

Design und Validierung eines Hygienescores für Kälber

von Laura Maria Kellermann

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität
München

Design und Validierung eines Hygienescores für Kälber

von Laura Maria Kellermann

aus Ostfildern

München 2020

Aus dem Zentrum für Klinische Tiermedizin der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München

Lehrstuhl für Innere Medizin und Chirurgie der Wiederkäuer

Arbeit angefertigt unter der Leitung von:
Univ.-Prof. Dr. Gabriela Knubben-Schweizer

Mitbetreuung durch:
Dr. Moritz Metzner

Gedruckt mit Genehmigung der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München

Dekan: Univ.-Prof. Dr. Reinhard K. Straubinger, Ph.D.

Berichterstatter: Univ.-Prof. Dr. Gabriela Knubben-Schweizer

Korreferent: Univ.-Prof. Dr. Michael H. Erhard

Tag der Promotion: 25. Juli 2020

Meiner Familie

INHALTSVERZEICHNIS

I.	EINLEITUNG	1
II.	LITERATURÜBERSICHT	3
1.	Definition des Begriffs „Score“	3
2.	Hygienescores für Kälber	4
3.	Validierung von Scores	9
	3.1. Grundlagen der Validierung.....	9
	3.2. Validierte Scores bei Kälbern	10
	3.3. Validierte Scores bei Kühen.....	11
III.	PUBLIKATION	13
IV.	DISKUSSION	21
V.	ZUSAMMENFASSUNG	27
VI.	SUMMARY.....	29
VII.	LITERATURVERZEICHNIS	31
VIII.	DANKSAGUNG	39

ABKÜRZUNGSVERZEICHNIS

CHS	Calf Hygiene Score
PraeRi	Prävalenzstudie Rind
RZ	Risk Zone
RZB	Risk Zone Belly
RZR	Risk Zone Rear
RZS	Risk Zone Side
SM	Standard Method
z. B.	Zum Beispiel

I. EINLEITUNG

Die Haltungsbedingungen von Milchkühen sind stark im Fokus von Politik, Landwirtschaft und Verbraucher*innen. Im Spannungsfeld zwischen ökonomischen, ethischen, gesundheitlichen und ökologischen Zielen sind Tierärzt*innen bestrebt, eine Entwicklung der Nutztierhaltung in Richtung Tierwohl und Krankheitsprävention zu unterstützen. Dies erstreckt sich auch auf den Bereich der Kälberaufzucht, um durch eine gesunde Nachzucht den Grundstein für leistungsfähige und gesunde Milchkühe zu legen.

Die Bedeutung der Hygiene zur Prävention von Krankheiten bei Milchkühen wurde schon in vielen Arbeiten untersucht. BARKEMA et al. (1998) stellten Zusammenhänge zwischen der Sauberkeit von Kühen und von den Ställen, in denen sie gehalten werden, mit Tankmilchzellgehalten her. PHILIPOT et al. (1994) zeigten, dass eine nasse, verunreinigte Aufstallung einen Risikofaktor für Lahmheit bei Milchkühen darstellt. Aufgrund solcher Feststellungen haben RENEAU et al. (2005) einen Hygienescore entwickelt, der die Verschmutzung an Euter, Bauch, Unter- und Oberschenkel sowie der Schwanzregion von Milchkühen beschreibt. Dadurch kann eine standardisierte Bewertung des hygienischen Zustandes der Kuh mit dem Risiko für erhöhte „Somatic Cell Scores“ in Zusammenhang gebracht werden. COOK (2002) beschreibt einen Score für die Verschmutzung am Euter, der mit dem Risiko für das Auftreten von Euterentzündungen korreliert.

Bei Kälbern fehlen bisher Nachweise für einen Zusammenhang zwischen dem Auftreten bestimmter Krankheiten und Verschmutzungen am Kalb. Ein für solche Untersuchungen notwendiges validiertes Scoring System war bisher auch nicht vorhanden. Um Verschmutzungen am Kalb zuverlässig zu erfassen und wiederzugeben, wird in der vorliegenden Arbeit ein Hygienescore für Kälber vorgestellt und seine Validierung beschrieben. Mit diesem belastbaren Score können in zukünftigen Studien die Verschmutzung von Kälbern objektiv festgestellt, Einflussfaktoren bestimmt und Zusammenhänge mit der Gesundheit von Kälbern untersucht werden.

II. LITERATURÜBERSICHT

1. Definition des Begriffs „Score“

Ein Score ist in der Medizin ein Wert einer meist eindimensionalen Skala, der nach einer bestimmten Untersuchung oder Begutachtung vergeben wird. Dadurch kann ein bestimmter Zustand oder eine bestimmte Eigenschaft eines Patienten schnell beurteilt werden, ohne dass weitere Informationen benötigt werden. Das Scoring kann verschiedene Bereiche umfassen, so kann ein „Gesundheitsscore“ beim Kalb z. B. die Körpertemperatur, das Vorhandensein von Nasenausfluss und die Kotkonsistenz umfassen. Für jeden Parameter werden festgelegte Punktzahlen vergeben, die zusammen einen Score ergeben, der den Gesundheitszustand des Tieres in einem Wert zusammenfasst.

Die Art der Daten ist je nach Score sehr unterschiedlich: Die zu beurteilende Eigenschaft kann in metrischen (z. B. Brustumfang in Zentimeter), ordinalen (z. B. Bewertung der Lahmheit von Schweregrad 1 bis 5) oder nominalen (z. B. Rasse: Fleckvieh) Werten dargestellt werden.

Dass ein Scoring eine klinische Untersuchung nicht ersetzen kann, ist selbstverständlich. Aber die Vergabe eines Scores ermöglicht den Vergleich verschiedener Patienten in dem durch den Score beschriebenen Bereich, was z. B. bei der Beurteilung des Tierwohls häufig genutzt wird (WELFARE QUALITY®, 2009). Dafür sollte jedoch nachgewiesen sein, dass der Score reproduzierbar zuverlässige und valide Beurteilungen ermöglicht (KNIERIM & WINCKLER, 2009). Dies kann statistisch nachgewiesen werden, indem die Intra- und Inter-Observer-Reliabilität und die Validität in Bezug auf einen Referenzwert berechnet werden (MEAGHER, 2009).

2. Hygienescores für Kälber

In Untersuchungsprotokollen von Studien zur Gesundheit oder Haltung von Kälbern finden sich viele verschiedene Scores, unter anderem auch „Hygienescores“ für Kälber, die deren Verschmutzung beschreiben. Häufig wird z. B. bei der Beurteilung der Gesundheit der Kälber der Zustand des Haarkleids miteinbezogen. Nur die wenigsten der Scores wurden auf ihre Wiederholbarkeit und Zuverlässigkeit getestet. Eine Übersicht über Scores zur Verschmutzung von Kälbern ist in Tabelle 1 wiedergegeben.

Tabelle 1: Übersicht verschiedener Scores zur Beurteilung der Verschmutzung von Kälbern

Erstautor, Jahr	Score Aufbau	Validiert
ROTH et al. (2009)	Verschmutzung durch Faeces, Skala mit drei Stufen	Nein
JORGENSEN et al. (2017)	Verschmutzung durch Faeces, Region beschrieben, Score 0-2	Nein
KLEIN-JÖBSTL et al. (2014)	Verschmutzung durch Faeces, prozentuale Beurteilung von drei Regionen	Nein
CALVO-LORENZO et al. (2016)	Verschmutzung durch Faeces, Skala mit fünf Stufen (nach Prozent)	Nein
TEROSKY et al. (1997)	3 Regionen und Total, Score 1-5	Nein
WEBSTER et al. (1985)	Fünf Regionen, Skala von 1-3	Nein
PANIVIVAT et al. (2004)	Score 1-4 an bis zu drei Regionen	Nein
GRAHAM et al. (2018)	Score 0-3 an bis zu vier Regionen	Nein
LUNDBORG et al. (2005)	Verschmutzung durch Faeces an acht Regionen, Score 1-4 nach Prozent, Mittelwert-Berechnung der schmutzigsten Regionen zu einem Tier-Gesamtscore und Herdenscore	Nein, aber Abgleich der Werte nach gemeinsamem Scoring

Um den Gesundheitsstatus von Kälbern zu bestimmen, wurden in einer Studie von ROTH et al. (2009) die Tiere täglich anhand eines Gesundheitsscores beurteilt. Dieser Score setzt sich aus der Summe der Punkte in den einzelnen Bereichen zusammen (Tabelle 2).

Tabelle 2: Tägliches Gesundheitsscoring für Kälber, ROTH et al. (2009)

Evaluiertes Parameter	Punktvergabe nach Ausprägung
General condition	Normal condition 0 Slightly impaired 1 Severely impaired 2
Incidence of cough	None 0 Sporadic coughing 1 Repeated coughing 2
Dirtiness caused by feces	Inconspicuous 0 Dry dirtiness 1 Moist or wet dirtiness 2
Ears	Normal 0 Lop-eared 1
Eyes	Inconspicuous 0 Slight watering 1
Nose, nasal discharge	No discharge 0 Watery discharge or dry nose 1 Purulent discharge 2
Navel	Inconspicuous 0 Swollen, without discharge 1 Swollen, with discharge 2

In den Ergebnissen der Studie korreliert der Gesundheitsscore, der auch eine Beurteilung der Verschmutzung am Kalb durch Kot miteinbezieht, mit der Anzahl tierärztlicher Behandlungen.

JORGENSEN et al. (2017) beschreiben die Auswirkungen von verschiedenen Parametern bei der Kälberhaltung mit Fütterungsautomaten. Es wurde zusätzlich zu anderen gesundheitsbezogenen Erhebungen ein „hide dirtiness score“ verwendet, um zu erfassen, ob ein Kalb Durchfall hat oder nicht. Der Score wurde nach MCGUIRK (2008) modifiziert, wobei McGuirk mit dem „fecal score“ ursprünglich die Konsistenz von Kälberkot beschrieb. Bei dem Score werden die Perianalregion, die Unterseite des Schwanzes und der Schwanzansatz beurteilt. Die Skala umfasst „0 = clean“, „1 = evidence of loose or abnormal fecal consistency“ und „2 = significant evidence of watery diarrhea“. Die Korrelation zwischen dem Scoring und dem tatsächlichen Auftreten von Durchfall wurde nicht überprüft. Je schneller die Milchmenge gesteigert wurde, desto gesünder und sauberer waren die Kälber. Die Sauberkeit der Kälber war zudem besser, wenn ihnen eine größere maximale Milchmenge zur Verfügung gestellt wurde. Im Winter waren die Kälber stärker verschmutzt und erzielten beim Scoring schlechtere Gesundheitswerte.

