

Kurzfristige Wiederholbarkeit über die Zeit von Einzelindikatoren und aggregierten Messgrößen des Wohlergehens in Milchviehbetrieben

Masterarbeit

an der Universität für Bodenkultur

Masterstudium: Nutztierwissenschaften

vorgelegt von
Gerhard Koch

betreut von
Univ.Prof. Dr.med.vet. Christoph Winckler
Mag. Dr. Marlene Kirchner

Wien, 05. November 2014

Danksagung

Es liegt nun an mir, mich bei all jenen zu bedanken, die mir das vorliegende Ergebnis ermöglichten. Ein Dank geht an Herrn Christoph Winckler und Frau Marlene Kirchner. Sie haben mir die Möglichkeit gegeben, diese Arbeit am Institut für Nutztierwissenschaften zu verfassen. Sie hatten stets Zeit, Geduld und immer passende Ratschläge für mich.

Ein weiterer Dank gilt meiner Studienkollegin Angelika Wassermann für die Erhebung der Daten in 15 Betrieben, für die Auswertung und auch für die Zusammenarbeit bei den Präsentationen.

Auch bei den 30 Betrieben, die Angelika und mir die Möglichkeit gegeben haben, in ihren Betrieben Daten zu erheben, möchte ich mich von ganzem Herzen bedanken.

Soweit personenbezogene Bezeichnungen nur in männlicher Form angeführt sind, beziehen sich diese auf Männer und Frauen in gleicher Weise.

Inhaltsverzeichnis

1	EINLEITUNG	5
2	FRAGESTELLUNG	6
3	LITERATURÜBERSICHT	7
3.1	Wohlergehen von Tieren (Animal Welfare).....	7
3.2	Beurteilung des Wohlergehens auf Betriebsebene	8
3.3	Reliabilität von Einzelindikatoren des Wohlergehens	13
3.4	Reliabilität von aggregierten Messgrößen des Wohlergehens	25
4	TIERE, MATERIAL UND METHODE	27
4.1	Betriebe	27
4.2	Haltungssysteme	28
4.3	Datenerhebungsmethode.....	28
4.4	Beobachtende Personen	28
4.5	Datenerhebung	29
4.6	Statistische Auswertung	34
5	ERGEBNISSE	37
5.1	Status des Wohlergehens in den Untersuchungsbetrieben	37
5.2	Kurzfristige Wiederholbarkeit über die Zeit von Einzelindikatoren und aggregierten Messgrößen des Wohlergehens	42
5.2.1	Alle Untersuchungsbetriebe	42
5.2.2	Untersuchungsbetriebe mit identischer bzw. abweichender Stichprobe bei der Einzeltieruntersuchung.....	47
6	DISKUSSION	49
6.1	Indikatoren des Wohlergehens.....	49
6.2	Aggregierte Messgrößen des Wohlergehens	56
7	FAZIT	58

8	ZUSAMMENFASSUNG.....	59
9	ABSTRACT	60
10	ABBILDUNGSVERZEICHNIS	61
11	TABELLENVERZEICHNIS.....	61
12	LITERATURVERZEICHNIS.....	63

1 Einleitung

Spezialisierung, Intensivierung und Mechanisierung sind für die Nutztiere häufig mit Änderungen in den Produktionsbedingungen verbunden. Quantität wurde zudem über die Qualität der einzelnen Produkte bzw. der Produktionsbedingungen gestellt. Diese Entwicklung führte zu einem steigenden Interesse der Bevölkerung am Wohlergehen von Nutztieren (Dörfler, 2007). Speziell Seuchen wie BSE, Maul- und Klauenseuche oder die Vogelgrippe waren einschlägige Ereignisse, die das Bewusstsein für das Wohlergehen der Tiere gesteigert haben. Konsumenten möchten über die Verhältnisse, in denen die Tiere gehalten werden, Bescheid wissen (Dörfler, 2007).

Vor diesem Hintergrund wurden weltweit verschiedene Beurteilungssysteme entwickelt, die das Wohlergehen von Nutztieren in den Betrieben beurteilen. Im Rahmen des EU-Projektes Welfare Quality[®] wurden Beurteilungssysteme für das Wohlergehen der Tiere entwickelt, die in landwirtschaftlichen Betrieben angewendet werden können. Dieses Konzept basiert vorwiegend auf der Erhebung tierbezogener Indikatoren (Welfare Quality[®], 2009). Tierbezogene Indikatoren sind direkte Parameter, die vom Verhalten und der Gesundheit des Tieres abgeleitet werden. Umweltbasierte Indikatoren sind hingegen indirekte Parameter, die sich auf die Haltungsbedingungen, das Management und dem Umgang mit den Tieren beziehen (Waiblinger et al., 2001).

