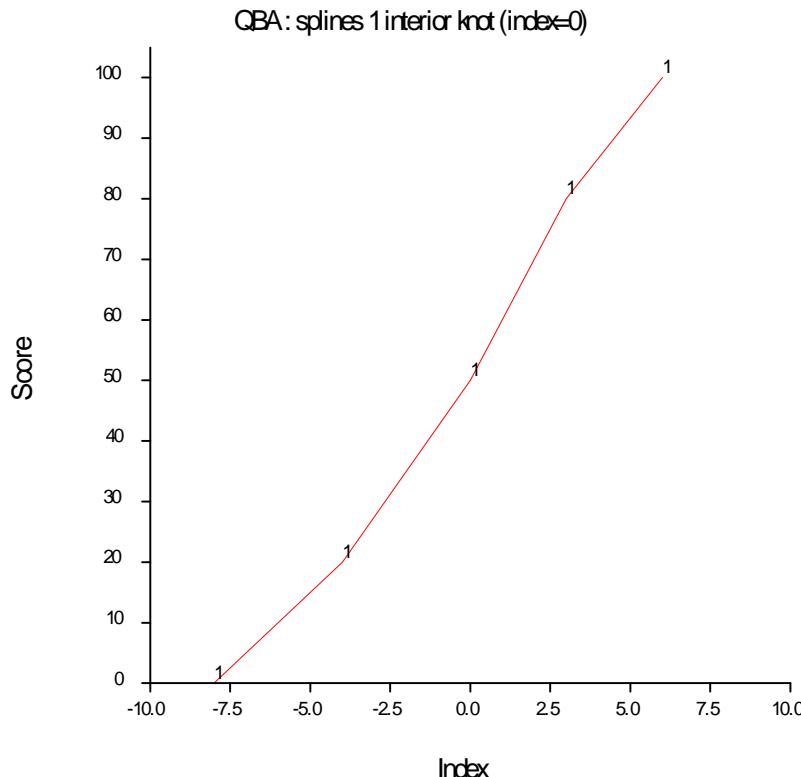


Figure 12.1: I-Spline function allowing attributing a fearfulness score to the index derived from Qualitative Behaviour Assessment

12.7 Calculation

The values (between 0 and 125) obtained by a farm for the 20 terms of the Qualitative Behaviour Assessment are turned into an index thanks to a weighted sum. The weights of the various terms in this sum are:

Terms	Weights
Active	-0.00376
Relaxed	0.01559
Fearful	-0.00930
Agitated	-0.01237
Depressed	-0.01081
Calm	0.00823
Tense	-0.01357
Indifferent	0.00649
Frustrated	-0.01482
Friendly	0.01239
Bored	-0.00508
Playful	0.00334
Lively	0.00476
Inquisitive	-0.00100
Boisterous	-0.00131
Uneasy	-0.01543
Sociable	0.01162
Apathetic	-0.01383
Happy	0.01241
Distressed	-0.01016

The constant to be included in this sum equals **0.445439**.

This index is then transformed into a score thanks to I-spline functions:

Knot(s): 1 interior knot at 0

Coefficients:

COEF	VALUE
a_from_0	50
b1_before_0	-10
b2_before_0	-1.25000
b1_from_0	11.66667
b2_from_0	-0.55556

with

When Index < 0 then Score = b1_before*Index + b2_before*Index*Index

When Index >= 0 then Score = a + b1_from*Index + b2_from*Index*Index

In addition the score can vary between 0 and 100 only hence

- if a calculation bring a value below 0 then Score = 0
- if a calculation bring a value above 100 then Score = 100

Annex: List of the experts consulted to construct the criteria

EXPERTS

Lastname	Firstname	Code
Bokkers	Eddie	1
Brscic	Marta	1
Cozzi	Giulio	1
Gottardo	Flaviana	1
Lensink	Joop B.	1
Leruste	Hélène	1
Van Reenen	Kees	1
Veissier	Isabelle	1

- 1. CATTLE
- 2. PIGS
- 3. POULTRY

Colophon



This report is an official deliverable of the Welfare Quality project.

Welfare Quality is a European research project focussing on the integration of animal welfare in the food quality chain: from public concern to improved welfare and transparent quality.

Welfare Quality is co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508.

Project Office Welfare Quality
Animal Sciences Group of Wageningen UR,
Edelhertweg 15
P.O Box 65, 8200 AB Lelystad,
The Netherlands
Phone +31 320 293503
Fax +31 320 238050
e-mail info@welfarequality.net
Internet www.welfarequality.net

The text of this report represents the authors' views and does not necessarily represent a position of the European Commission who will not be liable for the use made of such information.

ANNEX 2

Gedetailleerde resultaten vereenvoudigingsonderzoek WQ protocol vleeskalveren

1. Vereenvoudiging door verkorten waarnemingstijd

1.1 Resultaten op het niveau van de eindkwalificatie

a1) Totaal (gouden standaard) versus ochtend (vereenvoudiging):

GS1 Ochtend

	Excell	Enhanc	Accept	Not cl	NA	Margin			90 % Confidence interval		
									est	low	upp
Excellent	0	0	0	0	0	0					
Enhanced	0	154	0	0	12	166	%gelijk	100.0	98.4	100.0	
Acceptable	0	0	32	0	6	38	%sp	100.0	98.1	100.0	
Not classified	0	0	0	0	0	0	%se	100.0	91.1	100.0	
NA	0	0	0	0	20	20	%_fp	0.0	0.0	1.9	
Margin	0	154	32	0	38	224	%fn	0.0	0.0	8.9	

a2) Totaal (gouden standaard) versus middag (vereenvoudiging):