KLEIN-JÖBSTL et al. (2014) untersuchten einen Zusammenhang von Managementfaktoren und dem Auftreten von Kälberdurchfall. Die Verschmutzung

(nur mit Kot) von Bauch, Oberschenkeln und den distalen Gliedmaßen wurde in Prozent bei je fünf zufällig ausgesuchten Kälbern pro Betrieb erfasst. Die Ergebnisse zeigen keinen Zusammenhang zwischen der Verschmutzung der Kälber und dem Auftreten von Durchfall. Zu erwähnen ist, dass die Kälber auf 94 von 100 Betrieben als sauber beurteilt wurden. Das Reinigungsintervall der Abkalbebox wirkte sich auf das Durchfallrisiko aus. Starke Einflüsse hatten außerdem Vorerkrankungen mit einer Atemwegsinfektion, der Standort der Kälber und das Vorhandensein anderer Tierarten auf dem Betrieb.

Bei TEROSKY et al. (1997) wurden Kälber in Einzelboxen und in Anbindehaltung mit Trennwänden mit verschiedenen Breiten untersucht. Dabei wurde die Verschmutzung der Kälber nach einer Skala von 1-5 beurteilt. Der Score 1 steht für „very little manure“ und der Score 5 für „caked-on manure over majority of body section“. Die Zwischenstufen wurden nicht definiert. Die Scores werden je für die Vordergliedmaßen, den Bauch, die Hintergliedmaßen und für das gesamte Erscheinungsbild vergeben. Dieser Hygienescore wurde von CALVO-LORENZO et al. (2016) für eine Skala von 0-4 genauer definiert und auch modifiziert. Die Verschmutzung am Kalb wird nun über den ganzen Körper hinweg bewertet.

Score 0	No feces on calf coat
Score 1	Some dried feces on calf coat (~25% coverage)
Score 2	Little dried and wet feces on calf coat (~50% coverage)
Score 3	Substantial wet and caked on feces (~75% coverage)
Score 4	Extensive wet and caked on feces (~ >80% coverage)

In beiden Studien wurde untersucht, ob diverse Gesundheits- und Verhaltensparameter (u. a. auch die Verschmutzung der Kälber) mit unterschiedlichen Aufstallungsformen der Kälber in Beziehung stehen. Während bei TEROSKY et al. (1997) unter anderem das Wachstum, die Nabelgesundheit und die meisten Blutparameter nicht durch die Aufstallungsform beeinflusst wurden, beobachteten die Autoren bei zunehmender Breite (von 56 cm über 66 bis zu 76 cm Breite) in Boxen für nicht angebundene Kälber eine stärkere Verschmutzung als bei den angebondenen Kälbern. Fast 20 Jahre später stellten CALVO-LORENZO et al. (2016) allerdings fest, dass die Sauberkeit der Kälber

bei einem großzügigen Platzangebot (3,71 m² pro Kalb) am besten ist.

Eine Studie von WEBSTER et al. (1985) untersuchte den Einfluss verschiedener Kälberaufzuchtssysteme (u. a. Mutterkuhhaltung, Einzel- und Gruppenhaltung, verschiedene Zeitpunkte für das Absetzen) auf die Inzidenz von Krankheiten, Verschmutzung und Verletzungen der Kälber. Die Verschmutzung der Kälber wird auf beiden Seiten auf einer dreistufigen Skala an folgenden Regionen festgehalten:

- i) Beine, oberhalb der Fessel
- ii) Hinterbeine und Schwanz
- iii) Bauch
- iv) Rücken (Schlamm und Kot)
- v) Rücken (Schweiß oder Staub)
- vi) Kopf

Früh abgesetzte Kälber waren, unabhängig von der Aufstallungsform, weniger schmutzig als Mastkälber, die noch mit Milchaustauscher getränkt wurden.

PANIVIVAT et al. (2004) untersuchten in einer Studie das Wachstum und den Gesundheitszustand von Kälbern, die auf verschiedenen Einstreumaterialien gehalten wurden. Dabei wurde täglich zur Beurteilung der Verschmutzung der Kälber ein modifiziertes Scoring der „Pharmacia Animal Health Hygiene Scorecard“ für Milchkühe verwendet.

- | | |
|---------|--|
| Score 1 | Calf is clean (not dirty around thighs or body), only manure at lower ends of legs |
| Score 2 | Tail head region and back of calf are soiled with manure |
| Score 3 | Tail head region, thighs or legs are soiled with manure |
| Score 4 | Thighs, legs, and tail head region are soiled with manure |

Die Verschmutzung der Einstreu und die Verschmutzung der Kälber unterschied sich je nach verwendetem Material, und das Material hatte auch einen Einfluss darauf, wie oft die Kälber wegen Durchfall behandelt wurden. Der gleiche Score wurde auch in Studien von SUTHERLAND et al. (2013; 2014; 2017) verwendet. Es wurden die Zusammenhänge von Einstreumaterialien und der Gesundheit, der Physiologie, dem Verhalten, der Sauberkeit und der Thermoregulation von Kälbern untersucht. Einflüsse z. B. auf die Thermoregulation, das Liegeverhalten und die

Anzahl von *E. coli* auf der Hautoberfläche wurden nachgewiesen, jedoch nicht auf die sichtbare Verschmutzung, da die Kälber in allen drei Studien auf allen Einstreumaterialien als sehr sauber beurteilt wurden. GRAHAM et al. (2018) untersuchten, ob die Verschmutzung am Tier zuverlässig auf Durchfall hinweist. Dafür wurde der Score von PANIVIVAT et al. (2004) angepasst, so dass nun bei Score 3 und 4 sowohl auch der Schwanzansatz als auch das „hintere Ende“ des Kalbes (= back end) aufgezählt werden. Aus den Daten konnte, entsprechend der Studie von KLEIN-JÖBSTL et al. (2014), nur ein sehr schwacher Zusammenhang hergestellt werden. Die Autoren erklären abschließend, dass weder einmalige noch fortlaufende Evaluierungen der Verschmutzung der Kälber das tatsächliche Auftreten von Durchfall beschreiben können.

Eine sehr komplexe Beurteilung der Hygiene für einen Kälberbestand im Rahmen von wiederholten Besuchen beschreiben LUNDBORG et al. (2005). In der Studie wurde der Einfluss von Managementfaktoren und Umwelteinflüssen auf typische Kälbererkrankungen untersucht. Die Erhebungen umfassen auch den Hygienestatus der Kälber. Die jeweils jüngsten zwei Kälber der Betriebe wurden bei jedem Besuch ausgesucht bis zu einer Anzahl von zehn Kälbern pro Betrieb. Bei diesen Kälbern wurde bei jedem Betriebsbesuch an acht Körperregionen (Vorder- und Hinterklauen, Vorderbeine unterhalb des Carpus, Hinterbeine unterhalb des Sprunggelenks, Brustkorb, Oberschenkel, Bauch und Schwanz) die prozentuale Verschmutzung durch Faeces beurteilt. Von den Regionen wurde der Median der Region für jedes Kalb berechnet und die drei am stärksten verschmutzten Regionen (wobei nur die mehr verschmutzten Klauen miteinbezogen werden dürfen) zu einem medianen Kälberscore zusammengefasst. Der Median aller Kälberscores des Betriebes wurde als Herdenscore definiert. Die Autoren beschreiben eine Überprüfung der Zuverlässigkeit des Scorings folgendermaßen: Die Tierärzt*innen des Projektes vergaben an drei Terminen gemeinsam Scores, um sich in ihren Beurteilungen anzugleichen. Eine statistische Überprüfung wird nicht angegeben. In den Ergebnissen der Studie wird die Sauberkeit der Kälber nicht als Einflussfaktor auf Kälbererkrankungen erwähnt.

3. Validierung von Scores

3.1. Grundlagen der Validierung

Prinzipiell sollten nach MEAGHER (2009) in eine Validierung eines Scores folgende Aspekte miteinbezogen werden:

- Feststellung der Validität der Methode (also der Eignung zur Überprüfung der zu messenden Aspekte), z. B. durch den Vergleich mit Referenzwerten
- Eine Überprüfung der Intra- und Inter-Observer-Reliabilität
- Bei kontinuierlicher Datenerhebung: Eine Überprüfung der Test-Retest-Zuverlässigkeit

Dabei gibt es für die Berechnung der Reliabilität mehrere Verfahrensweisen, die je nach Art der erhobenen Werte verwendet werden und unterschiedliche Vor- und Nachteile aufweisen (STEMLER, 2004). Das Maß der Übereinstimmung wird z. B. in einem Kappa-Wert ausgedrückt. Aus diesem Wert lässt sich ableiten, ob es durch den Score gelingt eine subjektive Wahrnehmung von Untersuchern zu einer standardisierten Aussage zu formen. Für die Übereinstimmung bei medizinischen Fragestellungen kann für Kappa-Werte die Einteilung nach MCHUGH (2012) verwendet werden (Tabelle 3).

Tabelle 3: Interpretation von Kappa nach MCHUGH (2012)

Value of Kappa	Level of Agreement	% of Data that are Reliable
0 – 0,20	None	0 – 4 %
0,21 – 0,39	Minimal	4 – 15 %
0,40 – 0,59	Weak	15 – 35 %
0,60 – 0,79	Moderate	35 – 63 %
0,80 – 0,90	Strong	64 – 81 %
Mehr als 0,90	Almost Perfect	82 – 100 %

In Fachgebieten wie der Verhaltenskunde, die sich oft auf Beobachtungen schwer messbarer Aktionen bezieht, wird die Notwendigkeit zur Validierung deutlich wahrgenommen. Viele der durch Ethologen genutzten Erfassungssysteme für

Rinder sind nachweisbar zuverlässig anwendbar (KNIERIM & WINCKLER, 2009; VOGT et al., 2017). Verbreitete Systeme wurden aber auch schon durch Validierungen in ihren Aussagen eingeschränkt (BOKKERS et al., 2012). In der klinischen Beurteilung von Kälbern und Kühen könnte man fälschlicherweise annehmen, dass hier handfeste Parameter untersucht werden und daher wenig Abweichung vorkommt. Es mag daher überraschend sein, dass selbst in Messungen, wie der Körpertemperatur mit dem Thermometer, Abweichungen unter den Untersuchern festgestellt werden können (BURFEIND et al., 2010). Dies stellt in der Praxis nur ein geringes Problem dar, ist aber ein gutes Beispiel dafür, dass die Validität einer Methode nicht ohne Überprüfung angenommen werden darf. Auch wenn Scores wie der „Nesting-Score“ (LAGO et al., 2006) dem Nutzer unmissverständlich erscheinen, wäre eine Validierung eine ergebnisbestätigende Aussage, die von künftigen Studien getätigt werden sollte (MEAGHER, 2009).

3.2. Validierte Scores bei Kälbern

Wie einleitend erwähnt, wurde bisher kein Score für das Maß an Verschmutzung an Kälbern validiert. Auch wurden Gesundheits-Scoringsysteme für Kälber nach dem Kenntnisstand der Autorin nur selten auf die Übereinstimmung der Observer überprüft. Eine Erwähnung findet die Observer Reliabilität bei STUDDS et al. (2018). Demnach führten die Observer die Untersuchungen der Kälber so oft durch, bis ein guter Wert für die Übereinstimmung erbracht wurde. Dabei bezogen sich die Werte auf Untersuchungen des Nabels nach FECTEAU et al. (1997) und auf das Beurteilen von Durchfall nach MCGUIRK (2008).