Wichtige Eigenschaften von Indikatoren und aggregierten Messgrößen des Wohlergehens, wie sie beispielsweise im Welfare-Quality[®]-System Verwendung finden, sind die Aussagekraft, die einfache Durchführbarkeit und die Reliabilität (Zuverlässigkeit). Bei der Reliabilität werden verschiedene Aspekte in Betracht gezogen. Die Inter-Observer-Reliabilität bezieht sich auf die Übereinstimmung der Ergebnisse von zwei oder mehreren Beobachtern. Die Intra-Observer-Reliabilität betrachtet die Übereinstimmung innerhalb des gleichen Beobachters. Die Wiederholbarkeit über die Zeit bezieht sich auf die Übereinstimmung zwischen Beobachtungen von zumindest zwei verschiedenen Erhebungen desselben Beobachters (Scott et al., 2011). Diese Eigenschaft beruht darauf, dass eine Erhebung über einen längeren Zeitraum erhalten sein soll, außer es gibt Veränderungen in der Haltung oder im Management (Capdeville und Vessier, 2001; Winckler et al., 2007).

Bisher liegen nur wenige Untersuchungen zur Wiederholbarkeit über die Zeit für die Welfare Quality[®] Erhebungsprotokolle für Rinder vor. Neben der Untersuchung von Einzelindikatoren haben Kirchner et al. (2014) die Wiederholbarkeit des Welfare Quality[®] Erhebungsprotokolls bei Masttieren erhoben. 63 Betriebe wurden drei Mal in einem Intervall von einem Monat und sieben Monaten nach der ersten Erfassung der Daten besucht. Die vorliegende Untersuchung befasst sich nun mit der kurzfristigen Wiederholbarkeit über die Zeit des Welfare Quality[®] Erhebungsprotokolls für Milchkühe, d.h. im Abstand von wenigen Tagen. Zusätzlich

wurden die qualitative Beurteilung des Aufstehvorganges und verschiedene Indizes der Liegeplatznutzung einbezogen.

2 Fragestellung

Über die Wiederholbarkeit über die Zeit von Einzelindikatoren und aggregierten Messgrößen gibt es bisher nur wenige Erhebungen. Die vorliegende Untersuchung soll einen Beitrag zur Reliabilitätsprüfung der Erhebung des Wohlbefindens von Milchkühen leisten.

Dazu wurde folgender Fragestellung nachgegangen:

- Wie wiederholbar sind Einzelindikatoren und aggregierte Messgrößen des Wohlergehens bei Erhebungen im Abstand von vier Tagen?
- Welche Ursachen lassen sich für die Variabilität feststellen?

Folgende Einzelindikatoren und aggregierten Messgrößen wurden dazu erhoben und die kurzfristige Wiederholbarkeit über die Zeit mit der Korrelation nach Spearman berechnet:

- Einzelindikatoren sowie aggregierte Scores für Welfare-Kriterien und Welfare-Prinzipien des Welfare Quality[®] Erhebungsprotokolls für Milchkühe
- Qualitative Beurteilung des Aufstehvorganges
- Indizes der Liegeplatznutzung

3 Literaturübersicht

3.1 Wohlergehen von Tieren (Animal Welfare)

In den 1960er Jahren wurde das Interesse am Wohlergehen der Nutztiere in Europa geweckt (Dörfler, 2007). Ausschlaggebend dafür waren große Veränderungen in der landwirtschaftlichen Nutztierproduktion (Blokhuys et al., 1998). Landwirtschaftliche Betriebe hatten ihre Produktion spezialisiert, die Produktion wurde intensiviert und auch die Anzahl der Nutztiere pro Betrieb stieg an. Im Bereich Haltung und Management kam es zu großen Veränderungen durch Mechanisierung und technische Entwicklungen. Gegenüber der Quantität wurde die Qualität der Produkte mehr und mehr in den Hintergrund gerückt, wobei man sich in erster Linie auf das Angebot, den Preis und die Konkurrenz konzentrierte (Dörfler, 2007). In weiterer Folge waren auch Seuchen wie BSE, Maul- und Klauenseuche oder die Vogelgrippe einschlägige Ereignisse, die das Bewusstsein hinsichtlich Animal Welfare gesteigert haben (Korte et al., 2006). Das steigende Interesse der Öffentlichkeit hat zu großen Änderungen geführt, sodass die Lebensmittelsicherheit und der Umweltschutz eine tragende Rolle bei der Diskussion um die Nutztierproduktion spielen (Blokhuys, 2005). Tierwohl hat eine große Priorität für die Konsumenten in Europa. Es ist nicht mehr nur das Endprodukt wichtig, sondern auch die Verhältnisse, in denen das Tier gehalten wird (Miele und Roex, 2005).