GS1 Middag

	Excell	Enhanc	Accept	Not cl	NA	Margin			90% Confidence interval		
									est	low	upp
Excellent	0	0	0	0	0	0					
Enhanced	0	165	0	0	1	166	%gelijk	99.0	96.9	99.8	
Acceptable	0	2	35	0	1	38	%sp	100.0	98.2	100.0	
Not classified	0	0	0	0	0	0	%se	94.6	83.9	99.0	
NA	0	0	0	0	20	20	%_fp	0.0	0.0	1.8	
Margin	0	167	35	0	22	224	%fn	5.4	1.0	16.1	

a3) Totaal (gouden standaard) versus avond (vereenvoudiging):

GS1 Avond

	Excell	Enhanc	Accept	Not cl	NA	Margin			90 % Confidence interval		
									est	low	upp
Excellent	0	0	0	0	0	0					
Enhanced	0	161	1	0	4	166	%gelijk	98.5	96.2	99.6	
Acceptable	0	2	36	0	0	38	%sp	99.4	97.1	100.0	
Not classified	0	0	0	0	0	0	%se	94.7	84.3	99.1	
NA	0	0	0	0	20	20	%_fp	0.6	0.0	2.9	
Margin	0	163	37	0	24	224	%fn	5.3	0.9	15.7	

1.2 Resultaten op het niveau van Principes

Principle 4	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
	est.	lower	upper	est.	lower	upper	est.	lower	upper	
Morgen	%gelijk	100.0	98.4	100.0	97.8	95.1	99.3	100.0	98.4	100.0
	%se				98.5	95.2	99.7	100.0	98.4	100.0
	Rsp=0.994	100.0	98.4	100.0	96.4	89.2	99.4			

Middag		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%gelijk	100.0	98.5	100.0	98.0	95.5	99.3	100.0	98.5	100.0
	%se				97.9	94.6	99.4	100.0	98.5	100.0
Rsp=0.995	%sp	100.0	98.5	100.0	98.4	92.6	99.9			

Avond		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%gelijk	100.0	98.5	100.0	96.5	93.5	98.3	100.0	98.5	100.0
	%se				97.8	94.5	99.4	100.0	98.5	100.0
Rsp=0.995	%sp	100.0	98.5	100.0	93.5	85.8	97.8			

1.3. Resultaten op het niveau van Criteria

Criterium 9	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
	est.	lower	upper	est.	lower	upper	est.	lower	upper	
Morgen	%gelijk	97.3	94.4	98.9	79.6	74.1	84.3	83.9	78.8	88.1
	%se	80.0	34.3	99.0	65.3	55.3	74.4	87.8	82.9	91.7
	Rsp=0.65	97.8	95.0	99.2	89.2	83.1	93.6	35.7	15.3	61.0

Middag		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%gelijk	93.1	89.4	95.8	80.2	75.0	84.7	90.1	85.9	93.3
	%se				91.0	83.8	95.7	93.5	89.7	96.2
Rsp=0.79	%sp	93.4	89.7	96.0	73.4	66.1	79.8	52.9	31.1	74.0

Avond		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%gelijk	91.0	87.0	94.1	75.0	69.4	80.0	90.5	86.4	93.7
	%se				80.8	71.9	87.8	95.1	91.6	97.4
Rsp=0.68	%sp	91.3	87.2	94.4	71.3	63.8	78.0	41.2	21.2	63.6

Criterium 10	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
		est.	lower	upper	est.	lower	upper	est.	lower	upper
Morgen	%gelijk	100.0	98.4	100.0	97.8	95.2	99.3	88.7	84.2	92.3
	%se	100.0	71.7	100.0	93.9	84.9	98.3	88.7	83.2	92.9
	%sp	100.0	98.3	100.0	99.3	96.6	100.0	88.7	78.9	95.0

		est.	lower	upper	est.	lower	upper	est.	lower	upper
Middag	%gelijk	99.0	96.9	99.8	97.5	94.9	99.0	95.5	92.4	97.7
	%se	77.8	45.0	95.9	92.0	82.6	97.2	94.0	89.7	96.8
	%sp	100.0	98.5	100.0	99.3	96.9	100.0	100.0	94.5	100.0

		est.	lower	upper	est.	lower	upper	est.	lower	upper
Avond	%gelijk	100.0	98.5	100.0	96.0	92.9	98.0	95.5	92.3	97.6
	%se	100.0	71.7	100.0	96.0	87.9	99.3	97.3	93.9	99.1
	%sp	100.0	98.4	100.0	96.0	92.3	98.2	90.4	80.8	96.1

Criterium 12	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
		est.	lower	upper	est.	lower	upper	est.	lower	upper
Morgen	%gelijk	97.7	95.3	99.1	88.7	84.6	92.0	80.2	75.3	84.5
	%se	50.0	2.5	97.5	70.8	52.1	85.4	81.6	75.6	86.7
	%sp	98.2	95.9	99.4	90.9	86.8	94.0	77.3	68.0	85.0

		est.	lower	upper	est.	lower	upper	est.	lower	upper
Middag	%gelijk	100.0	98.7	100.0	96.4	93.6	98.2	86.5	82.1	90.1
	%se	100.0	22.4	100.0	95.8	81.7	99.8	85.7	80.1	90.2
	%sp	100.0	98.6	100.0	96.5	93.5	98.3	88.0	80.0	93.6

		est.	lower	upper	est.	lower	upper	est.	lower	upper
Avond	%gelijk	99.1	97.2	99.8	91.4	87.7	94.3	86.5	82.1	90.1
	%se	50.0	2.5	97.5	58.3	39.7	75.4	89.8	84.7	93.6
	%sp	99.5	97.9	100.0	95.5	92.2	97.6	80.0	70.9	87.2