Im Bereich der respiratorischen Erkrankungen wurden mehrfach Untersuchungen zur Validierung von Diagnosemethoden durchgeführt. Ein Score zur Feststellung von respiratorischen Erkrankungen von LOVE et al. (2014) wurde ausführlich auf seine Sensitivität und im Vergleich mit anderen Scores getestet. In der Einleitung stellt der Autor fest, dass kein anderes praktikables Scoringssystem in diesem Bereich bisher validiert wurde. Jedoch wurde in dieser Studie nicht überprüft, ob er auch bei der Anwendung durch verschiedene Tierärzt*innen die gleichen Ergebnisse liefert. Eine Studie von 2019 bewertet die Inter-Observer-Übereinstimmung bei der Lungenauskultation von Kälbern und zeigt auch einen Vergleich der Auskultation zu den pathologischen Befunden von Kälberlungen

(PARDON et al., 2019). Die Ergebnisse führen zu der Einschätzung, dass unreflektierte Beurteilungen der Kälber zu großen Fehlerspannen führen können. Zwei Studien überprüften die Übereinstimmung der Observer in Bezug auf zusammengesetzte Gesundheitsscorings für respiratorische Erkrankungen. Bei BUCZINSKI et al. (2016) erhoben dafür Studierende der Tiermedizin nach einer zweistündigen Einführung Körpertemperatur, Husten, Augen- und Nasenausfluss sowie die Ohrstellung der Kälber. Die Bewertung der Kälber bei AMRINE et al. (2013) erfolgte durch neun sehr erfahrene Tierärzt*innen über einen modifizierten Score von PERINO und APLEY (1998), der wörtliche Beschreibungen verwendet:

1 = usual behavior

2 = slight illness (mild signs of depression or cough)

3 = moderate illness (severe signs of depression, labored breathing, or cough)

4 = severe illness (moribund or failure to respond to human approach)

Die Autoren beider Studien stellten fest, dass die Reliabilität zwischen den Observern sehr gering ist.

3.3. Validierte Scores bei Kühen

In der Bestandsbetreuung und Einzeltierbeurteilung häufig benutzte validierte Scores für Kühe, die ausführlich in Bezug auf die Observer-Reliabilität validiert wurden, sind zum Beispiel der Body condition score (EDMONSON et al., 1989), der Teat end callosity score (NEIJENHUIS et al., 2000) und das Lameness scoring von WELLS et al. (1993), aus welchem der Sprecher Score (SPRECHER et al., 1997) entwickelt wurde.

Bei der Beurteilung der Hygiene bei Kühen untersuchten SCHREINER und RUEGG (2003) das Maß der Übereinstimmung zwischen den Beobachtungen. Da das gesamte Scoring durch eine Person erfolgte, wurde hier nur die Intra-Observer-Reliabilität mit einer Kappa-Statistik nach MARTIN et al. (1987) berechnet. Dieser Hygienescore kann zum Beispiel in der Bestandsbetreuung verwendet werden und soll Rückschlüsse auf die Gesundheit der untersuchten Tiere und das Management zulassen. Bei der Schlachttieruntersuchung werden auch Scores zur Hygienebeurteilung verwendet (ANONYM, 1997; VAN DONKERSGOED et al.,

1997), da Verschmutzungen am Schlachttier die Fleischqualität und Lebensmittelsicherheit beeinflussen können (MCEVOY et al., 2000). Ein Beispiel für die Validierung eines Scores am frisch geschlachteten Tier wurde von JORDAN et al. (1999) veröffentlicht, wobei der Fokus auf die Inter-Observer-Reliabilität gelegt wurde.

In einer Studie von DAHLHOFF (2014) wurden der Hygienescore nach PELZER et al. (2007), sowie ein Score modifiziert nach REUBOLD (2003) zu Verletzungen bei Kühen, auf ihre Reliabilität geprüft. Die Inter- und Intra-Observer Reliabilität wiesen eine mittlere bis gute Übereinstimmung auf.

Im Welfare Quality Protokoll (WELFARE QUALITY®, 2009) wird ein Hygienescore für Kühe dargestellt, für den VAN OS et al. (2018) eine Intra-Observer-Reliabilität mit mindestens 80 % feststellten. Es wurden dafür sowohl Beurteilungen von Fotos, Videoclips als auch live Beobachtungen mit je mindestens 20 Kühen miteinbezogen. In einer Übersicht zu den Methoden des Welfare Quality Protokolls kommen KNIERIM und WINCKLER (2009) zu der Feststellung, dass sowohl die Validität als auch die Reliabilität für viele weitere Bereiche nicht untersucht oder nachgewiesen wurden.

III. PUBLIKATION

DESIGN AND VALIDATION OF A HYGIENE SCORE FOR CALVES

L.M. Kellermann,* A. Rieger,* G. Knubben-Schweizer,* and M. Metzner*

* Clinic for Ruminants with Ambulatory and Herd Health Services at the Centre
for Clinical Veterinary Medicine, Ludwig-Maximilians-Universität (LMU)
Munich, Sonnenstrasse 16, 85764 Oberschleissheim, Germany

Journal of Dairy Science (2020), Vol. 103, Issue 4, p3622–3627



J. Dairy Sci. 103:3622–3627
<https://doi.org/10.3168/jds.2019-17536>
 © American Dairy Science Association®, 2020.

Short communication: Design and validation of a hygiene score for calves

L. M. Kellermann,*  A. Rieger,  G. Knubben-Schweizer,  and M. Metzner 

Clinic for Ruminants with Ambulatory and Herd Health Services, Centre for Clinical Veterinary Medicine, Ludwig-Maximilians-Universität Munich, Sonnenstrasse 16, 85764 Oberschleissheim, Germany

ABSTRACT

The objective of this study was the establishment and validation of a scoring system for calf dirtiness. Defined areas of the belly, side, and rear of the calves were scored according to the percentage of dirtiness: score 1 ranges from 0 to 10% of the area (no or little soiling), score 2 from over 10 to 30% (medium soiling), and score 3 applies to soiling of more than 30% of the area (heavy soiling). The scores of the individual regions were summed up to yield the calf hygiene score (CHS) ranging from 3 to 9. The validation of the CHS was performed by 5 veterinarians on 42 calves. It was validated for its inter- and intra-observer-reliability and against a standard method created by image processing of photographs of calves. The agreement between the observers and the standard method was weak to moderate with kappa values from 0.58 to 0.67. Inter- and intra-observer reliability resulted in a moderate to strong agreement with 29 of 36 kappa values between 0.60 and 0.89. The CHS was thus successfully validated as reliable and could be a useful tool for herd health management.

Key words: calf, hygiene, score, reliability, validation

Short Communication

The handling and housing of livestock is of key interest to politicians, farmers, and consumers. Veterinarians aim to improve animal welfare and to prevent disease, while balancing economic, ethical, and food safety interests. Cleanliness and hygiene are important factors in achieving these aims. Different methods have been used to describe the effect of hygienic conditions on calves' health. Roth et al. (2009) and Calvo-Lorenzo et al. (2016) described the dirtiness caused by feces by observing the amount and consistency of the feces. Jorgensen et al. (2017) additionally differentiated between different regions (perianal region, underside of the tail,

and tailhead). Panivivat et al. (2004) used a scoring system where tail head region, thighs, and legs were counted when soiled with manure. This scoring has been applied by Sutherland et al. (2013, 2014, 2017) and Graham et al. (2018). Other studies assessed different body regions and their degree of dirtiness (score or percentage): Terosky et al. (1997) and Klein-Jöbstl et al. (2014) focused on legs, thighs, and the belly, whereas Webster et al. (1985) also included the back and the head. Lundborg et al. (2005) used a very detailed scoring for 8 different body parts of calves, where the percentage of contamination results in a score. To prove that a scoring system produces reliable and valid data, it should be statistically validated, including tests for inter-observer and intra-observer reliability (Meagher, 2009). In dairy science we use various rating systems with a statistically proven functionality, for example, BCS (Edmonson et al., 1989) or lameness scoring (Wells et al., 1993). To the best of our knowledge, no statistically tested scoring system for the hygiene of calves has been published so far. The purpose of this study was therefore to design a calf hygiene score (CHS) that is easy to apply in the field and has a high inter- and intra-observer agreement. Furthermore, the CHS was validated against a standard method (SM) created in an image-based evaluation system. In this short communication, we explain the design of this CHS and present the results of the validation.

For the scoring system, 3 risk zones (RZ) are considered, and each RZ is assigned an individual score. The summation of the RZ results in the CHS. The 3 evaluated RZ are the belly (RZB), the side (RZS), and the rear (RZR; Figure 1). For the evaluation of RZB and RZS, the examiner is on the side of the calf at an angle of 90° and in the middle of the longitudinal axis of the calf, at a distance of 0.5 to 2 m. The examiner should be in a squatting position to keep the eye level at the height of the back line. For the evaluation of RZR, the examiner is at the same distance and posture, located behind the calf in the middle of its longitudinal axis. The RZB is defined as the visible area below a line from the elbow to the knee fold, cranially limited by the foreleg and caudally by the hindleg (Figure 1a). The RZS

Received September 2, 2019.

Accepted December 2, 2019.

*Corresponding author: l.kellermann@lmu.de

is the visible area caudally of a line orthogonal to the ground through the knee fold. The tail is excluded and only the lateral view is evaluated (Figure 1b). The RZR is the visible area from the tail attachment down to the proximal end of the tarsal joints and the complete tail (including the switch; Figure 1c). The evaluation of the percentage of dirtiness of the surfaces is expressed as a score from 1 to 3 for each individual RZ. Score 1 ranges from 0 to 10% of the area (no or little soiling), score 2 from over 10 to 30% (medium soiling), and score 3 applies to soiling of more than 30% of the area (heavy soiling). All types of soiling must be considered, except loose straw or loose particles in the haircoat, freshly defecated excrements, or a wet coat assumed to be due to rain or drinking water. The RZB and RZS are examined from the right and the left side of the calf and the results of the more soiled RZB and RZS are each documented, such that only the dirtier side of the calf contributes to the score. Schematic drawings were handed out as a scorecard to the examiners for reference (see Figure 1). The scores of the individual regions were summed up to yield the CHS ranging from 3 to 9.

A total of 42 Holstein calves were included in the study for evaluation of inter-observer reliability, 41 of them also for the evaluation of intra-observer reliability. One calf was excluded from the intra-observer reliability calculation because it was newborn and changed its appearance from wet to dry between the scoring rounds, so that a comparison within the different rounds of each observer could not be performed. The calves were suckling calves at the age of 0 to approximately 12 wk housed in individual pens or in group igloos, and all belonged to a commercially operating dairy farm in Bavaria. The evaluation took place on 1 d, with 5 veterinarians (observers). Each calf was rated by each observer individually 3 times in a different order within 4 h to create a data set for calculation of inter- and intra-observer reliability. Additionally, within this time window standardized photographs of all calves were taken for each RZ including both sides, from the same observation position as defined above. All 210 pictures were used to establish a SM for the validation. Photoshop CC 2017 (Adobe Systems Inc., San Jose, CA) was used to calculate the actual proportions of soiled surfaces in the photographed RZ of the calves. In brief, the specified area of the respective RZ and the soiled areas on it were marked manually with a polygon-tool and the numbers of pixels were displayed. The sum of pixels of all soiled areas was calculated and divided by the number of pixels of the RZ. The percentage of pixels belonging to the soiled area was assigned to a score and compared with the score awarded for this region by the observers. The verification of the schematic drawings of Figure 1 for a correct percentage of soiled area

within the respective RZ was performed with the same graphical method.