Das Brambell-Komitee war eine der ersten Kommissionen, die sich mit der Beurteilung des Wohlergehens von Tieren in der intensiven Tierhaltung auseinandergesetzt haben. Diese wurde eigens von der britischen Regierung eingesetzt (Webster, 1994; Webster, 2005). Im Brambell-Bericht von Dezember 1965 wurde erstmals das Konzept der Freiheiten genannt. Nutztiere sollen zumindest die Freiheit haben, „aufzustehen, sich hinzulegen, sich umzudrehen, sich zu putzen und ihre Glieder zu strecken“ (Anonym, 2009). Die vier Freiheiten waren somit der Anstoß für die Animal-Welfare-Debatte in Europa (Dörfler, 2007). Angelehnt an die vier Freiheiten wurde vom Farm Animal Welfare Council im Jahr 1993 eine modifizierte Form entwickelt.

Die fünf Freiheiten beinhalten:

1. Freiheit von Durst, Hunger und Unterernährung
2. Freiheit von falscher Haltung
3. Freiheit von Schmerz, Verletzungen und Krankheiten
4. Freiheit zum Ausleben normaler Verhaltensmuster
5. Freiheit von Angst und Stress (Webster, 2001)

3.2 Beurteilung des Wohlergehens auf Betriebsebene

Das grundsätzliche Ziel der Beurteilung des Wohlergehens auf Betriebsebene (On-Farm Welfare Assessment) ist es, das Wohlergehen der Tiere zu messen (Fraser und Sorensen, 2010). Für die Beurteilung des Wohlergehens auf Betriebsebene ist eine angemessene Auswahl an Messmethoden notwendig mit denen in weiterer Folge eine Gesamtbewertung vorgenommen werden kann (Knierim und Winckler, 2009). Jedoch ist dies sehr komplex, da verschiedene Einflussbereiche zu beurteilen sind. Die Einteilung der Parameter des Wohlergehens erfolgt in der Regel in direkte und indirekte Parameter. Unter direkten Parametern versteht man tierbezogene Parameter, die von der Gesundheit und dem Verhalten des Tieres abgeleitet werden können. Indirekte Parameter sind Faktoren der Umwelt, die das Tier beeinflussen. Dazu zählen die Haltungsbedingungen, das Management und der Umgang mit Tieren (Waiblinger et al., 2001). Die Ermittlung von umweltbasierten Parametern ist verhältnismäßig einfach und schnell durchzuführen. Dahingehend sind diese Parameter eine gute Ausgangsbasis, um Probleme des Wohlergehens zu lösen (Johnsen et al., 2001). Tierbezogene Parameter geben einen Aufschluss, wie erfolgreich sich das Tier mit seiner Umwelt auseinandersetzen kann. Über den Gesundheitszustand können auch Rückschlüsse auf zurückliegende Ereignisse gezogen werden (Hörning, 2001). Die Ermittlung von tierbezogenen Parametern ist aufwändiger als jene der umweltbasierten Parameter (Johnsen et al., 2001).

Ziele der Beurteilung des Wohlergehens auf Betriebsebene

Die Ermittlung des Wohlergehens von Tieren verfolgt eine Reihe an Zielen:

- Ermittlung der vorherrschenden Probleme des Wohlergehens
- Überprüfung von derzeitigen gesetzlichen Regelungen auf den landwirtschaftlichen Betrieben
- Aufzeigen der Risikofaktoren für die Beeinträchtigung des Wohlergehens
- Identifizierung und Priorisierung der Ziele für Interventionsprogramme
- Aufklärung verschiedener Stakeholder über die Notwendigkeit von Interventionsprogrammen
- Überprüfung der Wirksamkeit von Interventionsprogrammen
- Sammeln von Informationen, um in weiterer Folge Beratungen durchzuführen und ethische Urteile abzugeben (Whay, 2007).
- Entwicklung von Gütesiegeln für die Erzielung höherer Produktpreise (Johnsen et al., 2001)
- Verleihung von Zertifikaten (Knierim und Winckler, 2009)

Anforderungen an Beurteilungskriterien des Wohlergehens auf Betriebsebene

Damit Methoden für die Anwendung im On-Farm Welfare Assessment geeignet sind, müssen diese gewisse Anforderungen erfüllen. Die Methoden müssen aussagekräftig, einfach durchführbar und reliabel sein (Hörning, 2001).

a) aussagekräftig:

Es muss eine sichere Aussage über das Wohlergehen der Tiere möglich sein. Tierbezogene Parameter haben einen theoretischen Vorteil gegenüber umweltbasierten, da sie direkt am Tier gemessen werden (Knierim und Winckler, 2009). Es wird unterschieden zwischen *face validity*, *construct validity*, *content validity* und *criterion validity*. *Face validity* ist die offensichtliche Aussagekraft nach Einschätzung durch Experten. *Construct validity* gibt an, inwieweit aufgestellte Hypothesen mit einem Test bestätigt werden können. *Content validity* liegt vor, wenn die verschiedenen Abstufungen einer Methode alle Ausprägungen des Wohlergehens in diesem Bereich abdecken. *Criterion validity* zeigt, welchen Bezug die Methode zu anderen Welfare-Parametern hat (Scott et al., 2001; Knierim und Winckler, 2009).

b) einfach durchführbar:

Die Beurteilung des Wohlergehens auf Betriebsebene muss für einen ausgebildeten Beobachter leicht anwendbar sein und die benötigte Zeit für die Erhebung der Daten soll nicht zu lang sein (Hörning, 2001; Winckler, 2008).

c) reliabel

Es wird unterschieden zwischen der Übereinstimmung der Ergebnisse verschiedener Beobachter (Inter-Observer-Reliabilität), der Übereinstimmung bei aufeinanderfolgenden Messungen des gleichen Beobachters (Intra-Observer-Reliabilität) und der Übereinstimmung zwischen Beobachtungen von zumindest zwei verschiedenen Erhebungen desselben Beobachters (Wiederholbarkeit über die Zeit) (Scott et al., 2001). In der Literatur wird bei der Wiederholbarkeit über die Zeit („repeatability“) auch von der „test-retest reliability“ (Scott et al., 2001) oder „consistency over time“ (Plesch et. al., 2010) gesprochen. Die Wiederholbarkeit über die Zeit von Indikatoren des Wohlergehens hat eine hohe Priorität für die Beurteilung des Wohlergehens auf Betriebsebene, speziell bei der Durchführung von Zertifizierungen. Um Kosten zu sparen, sollten die Beurteilungen zumindest für sechs Monate gültig sein (Knierim und Winckler, 2009).

Welfare Quality® Projekt

Welfare Quality® ist das Akronym für das europäische Forschungsprojekt „Integration des Wohlergehens von Tieren in die Lebensmittelerzeugungskette: Vom öffentlichen Interesse zu verbessertem Wohlergehen der Tiere und transparenter Qualität“. Das Welfare Quality® Projekt war darauf ausgerichtet, das Wohlergehen der Tiere in die Produktionskette zu integrieren (Canali und Keeling, 2009) und wurde von 2004 bis 2009 durchgeführt. Insgesamt waren 44 Institutionen aus 13 verschiedenen europäischen Ländern und vier Länder aus Lateinamerika beteiligt (Uruguay, Brasilien, Chile, Mexiko) (Blokhus, 2008).

a) Ziele des Welfare Quality® Projektes

- Entwicklung von praktischen Strategien/Indikatoren für die Verbesserung des Wohlergehens von Tieren
- Entwicklung eines Protokolls für die Beurteilung des Wohlergehens von Tieren am Betrieb und bei der Schlachtung
- Entwicklung eines Protokolls, um Informationen aus der Beurteilung des Wohlergehens der Tiere für die Kennzeichnung von Produkten zu nutzen
- Kooperation von europäischen Experten, um ein multidimensionales Konzept für das Wohlergehen der Tiere zu entwickeln (Blokhus, 2008)

b) Bereiche des Welfare Quality® Projektes

Der erste Weg strebte die Verbesserung des Wohlergehens der Tiere durch innovative, wissensbasierte, praktische und spezifische Strategien an, die direkt auf den landwirtschaftlichen Betrieben oder an anderen Stellen der Primärproduktion wie beispielsweise an Zuchtstationen angewendet werden können.

Zweitens sollte das Wohlergehen der Tiere transparenter gemacht werden, um in weiterer Folge den Konsumenten mehr Informationen über das Produkt, das sie kaufen, zu geben. Dazu ist eine Beurteilung des Wohlergehens der Nutztiere auf Betriebsebene notwendig (Blokhus, 2005).