2. Vereenvoudiging door voorspellen abnormaal oraal gedrag uit pensontwikkeling

Bij deze vereenvoudiging is %abnormaal oraal gedrag vervangen door predictie van %abnormaal oraal gedrag uit %pens ontwikkelingsscore 3&4

Predicties zijn berekend met logistische regressie analyse op de logit getransformeerde pensontwikkelingsscore 3 &4. De predictie formules is:

$$1) \text{ pred%abn_all} = 100 / (1 + \text{EXP}(-\text{LP_all})), \text{ met}$$

$$\text{LP_all} = -2.2116 - 0.0771 * \text{Logit(score34)} \text{ en}$$

$$\text{Logit(score34)} = \log((\text{Score34} + 0.5) / (100 + 0.5 - \text{Score34})).$$

Estimates of parameters

Parameter	estimate	s.e.	t(189)	t pr.	antilog of estimate	
Constant	-2.2116	0.0637	-34.71	<.001	0.1095	
LogitScore34	-0.0771	0.0144	-5.35	<.001	0.9258	

2.1 Resultaten op het niveau van de eindkwalificatie

GS1

	Excell	Enhanc	Accept	Not cl	NA	Margin			90% Confidence interval		
									est	low	Upp
Excellent	0	0	0	0	0	0					
Enhanced	0	151	2	0	13	166	%gelijk	97.9	95.3	99.3	
Acceptable	0	2	36	0	0	38	%sp	98.7	95.9	99.8	
Not classified	0	0	0	0	0	0	%se	94.7	84.3	99.1	
NA	0	0	0	0	20	20	%_fp	1.3	0.2	4.1	
Margin	0	153	38	0	33	224	%fn	5.3	0.9	15.7	

2.2. Resultaten op het niveau van Principles

Principle 4	20.0			55.0			80.0		
	90% Conf. interval			90% Conf. interval			90% Conf. interval		
	est.	lower	upper	est.	lower	upper	est.	lower	upper
%gelijk	100.0	98.4	100.0	94.8	91.3	97.1	100.0	98.4	100.0
%se				94.8	90.4	97.5	100.0	98.4	100.0
Rsp=0.98	%sp	100.0	98.4	100.0	94.7	87.0	98.6		

NB: alle bedrijven in de steekproef hebben een score voor Principle 4 tussen 20 en 80.

2.3. Resultaten op het niveau van Criteria

Criterium 10	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
	est.	lower	upper	est.	lower	upper	est.	lower	upper	
Rsp =0.91	%gelijk	99.5	97.5	100.0	93.2	89.4	95.9	91.6	87.6	94.7
	%se	88.9	57.1	99.4	84.0	73.0	91.8	98.6	95.8	99.8
	%sp	100.0	98.4	100.0	96.4	92.7	98.6	67.4	53.9	79.1

3. Vereenvoudiging door het voorspellen van longpathologie uit klinische verschijnselen

In het WQ integratie model wordt voor pneumonia afgeleid uit pathologische afwijkingen aan longen met de pneumonia klassen (normaal , licht problematisch, ernstig probleem) gerekend. De pneumonia klassen zijn zowel voorspeld uit de %prevalentie van de drie respiratoire variabelen: coughing, abnormal breathing and nose discharge, alsmede uit de logistisch getransformeerde %prevalenties. De logistisch getransformeerden zijn ook gebruikt bij de voorspelling van %abnormaal oraal gedrag uit de logistisch getransformeerde %pensontwikkelingscore3&4. Voor de voorspelling van de pneumonia klassen zijn twee methoden gebruikt, nl.:

- (a) continuation ratio model waarbij op grond van dubbele logistische regressie gediscrimineerd wordt tussen de pneumonia klassen en
- (b) ordinale regressie onder het proportional odds model.

Continuation ratio model mogelijkheid 1: middels dubbele logistische regressie (1 vs 2,3 en vervolgens 2 vs 3) zijn voorspellingen voor de pneumonia eindklassen gemaakt uit de %prevalentie van de drie respiratoire aandoeningen.

De frequentieverdeling voor de uit de %pneumonia berekende pneumonia klasse en de met dubbele logistische regressie voorspelde pneumonia klasse is weergegeven in onderstaande tabel

KLASSE_PredPneu KLASSE_Pneu	1	2	3	Nobserved
1	-	23	17	40
2	-	51	24	75
3	-	26	68	94
Nobserved	-	100	109	209

De cijfers in de tabel laten zien dat er geen voorspellingen in klasse 1=normaal (pneumonia <10.5) zijn, hoewel 40 waarnemingen <10.5 en daarmee tot de klasse normaal behoren! Model discrimineert niet voor klasse 1

Continuation ratio model mogelijkheid 2: middels dubbele logistische regressie (3 vs 1,2 en vervolgens 1 vs 2) op de %prevalentie

KLASSE_PredPneu KLASSE_Pneu	1	2	3	Nobserved
1	0	31	9	40
2	0	63	12	75
3	1	45	48	94
Nobserved	1	139	69	209

Continuation ratio model mogelijkheid 1: met de respiratoire aandoeningen op *logit-schaal*.

KLASSE_PredPneu KLASSE_Pneu	1	2	3	Nobserved
1	0	22	18	40
2	0	47	28	75
3	0	18	76	94
Nobserved	0	87	122	209

Continuation ratio model mogelijkheid 2: met de respiratoire aandoeningen op *logit-schaal*.