A kappa statistic according to the method of Ruddat (Ruddat et al., 2014) using RStudio version 1.1.453 (RStudio Inc., Boston, MA) with R version 3.4.4 (R Core Team, Vienna, Austria) was calculated. This kappa method is suitable for ordinal data and for more than 2 observers. Kappa statistics for the ratings of the individual RZ and the CHS were calculated. The global kappas with 95% confidence interval are displayed in Table 1 for each observer (intra-observer) and in Table 2 for each round (inter-observer). The reference results generated by image editing were included as a sixth observer and represent the SM. Results of the respective exclusion test are also presented in Table 2 as a kappa statistic with 95% confidence interval to identify if the observers 1–5 agree with the SM. Kappa values can range from -1 to 1 . Cut-off values for kappa are arbitrary (Landis and Koch, 1977), but for health-related scores the interpretation according to McHugh (2012) can be used: kappa values above 0.80 represent strong observer agreement, values from 0.60 to 0.79 indicate moderate agreement, and values from 0.40 to 0.59 represent weak agreement. Lower values would indicate no or minimal agreement.

The results show that the application of the CHS was highly reliable: 29 of 36 determined kappa values show a moderate to strong inter- or intra-observer agreement. For the CHS, the average intra-observer agreement was a kappa of 0.83 and the average inter-observer agreement was 0.77. Within the RZ, only the RZB showed a noticeably weaker agreement. The validation against the SM showed a weak to moderate agreement. The mean value of all assigned CHS scores for each calf (from each observer for 3 rounds) was calculated and rounded to full digits. In comparison to these 42 CHS mean values, the CHS score assigned by the SM yielded 81% higher and 19% equal values. This was also performed for the scores of each RZ, which resulted in 26% higher and 74% equal scores for RZB and 60% higher and 40% equal scores for RZS and RZR. No SM score was lower than the mean value of the observers' CHS scores. Further evaluations of the SM scoring results show that 22 of 42 calves would obtain a lower score for the RZB or RZS or even both RZ if they would have been scored on the cleaner side.

We identified several factors that could contribute to the weaker agreements detected for the RZB and for the differences between the SM and observers. Especially for the RZB, the specified body position is very important. We cannot exclude the possibility that the observers changed the exact height of their eye position slightly during the assessment. Additionally, changes to the body position of the calf alter the observable sur-

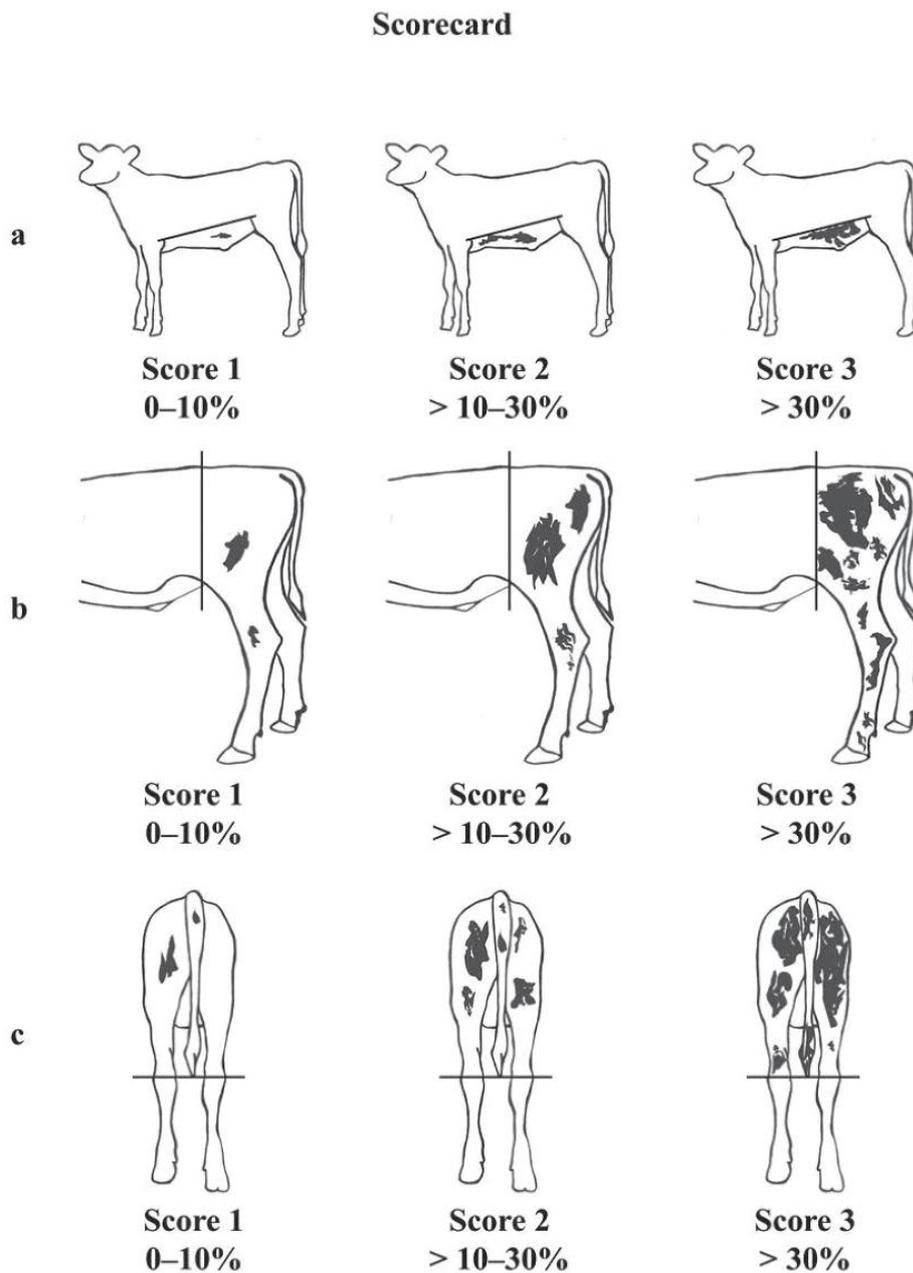


Figure 1. Scorecard for the use of the calf hygiene score (CHS). Each risk zone (a: belly, b: side, and c: rear) is assigned a score from 1 to 3. The CHS is the sum of all 3 risk zone ratings.

face particularly at the RZB. The imprecision and inaccuracy of human perception is a strong but inevitable source of error in optical scoring. The selected categories of percent were chosen because they correspond to

what we encountered in the field considering calves on farms. The advantage of a method using percent for assessments is that the number itself has an unambiguous value. Methods using the size of the hand amplify

Table 1. Global kappa values (95% CI) of the intra-observer validation of a calf hygiene score (CHS) by 5 observers in 3 rounds considering 3 risk zones: belly (RZB), side (RZS), and rear (RZR) of n = 41 calves

Assessed zone	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5
RZB	0.68 (0.46; 0.91)	0.51 (0.23; 0.78)	0.62 (0.40; 0.83)	0.78 (0.63; 0.93)	0.43 (0.13; 0.73)
RZS	0.79 (0.67; 0.91)	0.81 (0.70; 0.92)	0.75 (0.56; 0.93)	0.78 (0.68; 0.89)	0.56 (0.37; 0.74)
RZR	0.85 (0.76; 0.94)	0.84 (0.73; 0.94)	0.82 (0.72; 0.94)	0.82 (0.69; 0.94)	0.68 (0.55; 0.80)
CHS ¹	0.88 (0.80; 0.95)	0.87 (0.83; 0.91)	0.82 (0.70; 0.95)	0.89 (0.83; 0.95)	0.70 (0.58; 0.83)

¹CHS = sum of rating from RZB, RZS, and RZR.

the inaccuracy that both the hand and the evaluated animal can be of very different sizes. Descriptions such as “inconspicuous,” “little,” or simply “soiled” are not strictly determined and are highly subjective. To facilitate the visual assessment, charts have been used as an additional tool for reference (Edmonson et al., 1989; Cook, 2002; Reneau et al., 2005). The most common classification of burn wounds in human medicine is described by the Lund and Browder chart (Lund and Browder, 1944), where they divide the body surface in different parts of different importance. Likewise, we defined the areas to be assessed for the CHS precisely on the scorecard and with words, since the mere description of “hind” or “side” could lead to mistakes. The area of RZS and RZB overlap to a small extent. Due to the shape of the calf, we did not obtain clearer boundaries. The overlap was not taken into account in our calculations. Our experience has been that the scoring of newborn calves leads to difficulties (e.g. when they are still wet). We recommend excluding calves born in the last 24 h from this hygiene scoring. The colorization of calves involves the problem that dirt could be less visible on black parts. In order to keep the scoring simple, there were no measurements of the color proportions of the coat. To detect the condition of the coat, suf-

ficient lighting is necessary. We recommend performing the evaluation only if one could read a newspaper text at the calf’s location without any problems. The use of light sources such as flashlights or the evaluation in a dark environment have not been tested and may lead to different scoring results. A convincing scoring system should assess the hygienic condition of the calf as a whole. Mud, urine, milk residues, dust, and other materials should not be ignored because the impact on health is not clear. The definition of dirt for our scoring system is therefore not limited to soiling by feces and was clearly described. To obtain reliable results, it is recommended that the observer try to wipe loose parts off the coat once. Loose straw often hides excrement to which the straw adheres. Nevertheless, the selection of the relevant parts is sometimes challenging (e.g. when only the tips of the hairs are dirty or only slight color changes are visible). Those difficulties were the same for the evaluation of the pictures as for the observer on the farm. However, when creating the SM, we evaluated photographs and were able to zoom in and out in various ways, whereas the observer assessed moving calves in a short period. As a result, it is likely that in the analysis of photographs, even small dirt spots were easier to recognize and therefore a higher percentage of

Table 2. Global kappa values (95% CI) of the inter-observer validation of a calf hygiene score (CHS) by 5 observers in 3 rounds considering 3 risk zones: belly (RZB), side (RZS), and rear (RZR) of n = 42 calves, and a comparison with the standard method (SM)

Evaluation unit	RZB	RZS	RZR	CHS ¹
Round 1 (observer 1–5)	0.57 (0.41; 0.73)	0.67 (0.52; 0.81)	0.81 (0.72; 0.90)	0.82 (0.76; 0.90)
Round 2 (observer 1–5)	0.59 (0.41; 0.77)	0.69 (0.56; 0.81)	0.63 (0.50; 0.75)	0.74 (0.63; 0.85)
Round 3 (observer 1–5)	0.56 (0.43; 0.70)	0.70 (0.58; 0.81)	0.68 (0.57; 0.79)	0.76 (0.69; 0.84)
Total ² (observer 1–5)	0.58 (0.49; 0.66)	0.67 (0.59; 0.76)	0.70 (0.64; 0.77)	0.77 (0.71; 0.84)
Total (observer 1–5) plus SM ³	0.58 (0.49; 0.67)	0.60 (0.52; 0.68)	0.60 (0.53; 0.67)	0.66 (0.60; 0.73)

¹CHS = sum of rating from RZB, RZS, and RZR.