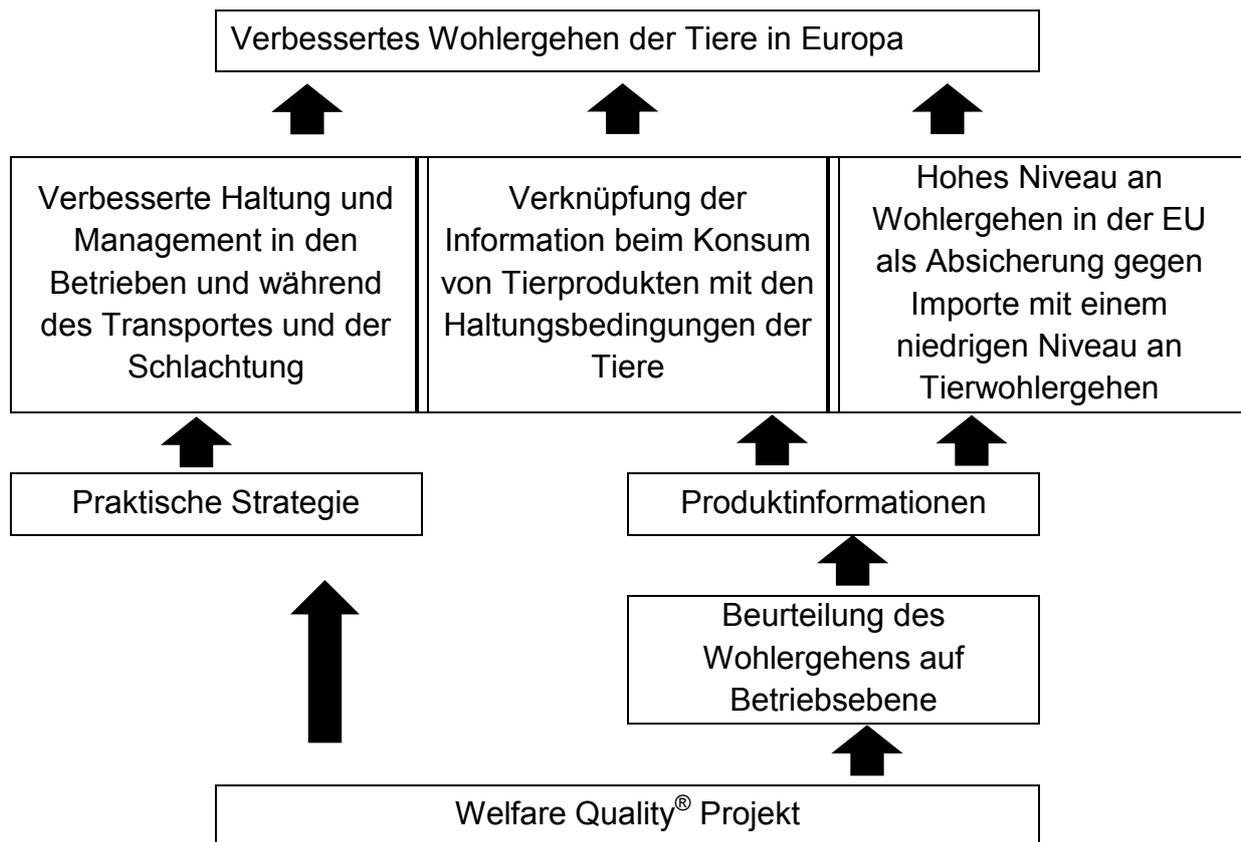


Abbildung 1: Wege zur Verbesserung des Wohlergehens der Tiere des Welfare Quality® Projektes (Blokhus, 2005)

c) Konzept Welfare Quality® Erhebungsprotokoll für Milchkühe

Wohlergehen ist multidimensional und kann nicht direkt gemessen werden. Daher ist es notwendig, eine Auswahl an Indikatoren in verschiedenen Disziplinen zu entwickeln (Blokhus, 2008). Die früheren Konzepte zur Beurteilung des Wohlergehens der Tiere beruhten auf umweltbasierten Parametern und auf der Erhebung von Ressourcen (Canali und Keeling, 2009). Im Zuge des Welfare Quality® Projektes wurde für die Beurteilung des Wohlergehens auf Betriebsebene ein eigenes Protokoll für Milchkühe entwickelt. Im Welfare Quality® Erhebungsprotokoll für Milchkühe sind verstärkt tierbezogene Parameter enthalten (Welfare Quality®, 2009). Ressourcen- oder managementbasierte Indikatoren ergänzen notwendige Informationen, wenn ein spezieller Aspekt des Wohlergehens nicht direkt am Tier bewertet werden kann. Um den aktuellen Status des Wohlergehens von Tieren in Bezug auf Verhalten, Gesundheit und Physiologie zu bewerten, sind tierbezogene Parameter wichtig (Canali und Keeling, 2009).