KLASSE_PredPneu KLASSE_Pneu	1	2	3	Nobserved
1	1	28	11	40
2	0	56	19	75
3	0	38	56	94
Nobserved	1	122	86	209

Proportional odds model: predictie van pneumonia klassen op basis van *ordinale regressie op de % prevalenties van coughing*

KLASSE_PredPneu KLASSE_Pneu	1	2	3	Nobserved
1	-	31	9	40
2	-	63	12	75
3	-	47	47	94
Nobserved	-	141	68	209

Predictie van pneumonia klassen op basis van ordinale regressie op de % prevalenties van de 3 respiratoire klinische variabelen geeft hetzelfde resultaat voor de pred klassen.

Nu met ordinale regressie op de logit getransformeerde prevalenties

$$\text{Logit} = \ln((\text{aantal}+0.5)/(\text{N-aantal} +0.5))$$

KLASSE_PredPneu	1	2	3	Nobserved
KLASSE_Pneu				
1	2	25	13	40
2	1	53	21	75
3	0	37	57	94
Nobserved	3	115	91	209

Regression analysis

Response variates: ordinal model for categories defined by pneuclass[1], pneuclass[2], pneuclass[3]

Distribution: Multinomial

Link function: Logit

Fitted terms: logitcough + logitabnbreath + logtnoseDischarge

Summary of analysis

Source	d.f.	meandeviance		approx	
		deviance	deviance	ratio	chi pr
Regression	3	28.2	9.416	9.42	<.001
Residual	204	408.0	2.000		
Total	207	436.2	2.107		

Estimates of parameters

Parameter	estimate	s.e.	antilog of		
			t(*)	t pr.	estimate
Cut-point 0/1	-4.227	0.761	-5.56	<.001	0.01460
Cut-point 1/2	-2.399	0.716	-3.35	<.001	0.09077
logitcough	0.521	0.157	3.32	<.001	1.684
logitabnbreath	-0.045	0.135	-0.34	0.737	0.9556
logtnoseDischarge	0.303	0.115	2.64	0.008	1.354

Accumulated analysis of deviance

Change	d.f.	meandeviance		approx	
		deviance	deviance	ratio	chi pr
+ logitcough	1	21.019	21.019	21.02	<.001
+ logitabnbreath	1	0.006	0.006	0.01	0.936
+ logtnoseDischarge	1	7.222	7.222	7.22	0.007
Residual	204	407.975	2.000		
Total	207	436.223	2.107		

Conclusie: vergelijken van de resultaten van de dubbele logistische regressie analyses en die voor analyse onder het drempel model leert dat

- met geen van de model varianten klasse 1 goed kan worden onderscheiden van de overige klassen;

- onderscheiden van klasse 2 en klasse 3 niet erg verschilt tussen de varianten.

De kwaliteit kentallen %overeenstemming, %specificiteit en %sensitiviteit zijn berekend voor alle bovengenoemde model varianten. De schattingen voor de kwaliteit kentallen voor de model varianten blijken niet veel te verschillen. Daarom zijn in deze bijlage alleen opgenomen de schattingen van de kwaliteit kentallen uit de ordinale regressie op de logit getransformeerde prevalenties van de respiratoire variabelen (drempelmodel).

3.1. Resultaten op het niveau van de eindkwalificatie

GS1 pneumonia

	Excell	Enhanc	Accept	Not cl	NA	Margin			90% Confidence interval		
	0	0	0	0	0	0			est	lower	upper
Excellent	0	0	0	0	0	0			est	lower	upper
Enhanced	0	163	3	0	0	166		%gelijk	98.5	96.2	99.6
Acceptable	0	0	38	0	0	38		%sp	98.2	95.4	99.5
Not classified	0	0	0	0	0	0		%se	100.0	92.4	100.0
NA	0	0	0	0	20	20		%_fp	1.8	0.5	4.6
Margin	0	163	41	0	20	224		%fn	0.0	0.0	7.6

3.1. Resultaten op het niveau van Principles

Principle 3	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
	est.	lower	upper	est.	lower	upper	est.	lower	upper	
%gelijk	100.0	98.7	100.0	91.9	88.2	94.7	99.1	97.2	99.8	
%se				95.6	91.9	97.9	99.5	97.8	100.0	
Rsp=0.89	%sp	100.0	98.7	100.0	82.3	72.4	89.7	87.5	52.9	99.4

3.1. Resultaten op het niveau van Criteria

Criterium 7	20.0			55.0			80.0			
	90% Conf. interval			90% Conf. interval			90% Conf. interval			
	est.	lower	upper	est.	lower	upper	est.	lower	upper	
%gelijk	99.6	97.9	100.0	95.5	92.6	97.6	98.7	96.6	99.6	
%se	100.0	76.2	100.0	99.0	97.0	99.8	100.0	98.6	100.0	
Rsp=0.86	%sp	99.5	97.8	100.0	46.7	24.4	70.0	57.1	22.5	87.1

4. Vereenvoudiging door halveren van de steekproef voor klinisch onderzoek

4.1. Resultaten op het niveau van de eindkwalificatie

GS1	
%agreement	50% (5%-95%)
-Halve steekproef	90.7(88.0-93.6)

%specificiteit	50% (5%-95%)
-Halve steekproef	94.0(91.6-96.7)

%sensitiviteit	50% (5%-95%)
-Halve steekproef	78.9(68.4-86.8)

4.1. Resultaten op het niveau van Principles

	20	55	80
Principle 1	50% (5%-95%)		
-%agreement	100.0(100.0-100.0)	93.7(91.0-95.5)	91.9(90.0-94.6)
-%specificiteit	100.0(100.0-100.0)	94.9(91.7-96.9)	93.5(87.1-96.8)
-%sensitiviteit		89.1(84.8-95.7)	91.8(89.3-94.7)