²Total = values for rounds 1, 2, and 3 combined.

³The SM was created by image processing of photographs of calves.

dirty areas was calculated. The kappa values of total (observer 1–5) plus SM demonstrate that the scoring still provides a sufficiently realistic image. Whilst the SM gives a more accurate estimate of absolute soiling, hygiene scoring is normally applied in order to assess animals relative to one another. Thus, the observer-based CHS is a useful tool for assessing hygiene in the field and is reliable as shown by the observer agreement values. Graham et al (2018) received a weighted kappa of only 0.22 for agreement between the scoring for diarrhea and the scoring for hide cleanliness. The CHS should therefore not be used to predict diarrhea in calves. Under field conditions, calves would often only be seen from one side. This side can be the cleaner or the dirtier side, and that for RZB and RZS. For reliable scoring it is necessary to examine both sides of the calves. Doing otherwise could change the scoring results to a lower CHS, as described in the results.

Goldberg et al. (2014) compared a mobile app with the chart previously mentioned by Lund and Browder (1944). Users marked the area by hand on a 3-dimensional model in the app, which then calculated the surface area. The results obtained were more accurate and the app was reported to be easier to apply than the conventional method. The development of an app could also be useful in the field of calf hygiene. The definitions of the RZ and other rules set up for the CHS could facilitate the creation of such an app in the future.

In conclusion, we have created an easy-to-use and accurate score for the hygiene of calves. By defining dirtiness, the regions that are assessed, and the method of how to score them, we obtained good observer agreement. Through this validation, the CHS is qualified as a reliable method to assess the hygienic condition of calves in monitoring herd health. The extent of the relationship between the score and calf health should be evaluated in further studies.

ACKNOWLEDGMENTS

Inga Ruddat (Hannover, Germany) provided us with valuable information about her methods of statistical analysis. We are grateful to Helen Fox (University of Cambridge alumnus) for her proofreading and advice. We also thank the observers of our validation for their time and high degree of concentration. The authors have not stated any conflicts of interest.

REFERENCES

- Calvo-Lorenzo, M. S., L. E. Hulbert, A. L. Fowler, A. Louie, L. J. Gershwin, K. E. Pinkerton, M. A. Ballou, K. C. Klasing, and F. M. Mitloehner. 2016. Wooden lutch space allowance influences male Holstein calf health, performance, daily lying time, and respiratory immunity. *J. Dairy Sci.* 99:4678–4692. <https://doi.org/10.3168/jds.2016-10888>.
- Cook, N. B. 2002. The influence of barn design on dairy cow hygiene, lameness, and udder health. Pages 97–103 in *Proc. Am. Assoc. Bov. Pract., Madison, WI. Am. Assoc. Bov. Pract., Rome, GA.*
- Edmonson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein dairy cows. *J. Dairy Sci.* 72:68–78. [https://doi.org/10.3168/jds.S0022-0302\(89\)79081-0](https://doi.org/10.3168/jds.S0022-0302(89)79081-0).
- Goldberg, H., J. Klaff, A. Spjut, and S. Milner. 2014. A mobile app for measuring the surface area of a burn in three dimensions: Comparison to the Lund and Browder assessment. *J. Burn Care Res.* 35:480–483. <https://doi.org/10.1097/BCR.0000000000000037>.
- Graham, A. N., D. L. Renaud, T. F. Duffield, and D. F. Kelton. 2018. Short communication: Calf cleanliness does not predict diarrhea upon arrival at a veal calf facility. *J. Dairy Sci.* 101:3363–3366. <https://doi.org/10.3168/jds.2017-14113>.
- Jorgensen, M. W., A. Adams-Progar, A. M. de Passillé, J. Rushen, S. M. Godden, H. Chester-Jones, and M. I. Endres. 2017. Factors associated with dairy calf health in automated feeding systems in the Upper Midwest United States. *J. Dairy Sci.* 100:5675–5686. <https://doi.org/10.3168/jds.2016-12501>.
- Klein-Jöbstl, D., M. Iwersen, and M. Drillich. 2014. Farm characteristics and calf management practices on dairy farms with and without diarrhea: A case-control study to investigate risk factors for calf diarrhea. *J. Dairy Sci.* 97:5110–5119. <https://doi.org/10.3168/jds.2013-7695>.
- Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. <https://doi.org/10.2307/2529310>.
- Lund, C. C., and N. C. Browder. 1944. The estimation of areas of burns. *Surg. Gynecol. Obstet.* 79:352–358.
- Lundborg, G. K., E. C. Svensson, and P. A. Oltenacu. 2005. Herd-level risk factors for infectious diseases in Swedish dairy calves aged 0–90 days. *Prev. Vet. Med.* 68:123–143. <https://doi.org/10.1016/j.prevetmed.2004.11.014>.
- McHugh, M. L. 2012. Interrater reliability: The kappa statistic. *Biochem. Med. (Zagreb)* 22:276–282. <https://doi.org/10.11613/BM.2012.031>.
- Meagher, R. K. 2009. Observer ratings: Validity and value as a tool for animal welfare research. *Appl. Anim. Behav. Sci.* 119:1–14. <https://doi.org/10.1016/j.applanim.2009.02.026>.
- Panivivat, R., E. B. Kegley, J. A. Pennington, D. W. Kellogg, and S. L. Krumpelman. 2004. Growth performance and health of dairy calves bedded with different types of materials. *J. Dairy Sci.* 87:3736–3745. [https://doi.org/10.3168/jds.S0022-0302\(04\)73512-2](https://doi.org/10.3168/jds.S0022-0302(04)73512-2).
- Reneau, J. K., A. J. Seykora, B. J. Heins, M. I. Endres, R. J. Farnsworth, and R. F. Bey. 2005. Association between hygiene scores and somatic cell scores in dairy cattle. *J. Am. Vet. Med. Assoc.* 227:1297–1301. <https://doi.org/10.2460/javma.2005.227.1297>.
- Roth, B. A., N. M. Keil, L. Gyax, and E. Hillmann. 2009. Influence of weaning method on health status and rumen development in dairy calves. *J. Dairy Sci.* 92:645–656. <https://doi.org/10.3168/jds.2008-1153>.
- Ruddat, I., B. Scholz, S. Bergmann, A. L. Buehring, S. Fischer, A. Manton, D. Prengel, E. Rauch, S. Steiner, S. Wiedmann, L. Kreienbrock, and A. Campe. 2014. Statistical tools to improve assessing agreement between several observers. *Animal* 8:643–649. <https://doi.org/10.1017/S1751731113002450>.
- Sutherland, M. A., M. Stewart, and K. E. Schütz. 2013. Effects of two substrate types on the behaviour, cleanliness and thermoregulation of dairy calves. *Appl. Anim. Behav. Sci.* 147:19–27. <https://doi.org/10.1016/j.applanim.2013.04.018>.
- Sutherland, M. A., G. M. Worth, C. Cameron, C. M. Ross, and D. Rapp. 2017. Health, physiology, and behavior of dairy calves reared on 4 different substrates. *J. Dairy Sci.* 100:2148–2156. <https://doi.org/10.3168/jds.2016-12074>.
- Sutherland, M. A., G. M. Worth, and M. Stewart. 2014. The effect of rearing substrate and space allowance on the behavior and physiol-

- ogy of dairy calves. *J. Dairy Sci.* 97:4455–4463. <https://doi.org/10.3168/jds.2013-7822>.
- Terosky, T. L., L. L. Wilson, C. L. Stull, and W. R. Stricklin. 1997. Effects of individual housing design and size on special-fed Holstein veal calf growth performance, hematology, and carcass characteristics. *J. Anim. Sci.* 75:1697–1703. <https://doi.org/10.2527/1997.7571697x>.
- Webster, A. J., C. Saville, B. M. Church, A. Gnanasakthy, and R. Moss. 1985. Some effects of different rearing systems on health, cleanliness and injury in calves. *Br. Vet. J.* 141:472–483. [https://doi.org/10.1016/0007-1935\(85\)90042-9](https://doi.org/10.1016/0007-1935(85)90042-9).
- Wells, S. J., A. M. Trent, W. E. Marsh, and R. A. Robinson. 1993. Prevalence and severity of lameness in lactating dairy cows in a sample of Minnesota and Wisconsin herds. *J. Am. Vet. Med. Assoc.* 202:78–82.

ORCID

- L. M. Kellermann  <https://orcid.org/0000-0002-3423-8717>
A. Rieger  <https://orcid.org/0000-0002-6826-4497>
G. Knubben-Schweizer  <https://orcid.org/0000-0002-0928-5933>
M. Metzner  <https://orcid.org/0000-0003-0752-6059>

IV. DISKUSSION

Diese Dissertationsarbeit hatte das Ziel, einen Hygienescore für Kälber zu entwerfen und die Anwendung zu validieren. Bei der Literaturrecherche wurden keine validierten Scores für die Verschmutzung von Kälbern gefunden, sowie nur vereinzelt validierte Scores im Bereich der Kälbergesundheit. Der neu entworfene Kälber-Hygienescore (Calf Hygiene Score, CHS) beinhaltet klare Definitionen sowie eine Scorecard, so dass der Score für Anwender*innen verständlich und leicht zu merken ist. Es wurde nicht nur die Übereinstimmung der Observer geprüft, sondern auch eine Methode zur Etablierung eines Standards (SM) umgesetzt und mit den Scores der Observer verglichen. Die Kappa-Werte der Inter- und Intra-Observer Reliabilität zeigen, dass der CHS zuverlässig angewendet werden kann. Die Aussage des Scores wird vorrangig durch die unumgänglichen Nachteile einer optischen Erfassung sowie durch die unterschiedlichsten Licht- und Platzverhältnisse im Feld eingeschränkt. Auf diese möglichen Herausforderungen bei der Anwendung wurde in der Publikation eingegangen.