Das Welfare Quality® Erhebungsprotokoll für Milchkühe besteht aus einer Kombination von tierbezogenen, ressourcen- und managementbasierten Indikatoren (Canali und Keeling, 2009). Aus den verschiedenen Indikatoren werden in Summe zwölf Welfare-Kriterien gebildet. Diese zwölf Welfare-Kriterien bilden wiederum im nächsten Schritt die Grundlage für vier Welfare-Prinzipien, nämlich Gute Fütterung,

Gute Haltung, Gute Gesundheit und Artgemäßes Verhalten. Schlussendlich wird aus den vier Welfare-Prinzipien eine Gesamtbeurteilung für den Betrieb vorgenommen. Die Gesamtbeurteilung zeigt schließlich eine Bewertung in vier Kategorien: ausgezeichnet („excellent“), gehobener Standard („enhanced“), akzeptabel („acceptable“) und nicht klassifiziert („not classified“). Eine Aufstellung der verschiedenen Indikatoren, der zwölf Welfare-Kriterien und der vier Welfare-Prinzipien ist in Tabelle 1 ersichtlich (Welfare Quality®, 2009).

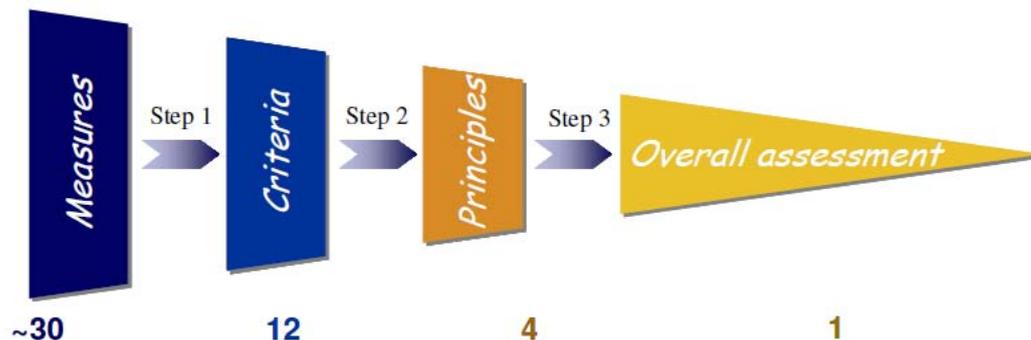


Abbildung 2: Schritte von links nach rechts von den Indikatoren („Measures“), Welfare-Kriterien („Criteria“), Welfare-Prinzipien („Principles“) zur Gesamtbeurteilung („Overall Assessment“) (Welfare Quality®, 2009)

Tabelle 1: Indikatoren, Welfare-Kriterien und Welfare-Prinzipien des Welfare Quality® Erhebungsprotokolls für Milchkühe (Welfare Quality®, 2009)

Welfare-Prinzipien	Welfare- Kriterien	Indikatoren
Gute Fütterung	1. Abwesenheit von andauerndem Hunger	Body Condition Score
	2. Abwesenheit von andauerndem Durst	Wasserversorgung, Sauberkeit der Wasserstellen, Wasserdurchfluss, Funktionsfähigkeit der Wasserstellen
Gute Haltung	3. Liegekomfort	Abliegezeit, Kollisionsfrequenz während des Abliegevorganges, teilweise oder komplett außerhalb der Liegefläche liegende Tiere, Sauberkeit von Unterbein, Hinterhand und Euter

	4.	Thermischer Komfort	Keine Indikatoren
	5.	Bewegungsfreiheit	Anbindehaltung, Zugang zu Auslaufflächen oder Weide
Gute Gesundheit	6.	Abwesenheit von Verletzungen	Lahmheit, Integumentschäden
	7.	Abwesenheit von Krankheiten	Husten, Nasenausfluss, Augenausfluss, erschwerte Atmung, Diarrhö, Vulvausfluss, subklinische Mastitis, Mortalität, Schwergewburten, festliegende Kühe
	8.	Abwesenheit von Schmerzen	Enthornung, Kupieren der Schwänze
	9.	Ausdruck von Sozialverhalten	Agonistisches Verhalten
Artgemäßes Verhalten	10.	Ausdruck anderer Verhaltensweisen	Weidegang
	11.	Mensch-Tier-Beziehung	Ausweichdistanz
	12.	Positiver emotionaler Zustand	Qualitative Verhaltens erfassung

3.3 Reliabilität von Einzelindikatoren des Wohlergehens

Im Weiteren wird ein Überblick der Literatur zu den verschiedenen Indikatoren hinsichtlich Reliabilität gegeben. Bei der Reliabilität wurden verschiedene Aspekte berücksichtigt, sofern dazu Veröffentlichungen vorlagen: Inter-Observer-Reliabilität, Intra-Observer-Reliabilität und Wiederholbarkeit über die Zeit.