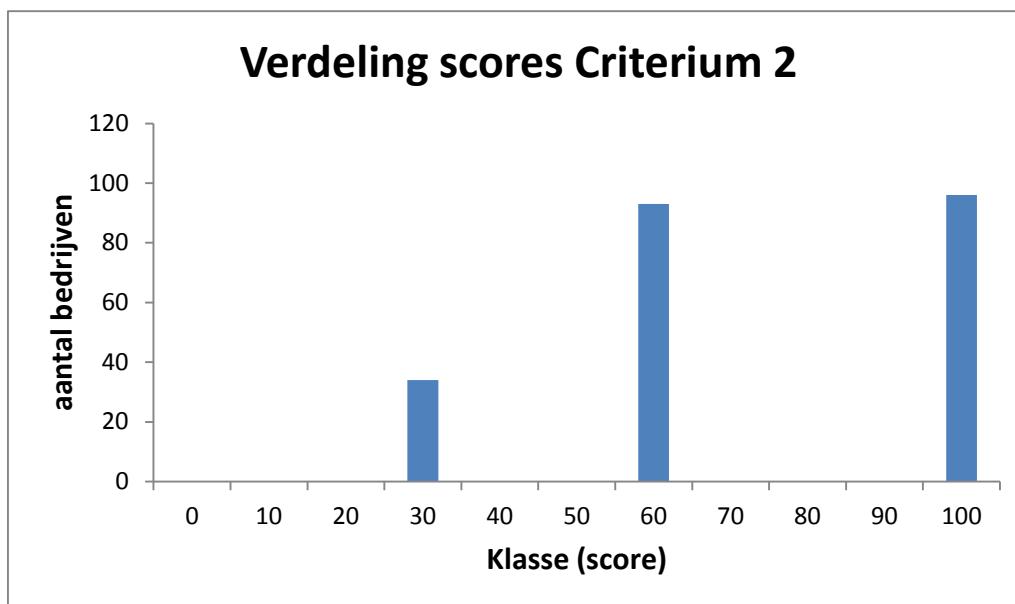
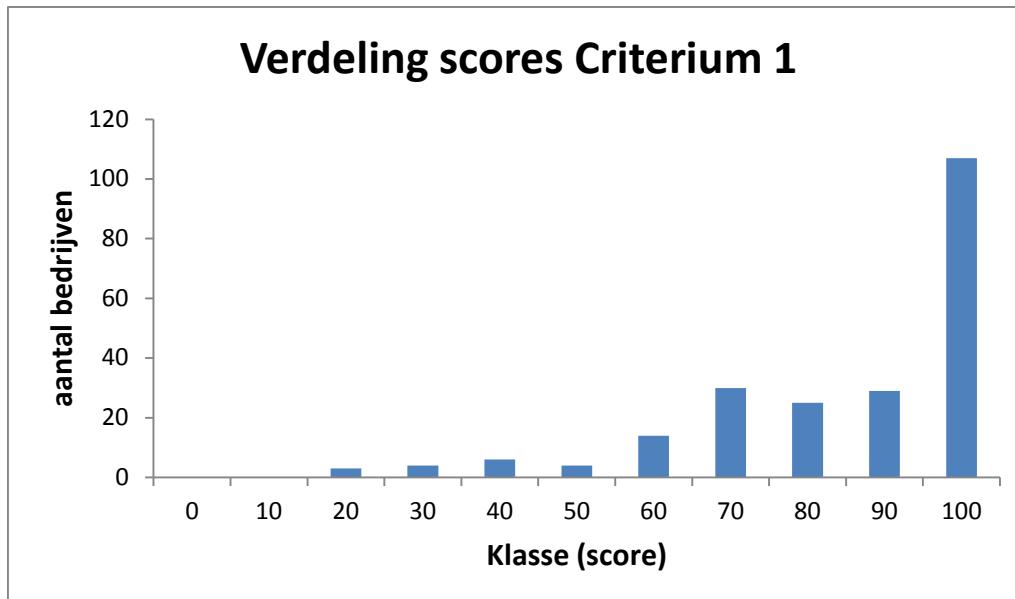
Principle 2	50% (5%-95%)		
-%agreement	99.5(98.6-100.0)	93.7(91.9-95.9)	100.0(100.0-100.0)
-%specificiteit	99.5(99.1-100.0)	96.7(94.8-98.3)	
-%sensitiviteit	83.3(33.3-100.0)	82.9(75.6-90.2)	100.0(100.0-100.0)

Principle 3	50% (5%-95%)		
-%agreement	100.0(100.0-100.0)	88.7(85.6-91.0)	97.3(96.2-98.6)
-%specificiteit	100.0(100.0-100.0)	80.6(74.2-87.1)	50.0(25.0-81.2)
-%sensitiviteit		91.2(88.8-94.4)	99.1(98.1-100.0)