In der Literaturübersicht ist auffällig, dass nur wenige der aufgefundenen Systeme zur Bewertung von Gesundheits- und Verhaltensparameter vollständig validiert wurden. Dies ist eine Aufgabe, die in Zukunft stärker angegangen werden sollte. Die Anwendung von Scores kann enorme Bedeutung für die beurteilten Tiere, für die Landwirt*innen und sogar auf die Gesetzgebung haben. Bei Nutzung durch den/die bestandsbetreuende/n Tierarzt/Tierärztin werden nach Ergebnis des Scorings (z. B. beim Lahmheits-Scoring) Tiere genauer untersucht, behandelt, oder nicht weiter untersucht. Dies kann einen direkten Einfluss auf die Tiergesundheit und damit auch das Wohlbefinden des Tieres haben, wie auch eine Fehleinschätzung bei falschem Scoring zu vermeidbarem Leiden führen kann (z. B. bei einem Hygienescore). Auszahlungen von Subventionen und Prämien können an Tierwohl Monitoring und Kontrollmaßnahmen gebunden werden, in denen Scores verwendet werden. Werden Scores in wissenschaftlichen Studien ohne Überprüfung ihrer Validität und Reliabilität verwendet, kann dies die Studienergebnisse verzerren. Je nach Zweck der Studie kann dies in Haltungs- und Managementempfehlungen resultieren, oder sogar zur Grundlage von Gesetzen und Verordnungen werden. Weiterführende Studien bauen auf diesem Wissen auf, so dass sich diese Problematik immer weiterzieht. KNIERIM und WINCKLER (2009)

beziehen sich vorrangig auf Erhebungen des WELFARE QUALITY® (2009) Protokolls mit der Feststellung, dass

1. die Zuverlässigkeit für jeden Parameter erhoben werden sollte,
2. ein Standard für die Bewertung der Reliabilität geschaffen werden sollte und dass
3. es Mindestanforderungen in diesem Bereich geben sollte, die ein Parameter nicht unterschreiten darf.

Diese Forderungen treffen auch auf Erhebungen zur Tiergesundheit zu. So könnten bestehende Scoring-Systeme bestätigt, verbessert oder durch neue Methoden ersetzt werden.

Punkt 3. bezieht sich unter anderem darauf, dass die Beurteilung der Ergebnisse zur Übereinstimmung der Observer selbst nicht standardisiert ist. Für den CHS wurde eine Kappa-Berechnung nach RUDDAT et al. (2014) verwendet. Dieses Kappa ist nicht verzerrt und für die Ordinalskala geeignet. Kappa-Werte können Werte von -1 bis +1 aufweisen, und je höher der Wert ist, desto höher ist das Maß der Übereinstimmung. 0 ist das Ereignis, was bei einer völlig zufälligen Bewertung eintreffen würde. Eine weitere Unterteilung der Kappa-Werte wird aus statistischer Sicht als willkürlich angesehen (LANDIS & KOCH, 1977). RUDDAT et al. (2014) schlagen vor, die Mindestanforderung der Kappa-Werte abhängig von der zu messenden Variablen und der vorgefundenen Observer-Übereinstimmung vorzunehmen. Die Kappa-Werte des CHS wurde nach einem Vorschlag von MCHUGH (2012) beurteilt. McHugh begründet die Einteilung damit, dass jede Abweichung von +1 auch das Maß der Ungenauigkeit darstellt. Aus (tier-)medizinischer Sicht ist eine Übereinstimmung von 60 % und weniger problematisch, da somit 40 % und mehr der Ergebnisse nicht übereinstimmen. Je nach bewertetem Kriterium ist das eine nicht akzeptable Fehlerquote, z. B. bei der Einschätzung ob ein Kalb schwer krank ist oder nicht. Im Laborbereich würden Methoden mit so einer hohen Ungenauigkeit auf starke Qualitätsmängel hinweisen. Zwar sind Kappa-Werte per se niedriger als die Angabe in Prozent, dennoch plädiert die Autorin dafür, nicht die empfohlenen Einschätzungen von COHEN (1960) anzuwenden, bei denen ein Kappa von 0,40 (bis 0,60) schon für eine moderate Übereinstimmung steht. Die von McHugh für den CHS übernommene Interpretation ist strenger und es wird erst ab einem Kappa von 0,60 eine moderate Übereinstimmung angenommen. Die Eigenschaft „Verschmutzung“ kann beim

Kalb sehr unterschiedlich ausgeprägt sein. Anhand der Werte fällt dennoch nur ein Observer auf (Observer 5), der in seiner Intra-Observer Beurteilung inkonsistenter als die anderen wirkt. Die Interpretation der Kappa-Werte nach RUDDAT et al. (2014) könnte in Anbetracht dieser Tatsachen auch zu einer ähnlich strengen Kategorisierung führen, da die Werte der Übereinstimmung insgesamt groß sind.

Um die Validität des CHS nachzuweisen, wurde dieser mit einer Standard Methode (SM) verglichen. Diese SM könnte theoretisch als Goldstandard bezeichnet werden, da momentan keine bessere Methode vorhanden ist, die Verschmutzung an Kälbern exakt wiederzugeben. Sie wurde mit einer sorgfältigen, möglichst einheitlichen (da durch eine einzelne Person durchgeführten) Arbeitsweise erstellt. Jedoch wurde die SM selbst nicht validiert. Zudem soll und kann sie auch nicht als tatsächlich durchführbarer Goldstandard eingesetzt werden, da die SM die Beurteilung von Kälbern durch die händische Auswertung von Fotografien mit einer Bildverarbeitungssoftware umfasst. Dies ist in der praktischen Anwendung nicht möglich. Die SM dient somit lediglich als Nachweis, dass der CHS auch das misst was er messen soll: die definierte Verschmutzung am Kalb. Durch technische Entwicklungen könnte jedoch sowohl die SM optimiert werden als auch ein neuer Goldstandard entstehen. In der Biologie werden Bildverarbeitungsprogramme verwendet, um Farbunterschiede von Zellkulturen herauszustellen und die Flächen zu berechnen (SCHINDELIN et al., 2012). Ob diese Programme auf unter Feldbedingungen erstellten Fotografien von gefleckten Kälbern angewendet werden könnten, ist nicht bekannt. Zukünftige Entwicklungen in der Kamerasensorik könnten aber auch Verschmutzungen nicht nur durch Farb- sondern auch durch Dichte- und Temperaturunterschiede wahrnehmen. Da jedoch auch dann Bewertungen vorrangig aus der Sicht des Beobachters/der Beobachterin geschehen, sind die Definitionen der Risikozonen (RZ) auch für zukünftige Erhebungen sinnvoll, sowie auch die Umwandlung der prozentualen verschmutzten Flächen in Scores.

Das Scoring des Haarkleids an den Karpalgelenken stellt höchstwahrscheinlich auch eine Möglichkeit dar, um auf die Haltungsbedingungen rückschließen zu können. Die Idee zu dieser Methode stammt von Herrn Pelzer (Leiter des Sachbereiches Rinderhaltung des Versuchs- und Bildungszentrums Haus Düsse der Landwirtschaftskammer Nordrhein-Westfalen) und wurde sowohl bei der Validierung als auch der Datenerhebung bei den in der PraeRi untersuchten Kälbern

im bayerischen Studiengebiet angewandt. Eine Auswertung dazu steht noch aus. Die Autorin hat sich dazu entschieden, diesen Parameter getrennt von dem in dieser Arbeit dargestellten Scoring auszuwerten, da er sich erheblich von dem CHS unterscheidet. Die Bewertung erfolgt nicht nach Prozent der verschmutzten Fläche, da die Karpalgelenke dafür eine zu kleine Fläche bieten. Es stehen zudem Untersuchungen aus, ob die Verschmutzung am Carpus mit der Verschmutzung des Kalbes korreliert. Möglicherweise ist diese Methode sensibler als der CHS, da das Karpalgelenk beim Abliegen und Aufstehen in den Untergrund hineingedrückt wird und auch bei frisch eingestreuten Bereichen eine darunter liegende feuchte Schicht das Haarkleid optisch verändern kann.

In der Umsetzung wird die Verwendung von Scores schnell ungenau, da diese auch in der einfachsten Version als aufwändig empfunden werden. Die in der Publikation aufgeführten Regeln und Empfehlungen sollten jedoch möglichst genau eingehalten werden. Schlechtere Lichtverhältnisse oder eine falsche Position des Beobachters können sich stark auf die Ergebnisse des Scorings auswirken. Gerade bei der Durchführung von Studien sollten sich die Observer regelmäßig abgleichen und ihre Genauigkeit prüfen. Schon PERINO und APLEY (1998) stellten fest, dass die Anwendung von Beurteilungsschemata kompromisslos sein muss, damit die Ergebnisse für möglicherweise ökonomisch (und dadurch auch tierwohl-) relevante Aussagen herangezogen werden können.

Durch die Anwendung des CHS in der deutschlandweiten Prävalenzstudie Rind (PraeRi) ergeben sich vielfältige Möglichkeiten, um seine Aussagekraft und Anwendung noch zu erweitern. In dem Verbundprojekt „PraeRi: Tiergesundheit, Hygiene und Biosicherheit in deutschen Milchviehbetrieben – eine Prävalenzstudie“ untersuchten Wissenschaftlerinnen und Wissenschaftler der Stiftung Tierärztliche Hochschule Hannover, der Freien Universität Berlin und der Ludwig-Maximilians-Universität München die Tiergesundheit in 750 deutschen Milchkuhbetrieben. Das Projekt läuft von 2016 bis 2020 und wird vom Bundesministerium für Ernährung und Landwirtschaft gefördert. Durch die große Anzahl an Betrieben und deren repräsentative Verteilung kann die Studie aufzeigen, wie verschmutzt die Kälber in Deutschland sind. Man kann für einen Betrieb aus allen gescorten Kälbern einen Betriebs-CHS berechnen (Mittelwert), dann ist diese Aussage sowohl auf der Einzeltier- als auch auf der Betriebsebene möglich. Mit den Daten zur Verteilung der vergebenen Scores kann eine Stichprobenberechnung

durchgeführt werden. Die Berechnung von Mindesttierzahlen für Stichproben ist wesentlich, da der Zeitfaktor eine Hauptrolle bei der Praktikabilität von Scores spielt (KNIERIM & WINCKLER, 2009). Der Tierarzt/die Tierärztin oder auch der Landwirt/die Landwirtin selbst kann dann sowohl Einzeltiere als auch bei größeren Betrieben durch repräsentative Stichproben den gesamten Kälberbestand eines Betriebes bewerten. Diese Bewertung kann mit der Verteilung des CHS in der PraeRi verglichen werden. Dadurch kann man die Betriebe künftig benchmarken. Als zu stark verschmutzt kann eine Positionierung im obersten Quartil angesehen werden, also wenn der CHS des Kalbes oder des Kälberbestands zu den 25 % höchsten CHS Scores gehören. Benchmarking hat sich als sehr wirksam erwiesen, um Landwirt*innen eine objektive Sicht auf ihren Betrieb zu ermöglichen und ihnen einen zusätzlichen Anreiz zu geben, sich zu verbessern (SUMNER et al., 2018).