a) Indikator Body Condition Score (BCS)

Der BCS reflektiert den Körperfettgehalt (Wagner et al., 1988) und damit den Ernährungszustand eines Tieres. BCS zeigt nicht nur den aktuellen Ernährungszustand des Tieres, sondern gibt auch Auskunft über die vergangene Zeit. BCS ist ein wichtiger Maßstab für die Gesundheit des Tieres. Beispielsweise haben zu fette Kühe in der Trockenzeit ein erhöhtes Risiko für Schwergewburten, Fettleber oder eine verzögerte Rückkehr der Brunst (Reid et al., 1986). Auch zu Beginn der Laktation können Kühe aufgrund der hohen Leistung und einer negativen Energiebalance Körperkondition verlieren und dies kann sich nachteilig auf die Gesundheit und die Fruchtbarkeit auswirken (Butler, 2003). Eine niedrige

Körperkondition ist ein Indikator, dass der Energiebedarf des Tieres nicht mit jener des aufgenommenen Futters übereinstimmt. BCS wird im Allgemeinen als wichtiges Werkzeug im Management von Mast- und Milchrinderherden verwendet (Leach et al., 2009a).

Inter-Observer-Reliabilität:

Bei Leach et al. (2009) wurde die Inter-Observer-Reliabilität bei Mastrindern und Milchkühen festgestellt. Die Beurteilung erfolgte durch drei Personen mittels Fotos. Zwei Personen waren Experten im Bereich Body Condition Scoring; die dritte Person hatte keine Erfahrung mit der Beurteilung von BCS. Die Bewertung erfolgte nach einer Drei-Stufen-Skala mit zu mager, normal und zu dick. Vorab wurde allen Personen Trainingsmaterial zur Verfügung gestellt. Die Übereinstimmung der zwei Experten ergab 79 Prozent für Milchkühe und für Mastrinder 90 Prozent. Die Übereinstimmung von allen drei Beobachter war mit 67 Prozent bei den Mastrindern und 65 Prozent bei den Milchkühen deutlich geringer (Leach et al., 2009a).

Wiederholbarkeit über die Zeit:

Kirchner et al. (2014) haben die Wiederholbarkeit über die Zeit für die Körperkondition über kürzere und längere Zeit bei Maststieren untersucht. Die Erhebung erfolgte in Summe drei Mal. Ein Monat und sieben Monate nach der ersten Erhebung erfolgten die zweite und dritte Erhebung. Die Rangkorrelation zwischen der ersten und zweiten Erhebung betrug 0,29 (n=59 Betriebe). Zwischen der zweiten und dritten Erhebung mit einem Intervall von sechs Monaten ergab sich eine Korrelation von 0,68 (n=49; Kirchner et al., 2014).

b) Indikatoren zur Ressource Wasser

Wasser ist das wichtigste Futtermittel für Rinder (Dyer, 2012). Das Gesamtgewicht der Milchkühe besteht zu 56 bis 81 Prozent aus Wasser. Die Hauptkomponente von Milch und der Körperausscheidungen ist Wasser (Murphy, 1992). Daraus resultiert, dass Milchkühe einen erheblichen täglichen Bedarf an Wasser haben (Linn und Raeth-Knight, o.J.). Ein freier Zugang zu Wasser mit hoher Qualität hat daher große Bedeutung (Dyer, 2012). Bei schlechter Wasserqualität sinkt die Aufnahme von Wasser. Dies führt zu einer reduzierten Milchleistung und einem schlechteren Gesundheitsstatus (Linn und Raeth-Knight, o.J.). Der Bedarf an Wasser pro Tier ist abhängig von der Größe des Tieres, vom Laktationsstadium, Körperkondition und der Umgebungstemperatur (Dyer, 2012).

Wiederholbarkeit über die Zeit:

Wie in der bereits näher beschriebenen Studie von Kirchner et al. (2014) wurde die Wiederholbarkeit über die Zeit der wasserbezogenen Ressourcen bei Maststieren erhoben. Für die Indikatoren ‚% Gruppen mit ausreichenden Wassertränken‘ und ‚% Gruppen mit zumindest zwei Wassertränken‘ lag eine sehr hohe Korrelation für die Wiederholbarkeit über kurze und längere Zeit vor (siehe Tabelle 2). Für den Indikator ‚Prozentueller Anteil an Gruppen mit verschmutzten Tränken‘ konnte keine Korrelation berechnet werden, da das Auftreten zu selten war (Kirchner et al., 2014).