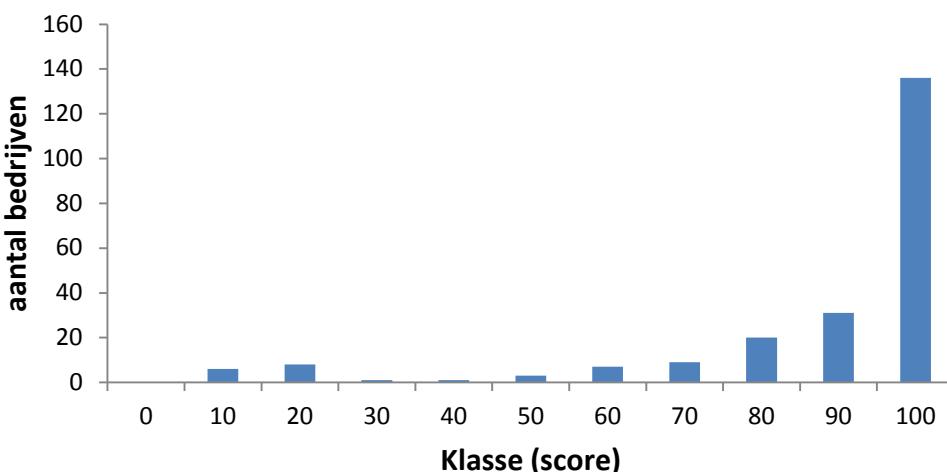
Principle 4	50% (5%-95%)		
-%agreement	100.0(100.0-100.0)	88.5(84.3-91.9)	100.0(100.0-100.0)
-%specificiteit	100.0(100.0-100.0)	85.5(77.4-90.3)	
-%sensitiviteit		89.4(86.3-93.7)	100.0(100.0-100.0)

ANNEX 3

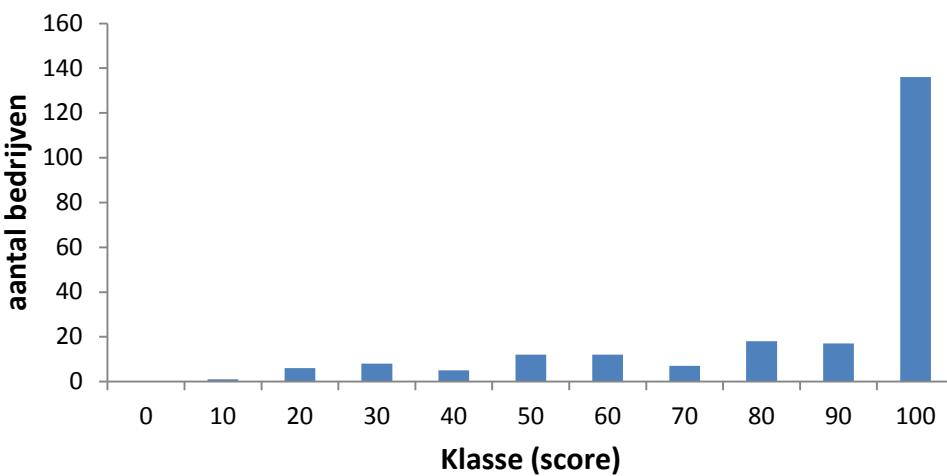
Verdeling van bedrijven over scores voor Criteria en Principles