Je nach Management, Arbeitsbelastung und Einstellung von Tierhalter*innen leben die Kälber in einer entsprechend mehr oder weniger stark verschmutzten oder sauberen Umgebung. In der Tierschutz-Nutztierhaltungsverordnung (Abschnitt 2 § 5) steht: „1. Kälber dürfen nicht mehr als unvermeidbar mit Harn oder Kot in Berührung kommen; ihnen muss im Stall ein trockener Liegebereich zur Verfügung stehen.“ Hieraus resultiert erstens, dass die Haltungsumgebung der Kälber auf das Vorhandensein trockener Liegebereiche geprüft werden muss. Zweitens muss auch objektiv beurteilt werden, wie sehr ein Kalb mit Kot und Harn in Berührung gekommen ist. Weiterhin stellt sich die Frage, wie „nicht mehr als unvermeidbar“ definiert werden kann. Für diese Beurteilungen bietet sich der CHS und das oben beschriebene Benchmarking der Betriebe an. In Studien zur Kälbergesundheit ist die optische Bewertung der Verschmutzung der Einstreu kritisch zu sehen, da dafür keine validierte Methode bekannt ist. Besser eignen sich Untersuchungen zum Risiko bei verschiedenen Einstreumaterialien und der Nachweis der bakteriellen Kontamination (MCGUIRK, 2008), welcher jedoch aus Kostengründen selten angewendet wird. Möglicherweise kann durch den CHS die Bewertung der Einstreu qualitativ aufgewertet werden, da er längerfristige Verschmutzungen bewertet, auch wenn z. B. direkt vor dem Scoring die Einstreu ausgewechselt wurde. Aus den Daten der PraeRi könnte dafür untersucht werden, ob die erhobene Verschmutzung der Einstreu mit dem CHS korreliert.

Es ist im Rahmen des Kälberscorings der PraeRi mitunter aufgefallen, dass die

Tiere weniger verschmutzt waren als es der Umgebung nach anzunehmen gewesen wäre. Nicht zu vernachlässigen ist die Möglichkeit der Kälber zur Kompensation verschmutzter Haltungseinrichtungen. Gesunde Kälber können eine starke Kompensationsfähigkeit besitzen und sich gezielt durch Lecken selbst reinigen, während erkrankte Kälber dagegen ihre Körperpflege verringern (BORDERAS et al., 2008). Kälber suchen sich die trockensten Bereiche der Liegeflächen aus (CAMILOTI et al., 2012), wobei in einer Gruppe dann wahrscheinlich erkrankte oder schwache Kälber vorrangig auf verschmutzten Flächen liegen müssen. Werden Kälber auf sehr kleinem Raum gehalten, können sie diese Möglichkeiten nicht wahrnehmen und entsprechend gehaltene Tiere wiesen in Studien auch eine größere Verschmutzung auf (CALVO-LORENZO et al., 2016). Diese Zusammenhänge sollten bei künftigen Untersuchungen zum CHS und Erkrankungen bei Kälbern beachtet werden, denn die Kausalität kann theoretisch in beide Richtungen erfolgen: Erkrankte Kälber könnten stärker verschmutzt sein, da sie sich weniger sauber halten können. Andersherum könnten aber auch Kälber erkrankt sein, weil sie in einer sehr verschmutzten Haltungsumgebung leben. Ein möglicher Einfluss sowohl auf die Gesundheit von Kälbern als auch auf ihre Verschmutzung kann das Management des Betriebs sein. So muss nicht die Verschmutzung selbst einen Einfluss haben, sondern sie ist ein Zeichen von falschem Management, was sich natürlich auch auf andere Bereiche der Kälberhaltung und -gesundheit auswirkt. Der CHS kann somit eine Möglichkeit darstellen das Haltungsmanagement auf dem Betrieb zu beurteilen.

V. ZUSAMMENFASSUNG

Laura Kellermann (2020)

Design und Validierung eines Hygienescores für Kälber

Eine gute Kälberhaltung ist das Fundament einer gesunden Milchviehherde. Das Aufzuchtmanagement von Kälbern wirkt sich auf ihr Wachstum, ihre Gesundheit und ihr Wohlbefinden aus. Wissenschaftliche Studien können den Einfluss von verschiedenen Faktoren sichtbar machen und es Landwirt*innen und Tierärzt*innen somit ermöglichen, wichtige Bereiche hinsichtlich Tierwohl und Krankheitsprävention weiterzuentwickeln. Ein Aspekt ist die saubere Aufstallung von Kälbern, damit sie möglichst wenig dreckig werden und trockene Liegeflächen zur Verfügung haben. Eine Überprüfung des Hygienezustands kann durch die Beurteilung der Liegeflächen oder aber am Tier direkt erfolgen. Eine solche Beurteilung sollte grundsätzlich statistisch überprüft sein, damit man nachvollziehen kann ob daraus valide und zuverlässig reproduzierbare Daten entstehen.

Um Verschmutzungen an Kälbern beurteilen zu können, wurde ein Hygienescore erstellt und validiert. Es werden drei Risikozonen betrachtet und die Scores der drei Zonen ergeben summiert den Calf Hygiene Score (CHS). Die definierten Risikozonen befinden sich an Bauch, Flanke und kaudaler Seite des Kalbes und werden nach dem Prozentsatz der Verschmutzung bewertet: Score 1 wird bei keiner oder geringer Verschmutzung (von 0 bis 10 % der Fläche) vergeben, Score 2 bei mittlerer Verschmutzung von über 10 bis 30 %. Score 3 steht für starke Verschmutzungen von mehr als 30 % der Fläche. Der aus den Bewertungen der Risikozonen resultierende CHS kann einen Wert von 3 bis 9 annehmen. Die Validierung des CHS wurde von fünf Tierärzt*innen ausgeführt, die jeder 42 Kälber beurteilten. Es wurde die Inter- und Intra-Observer-Reliabilität untersucht sowie eine Validierung anhand einer Standard Methode vorgenommen. Dafür wurde durch Bildbearbeitung von Fotos der Kälber festgelegt, wie stark die tatsächliche Verschmutzung der Risikozonen ist. Die Übereinstimmung zwischen den Observern und der Standardmethode war schwach bis moderat mit Kappa-Werten von 0,58 bis 0,67. Hierbei ist die größte Fehlerquelle vermutlich die

optische Erfassung von Flächen, mit der die menschliche Wahrnehmung große Schwierigkeiten hat. Die Werte der Inter- und Intra-Observer-Reliabilität zeigen eine moderate bis starke Übereinstimmung mit 29 von 36 Kappa-Werten zwischen 0,60 und 0,89. Der Hygienescore für Kälber wurde somit als zuverlässig validiert und kann zukünftig ein nützliches Werkzeug für die Bestandsbetreuung darstellen.

Im Hinblick auf das Wohlbefinden der Kälber und im Einklang mit der aktuellen Tierschutzgesetzgebung ist die Beurteilung von Kälbern mit dem CHS sinnvoll. Tierhalter*innen entstehen jedoch höhere Kosten und ein größerer Zeitaufwand, je sauberer die Kälber sein sollen. Durch die Ergebnisse künftiger Untersuchungen zum Zusammenhang zwischen Gesundheit und Verschmutzung der Kälber könnten Berater*innen und Tierärzt*innen weitere überzeugende Argumente für den Dialog mit Landwirt*innen erhalten.

VI. SUMMARY

Laura Kellermann (2020)

Design and validation of a hygiene score for calves

Proper rearing of calves is the foundation for a healthy dairy herd. The management of calves has an impact on their growth, health and well-being. Scientific studies can reveal the influence of various factors and thus enable farmers and veterinarians to improve important areas regarding animal welfare and disease prevention. One aspect is the clean housing of calves so that they remain as clean as possible and have dry lying areas available. The hygiene status can be monitored by assessing the lying areas or directly on the animal. Such an assessment should always be statistically verified, so that it can be determined whether valid and reliable data can be obtained.

To allow an assessment of the dirtiness of calves, a hygiene score was created and validated. Three risk zones are considered and the scores of the three zones add up to the Calf Hygiene Score (CHS). The defined risk zones are located on the belly, the flank and the rear of the calf and are evaluated according to the percentage of contamination: Score 1 is awarded for no or little soiling (from 0 to 10 % of the area), Score 2 for medium soiling of over 10 % up to 30 %. Score 3 stands for heavy soiling of more than 30 % of the area. The CHS resulting from the assessments of the risk zones can range from 3 to 9. The validation of the CHS was carried out by five veterinarians, who scored 42 calves each. The inter- and intra-observer reliability was examined and a validation using a standard method was performed. For this purpose, the degree of actual soiling of the risk zones was determined by image processing of photos of the calves. The agreement between the observers and the standard method was weak to moderate with kappa values of 0.58 to 0.67. The main source of error is probably the optical measurement of surfaces, which is very difficult for the human perception. The values of the inter- and intra-observer reliability show a moderate to strong agreement with 29 of 36 Kappa values between 0.60 and 0.89. The hygiene score for calves has thus been successfully validated as reliable and can be used in the future as a useful tool for herd management.

With regards to the well-being of the calves and in accordance with current animal welfare legislation, the assessment of calves with the CHS is useful. However, the cleaner the calves, the higher the costs and the more time an animal owner must spend on keeping them clean. The results of future research into the link between calf health and dirtiness could provide consultants and veterinarians with further compelling arguments for the dialogue with the farmer.

VII. LITERATURVERZEICHNIS

Amrine DE, White BJ, Larson R, Anderson DE, Mosier DA, Cernicchiaro N. Precision and accuracy of clinical illness scores, compared with pulmonary consolidation scores, in Holstein calves with experimentally induced *Mycoplasma bovis* pneumonia. *American Journal of Veterinary Research* 2013; 74: 310-5.

Anonym (1997) Clean Cattle Policy. Ed Department of Agriculture FaF, Dublin

Barkema HW, Schukken YH, Lam TJ, Beiboer ML, Benedictus G, Brand A. Management practices associated with low, medium, and high somatic cell counts in bulk milk. *Journal of Dairy Science* 1998; 81: 1917-27.

Bokkers EAM, Vries M, Antonissen I, Boer IJM. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Animal Welfare* 2012; 21: 307-18.

Borderas TF, de Passille AM, Rushen J. Behavior of dairy calves after a low dose of bacterial endotoxin. *Journal of Animal Science* 2008; 86: 2920-7.

Buczinski S, Faure C, Jolivet S, Abdallah A. Evaluation of inter-observer agreement when using a clinical respiratory scoring system in pre-weaned dairy calves. *New Zealand Veterinary Journal* 2016; 64: 243-7.

Burfeind O, Keyserlingk M, Weary D, Veira D, Heuwieser W. Short communication: Repeatability of measures of rectal temperature in dairy cows. *Journal of Dairy Science* 2010; 93: 624-7.

Calvo-Lorenzo MS, Hulbert LE, Fowler AL, Louie A, Gershwin LJ, Pinkerton KE, Ballou MA, Klasing KC, Mitloehner FM. Wooden hutch space allowance influences male Holstein calf health, performance, daily lying time, and respiratory immunity. *Journal of Dairy Science* 2016; 99: 4678-92.

Camiloti TV, Fregonesi JA, von Keyserlingk MAG, Weary DM. Short communication: Effects of bedding quality on the lying behavior of dairy calves. *Journal of Dairy Science* 2012; 95: 3380-3.

Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 1960; 20: 37-46.

Cook NB. The influence of barn design on dairy cow hygiene, lameness, and udder health. *Proc. of the 35th Ann. Conv. Amer. Assoc. Bov. Pract.*, Madison, WI 2002: Pages 97-103.

Dahlhoff K (2014) Beratung von milchviehhaltenden Betrieben auf der Basis von Verhaltens- und Erscheinungsparametern ihrer Milchkühe In: *Landwirtschaftlichen Fakultät. Rheinische Friedrich-Wilhelms-Universität Bonn.*

Edmonson AJ, Lean IJ, Weaver LD, Farver T, Webster G. A body condition scoring chart for Holstein dairy cows. *Journal of Dairy Science* 1989; 72: 68-78.

Fecteau G, Pare J, Van Metre DC, Smith BP, Holmberg CA, Guterbock W, Jang S. Use of a clinical sepsis score for predicting bacteremia in neonatal dairy calves on a calf rearing farm. *The Canadian Veterinary Journal* 1997; 38: 101.

Graham AN, Renaud DL, Duffield TF, Kelton DF. Short communication: Calf cleanliness does not predict diarrhea upon arrival at a veal calf facility. *Journal of Dairy Science* 2018; 101: 3363-6.

Jordan D, McEwen SA, Wilson JB, McNab WB, Lammerding AM. Reliability of an ordinal rating system for assessing the amount of mud and feces (tag) on cattle hides at slaughter. *Journal of Food Protection* 1999; 62: 520-5.

Jorgensen MW, Adams-Progar A, de Passillé AM, Rushen J, Godden SM, Chester-Jones H, Endres MI. Factors associated with dairy calf health in automated feeding systems in the Upper Midwest United States. *Journal of Dairy Science* 2017; 100:

5675-86.

Klein-Jöbstl D, Iwersen M, Drillich M. Farm characteristics and calf management practices on dairy farms with and without diarrhea: A case-control study to investigate risk factors for calf diarrhea. *Journal of Dairy Science* 2014; 97: 5110-9.

Knierim U, Winckler C. On-farm welfare assessment in cattle: Validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare* 2009; 18: 451-8.

Lago A, McGuirk SM, Bennett TB, Cook NB, Nordlund KV. Calf respiratory disease and pen microenvironments in naturally ventilated calf barns in winter. *Journal of Dairy Science* 2006; 89: 4014-25.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.

Love WJ, Lehenbauer TW, Kass PH, Van Eenennaam AL, Aly SS. Development of a novel clinical scoring system for on-farm diagnosis of bovine respiratory disease in pre-weaned dairy calves. *PeerJ* 2014; 2: e238.

Lundborg GK, Svensson EC, Oltenacu PA. Herd-level risk factors for infectious diseases in Swedish dairy calves aged 0–90 days. *Preventive Veterinary Medicine* 2005; 68: 123-43.

Martin SW, H. MA, P. W. *Veterinary Epidemiology*. Iowa State Univ. Press, Ames 1987;

McEvoy JM, Doherty AM, Finnerty M, Sheridan JJ, McGuire L, Blair IS, McDowell DA, Harrington D. The relationship between hide cleanliness and bacterial numbers on beef carcasses at a commercial abattoir. *Letters in Applied Microbiology* 2000; 30: 390-5.

McGuirk SM. Disease management of dairy calves and heifers. *Veterinary Clinics of North America: Food Animal Practice* 2008; 24: 139-53.

McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica* 2012; 22: 276-82.

Meagher RK. Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science* 2009; 119: 1-14.

Neijenhuis F, Barkema HW, Hogeveen H, Noordhuizen JPTM. Classification and Longitudinal Examination of Callused Teat Ends in Dairy Cows. *Journal of Dairy Science* 2000; 83: 2795-804.

Panivivat R, Kegley EB, Pennington JA, Kellogg DW, Krumpelman SL. Growth performance and health of dairy calves bedded with different types of materials. *Journal of Dairy Science* 2004; 87: 3736-45.

Pardon B, Buczinski S, Deprez PR. Accuracy and inter-rater reliability of lung auscultation by bovine practitioners when compared with ultrasonographic findings. *Veterinary Record* 2019; 185: 109-.

Pelzer A, Cielejewski H, Bayer K, Büscher W, Kaufmann O (2007) Cows and more – Was die Kühe uns sagen. Bonitieren – Bewerten – Beraten mit System. 8. Tagung: Bau, Technik und Umwelt in der landwirtschaftlichen Nutztierhaltung. Bonn

Perino LJ, Apley MD. Clinical trial design in feedlots. *Veterinary Clinics of North America: Food Animal Practice* 1998; 14: 343-65.

Philipot JM, Pluvinage P, Cimarosti I, Sulpice P, Bugnard F. Risk factors of dairy cow lameness associated with housing conditions. *Veterinary Research* 1994; 25: 244-8.

Reneau JK, Seykora AJ, Heins BJ, Endres MI, Farnsworth RJ, Bey RF. Association

between hygiene scores and somatic cell scores in dairy cattle. *Journal of the American Veterinary Medical Association* 2005; 227: 1297-301.

Reubold H (2003) Signum-Test. / Bewertung der Tiergerechtheit von Liegeboxen.
. In: *Vet.-Med. Report* 14-6

Roth BA, Keil NM, Gygax L, Hillmann E. Influence of weaning method on health status and rumen development in dairy calves. *Journal of Dairy Science* 2009; 92: 645-56.

Ruddat I, Scholz B, Bergmann S, Buehring AL, Fischer S, Manton A, Prengel D, Rauch E, Steiner S, Wiedmann S, Kreienbrock L, Campe A. Statistical tools to improve assessing agreement between several observers. *Animal* 2014; 8: 643-9.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 2012; 9: 676-82.

Schreiner DA, Ruegg PL. Relationship between udder and leg hygiene scores and subclinical mastitis. *Journal of Dairy Science* 2003; 86: 3460-5.

Sprecher DJ, Hostetler DE, Kaneene JB. A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance. *Theriogenology* 1997; 47: 1179-87.

Stemler S. A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. *Practical Assessment, Research, and Evaluation* 2004; 9: 1-19.

Studds MJ, Deikun LL, Sorter DE, Pempek JA, Proudfoot KL. Short communication: The effect of diarrhea and navel inflammation on the lying behavior of veal calves. *Journal of Dairy Science* 2018; 101: 11251-5.

Sumner CL, von Keyserlingk MAG, Weary DM. How benchmarking motivates farmers to improve dairy calf management. *Journal of Dairy Science* 2018; 101: 3323-33.

Sutherland MA, Stewart M, Schütz KE. Effects of two substrate types on the behaviour, cleanliness and thermoregulation of dairy calves. *Applied Animal Behaviour Science* 2013; 147: 19-27.

Sutherland MA, Worth GM, Stewart M. The effect of rearing substrate and space allowance on the behavior and physiology of dairy calves. *Journal of Dairy Science* 2014; 97: 4455-63.

Sutherland MA, Worth GM, Cameron C, Ross CM, Rapp D. Health, physiology, and behavior of dairy calves reared on 4 different substrates. *Journal of Dairy Science* 2017; 100: 2148-56.

Terosky TL, Wilson LL, Stull CL, Stricklin WR. Effects of individual housing design and size on special-fed Holstein veal calf growth performance, hematology, and carcass characteristics. *Journal of Animal Science* 1997; 75: 1697-703.

van Donkersgoed J, Jericho KW, Grogan H, Thorlakson B. Preslaughter Hide Status of Cattle and the Microbiology of Carcasses. *Journal of Food Protection* 1997; 60: 1502-8.

Van Os JMC, Winckler C, Trieb J, Matarazzo SV, Lehenbauer TW, Champagne JD, Tucker CB. Reliability of sampling strategies for measuring dairy cattle welfare on commercial farms. *Journal of Dairy Science* 2018; 101: 1495-504.

Vogt A, Aditia EL, Schlechter I, Schütze S, Geburt K, Gauly M, König von Borstel U. Inter- and intra-observer reliability of different methods for recording temperament in beef and dairy calves. *Applied Animal Behaviour Science* 2017; 195: 15-23.

Webster AJ, Saville C, Church BM, Gnanasakthy A, Moss R. Some effects of different rearing systems on health, cleanliness and injury in calves. *British Veterinary Journal* 1985; 141: 472-83.

Welfare Quality® (2009) Welfare Quality® assessment protocol for cattle. Welfare Quality® Consortium, Lelystad, Netherlands.

Wells SJ, Trent AM, Marsh WE, Robinson RA. Prevalence and severity of lameness in lactating dairy cows in a sample of Minnesota and Wisconsin herds. *Journal of the American Veterinary Medical Association* 1993; 202: 78-82.

VIII. DANKSAGUNG

Ich möchte mich vor allem bei Moritz bedanken, der mich durch diese Dissertation geführt hat. Ohne dich wäre das nichts geworden. Dein offenes Ohr, deine direkten Rückmeldungen, deine viele Zeit, und auch der Druck, den du mir gegeben hast, waren nötig damit ich jetzt am Ende meiner Dissertation stehe. Aus vollstem Herzen: Danke! Was ich von dir gelernt habe, wird mich auch nach der Doktorarbeit in meinem Leben begleiten. Ich hoffe, dass du an mich als deine letzte Doktorandin auch mit Freude zurückdenken kannst und dass wir uns auch in Zukunft nicht aus den Augen verlieren.

Liebe Gabi, du hast es mir ermöglicht, Mitarbeiterin im PraeRi Projekt zu werden und gleichzeitig mit dir als meine Betreuerin eine Dissertation verfassen zu können. Das waren perfekte Rahmenbedingungen und bin sehr dankbar für diese Zeit. Ich hoffe, dass du zufrieden bist mit dem, was ich daraus gemacht habe!

Liebe Anna, deine Geduld ist einfach unglaublich. Danke für die Stunden, die du dir mit den Kappa-Berechnungen, den E-Mails und Terminen mit mir um die Ohren geschlagen hast. Du weißt, du bekommst bald wieder Post von mir :)

Freddy, Corinna, Philip, Alex, Anne: „Wenn aus Kollegen Freunde werden, sind es Frollegen“ :) Und das ist wahrlich so geworden! Ihr seid mir in der PraeRi (und auch außerhalb, bei unseren legendären Fortbildungen und anderen schönen Aktionen) sehr ans Herz gewachsen. Auch wenn unsere beruflichen Wege sich schon getrennt haben oder bald trennen werden, hoffe ich, dass wir uns als Freunde erhalten bleiben. Freddy, wir essen auch weiterhin am 1. September gemeinsam Kuchen! Corinna, danke für die ganzen Antworten auf all die tausend Fragen, mit denen ich dich zum Paper und zur Diss gelöchert habe. Du hast mir sehr viel Mut gemacht, als du vor mir alles durchlaufen hast und mir gezeigt hat wie gut es am Ende ausgeht, wenn man nur weiter macht.

Mein Dank gilt auch meinen Eltern. Ihr habt mir alles mitgegeben, was diese Doktorarbeit überhaupt erst ermöglicht hat. Abschließend möchte ich mich bei Johannes und Havanna bedanken. Ihr seid ausschlaggebend dafür gewesen, dass ich mich endlich daheim zum Schreiben hingesetzt habe. Danke für die Unterstützung in den Momenten, in denen ich sonst nicht weiter gemacht hätte.