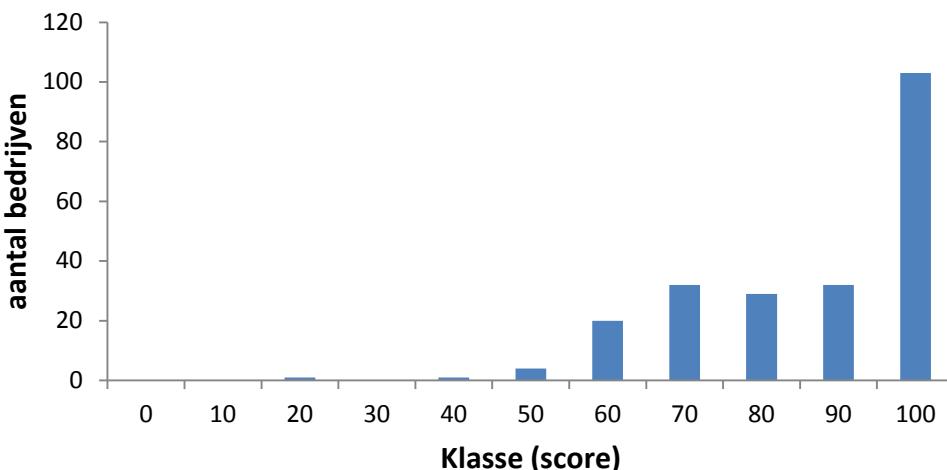
Verdeling scores Criterium 3



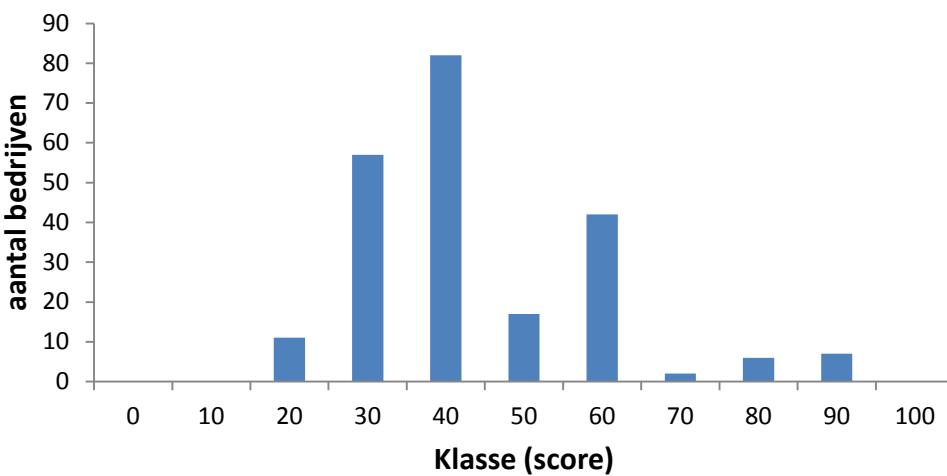
Verdeling scores Criterium 4



Verdeling scores Criterium 6

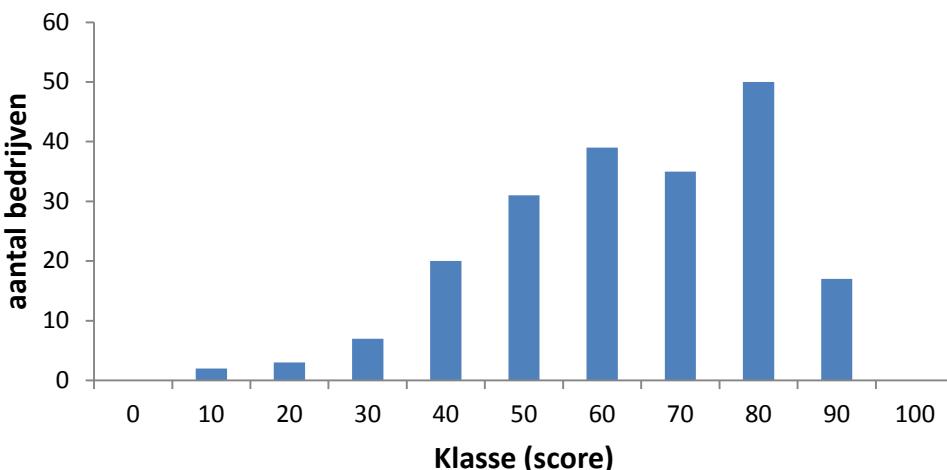


Verdeling scores Criterium 7

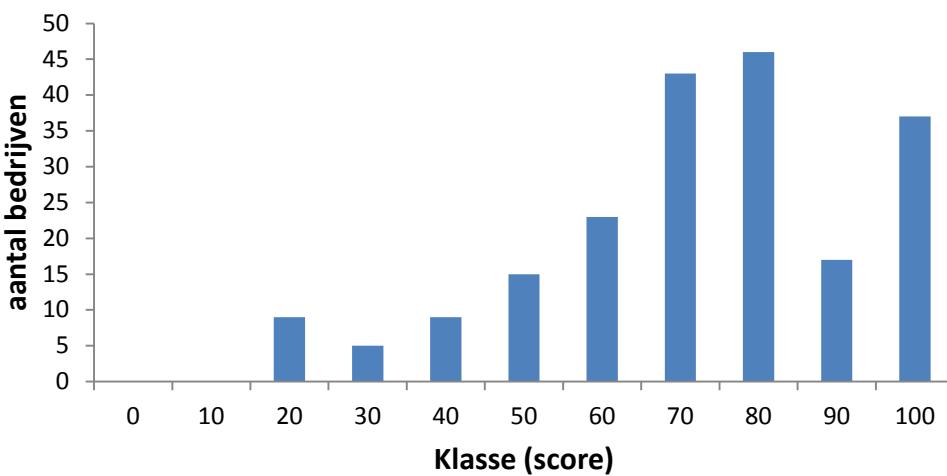


NB: Voor alle bedrijven had Criterium 8 de waarde 100 (geen ingrepen).

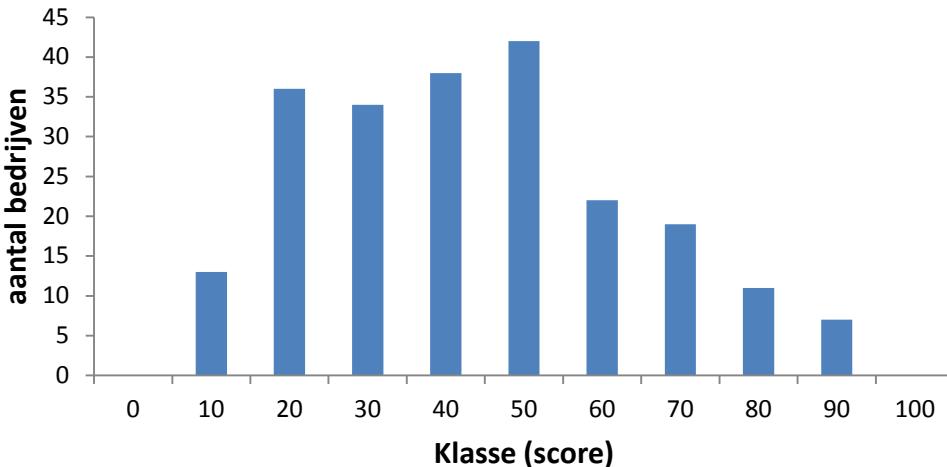
Verdeling scores Criterium 9



Verdeling scores Criterium 10

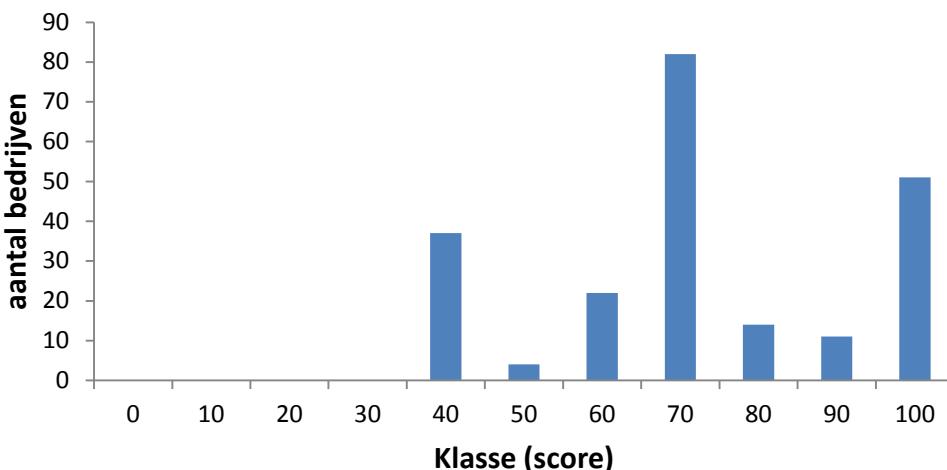


Verdeling scores Criterium 11

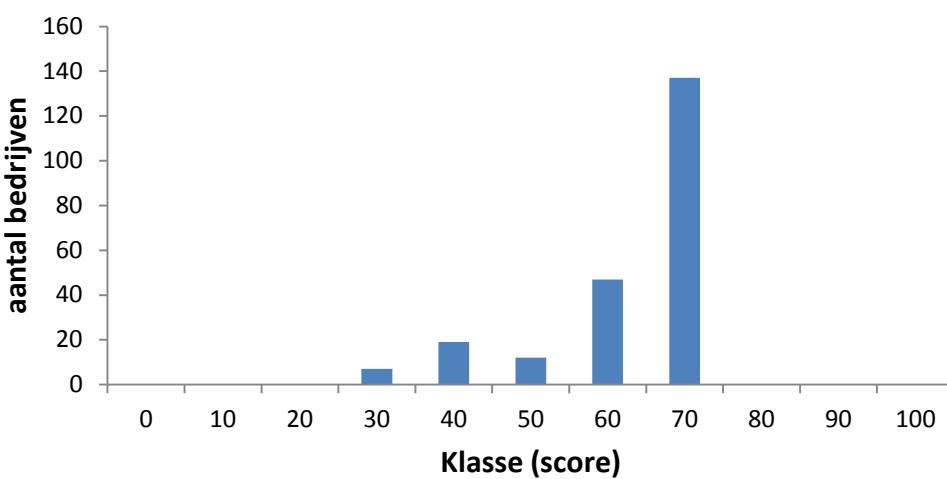


NB: Criterium 12 wordt helemaal bepaald door de uitkomsten van Qualitative Behavioural Assessment (QBA). QBA is in het onderzoek met vleeskalveren maar zeer beperkt uitgevoerd (alleen op de Italiaanse bedrijven, d.w.z. op slechts 24 bedrijven). Als waarde voor Criterium 12 is daarom, voor elk individueel bedrijf, het maximum van de waarden van Criterium 9, Criterium 10 en Criterium 11 gebruikt.

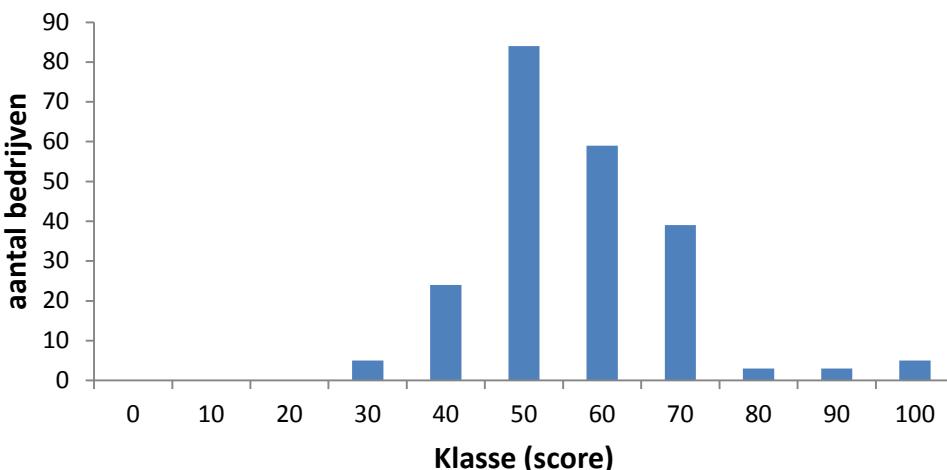
Verdeling scores Principle 1



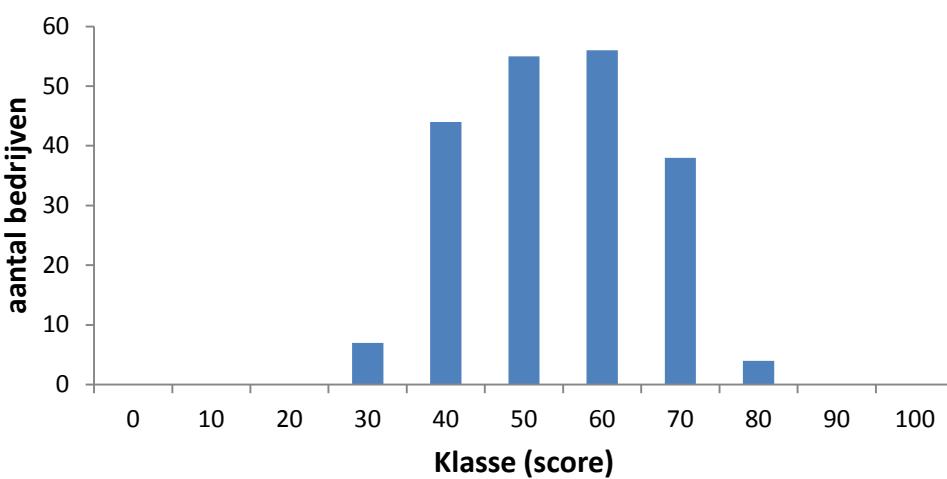
Verdeling scores Principle 2



Verdeling scores Principle 3



Verdeling scores Principle 4





Wageningen UR Livestock Research

Edelhertweg 15, 8219 PH Lelystad **T** 0320 238238 **F** 0320 238050

E info.livestockresearch@wur.nl | www.livestockresearch.wur.nl