Tabelle 2: Wiederholbarkeit über die Zeit von zwei Indikatoren zur Ressource Wasser bei Maststieren (Kirchner et al., 2014)

Indikator	r_s^a	r_s^b
% Gruppen mit ausreichenden Wassertränken	0,95	0,91
% Gruppen mit zumindest zwei Wassertränken	0,94	0,94

^a Korrelation nach Spearman mit einem Intervall von einem Monat

^b Korrelation nach Spearman mit einem Intervall von sechs Monaten

c) Indikatoren zum Ruheverhalten

Liegen ist ein Verhalten mit höchster Priorität (Munksgaard et al., 2005). Rinder ruhen hauptsächlich während des Liegens. Beeinträchtigungen im Ruheverhalten können zu mangelhafter Erholung und Frustration (Munksgaard und Simonsen, 1996), steigendem Lahmheitsrisiko (Singh et al., 1994) sowie Integumentschäden (Wechsler et al., 2000) führen.

Inter-Observer-Reliabilität:

Brörkens et al. (2009) haben in den Jahren 2005 und 2006 viermal Datenerhebungen zum Ruheverhalten durchgeführt. Die untersuchten Milch- und Stiermastbetriebe befanden sich in Deutschland, Österreich und Italien. Anhand von Videoaufzeichnungen erzielten zunächst zwei Beobachter bei 34 Abliegevorgängen von Milchkühen eine Korrelation von 0,85. Bei Live-Beobachtungen von 13 Abliegevorgängen durch zwei Beobachter in zwei verschiedenen Milchviehbetrieben wurde eine Inter-Observer-Reliabilität von 0,98 erreicht. Dies wurde durch weitere Live-Beobachtungen anhand von 21 Abliegevorgängen in zwei Milchviehbetrieben bestätigt. Für Beobachter A und B betrug die Korrelation 0,98. Videobeobachtungen von weiteren 65 Abliegevorgängen durch zwei Beobachter erbrachten eine Inter-Observer-Reliabilität von 0,95 (Brörkens et al., 2009).

Brörkens et al. (2009) haben außerdem die Inter-Observer-Reliabilität bezüglich der Beurteilung, ob Tiere teilweise oder komplett außerhalb der Liegefläche liegen,

getestet. Als Grundlage der Bewertung dienten 57 Bilder. Die Korrelation war mit 0,75 als akzeptabel zu bewerten (Brörkens et al., 2009).

Wiederholbarkeit über die Zeit:

Brörkens et al. (2009) und Plesch et al. (2010) haben die Wiederholbarkeit über die Zeit von verschiedenen Indikatoren zum Ruheverhalten in Milch- und Stiermastbetrieben untersucht. Die Erhebung erfolgte bezüglich der mittleren Abliegezeit, des Anteils an Tieren, die mit der Haltungseinrichtung beim Abliegen kollidieren, sowie des Anteils an Tieren, die teilweise oder komplett außerhalb der Liegefläche liegen. Die Betriebsbesuche der 19 Stiermastbetriebe und 35 Milchviehbetriebe erfolgten von August 2005 bis April 2006 in Österreich und Deutschland. Die Datenerhebung pro Betrieb erfolgte drei Mal mit einem Zeitintervall von 60 und 180 Tagen (± 10 Tage) nach der ersten Erhebung. Für die statistische Berechnung wurde Kendalls Rangkorrelationskoeffizient (W) gewählt. Für die Abliegezeit wurde eine Wiederholbarkeit über die Zeit bei den Milchkühen von 0,78 und bei den Mastrindern von 0,82 festgestellt. Für den Indikator ‚% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren‘ ließ sich eine Wiederholbarkeit über die Zeit von 0,95 feststellen. Für die Mastrinder konnte keine Berechnung erfolgen. Bei ‚% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen‘ betrug die Korrelation für Milchkühe 0,87 (Brörkens et al., 2009; Plesch et al., 2010).

Kirchner et al. (2014) ermittelten die Wiederholbarkeit über die Zeit für die ‚durchschnittliche Abliegezeit‘ bei Maststieren. Bei Betriebsbesuchen im Abstand von einem Monat ergab sich eine Korrelation von 0,70 ($n=55$ Betriebe). Mit einem Intervall von 6 Monaten ging die Konsistenz weiter zurück ($r_s=0,45$; $n=42$).

d) Indikatoren zur Sauberkeit

Hinweise aus der Verhaltensforschung deuten darauf hin, dass Rinder eine schmutzige Umwelt meiden (Phillips und Morris, 2002). Die Hygiene der Haltungsumwelt steht in einem engen Zusammenhang mit Infektionskrankheiten wie Mastitis (Schukken et al., 1990) und Klauenentzündungen (Roderiguez-Lainz et al., 1996). Die Sauberkeit der Zitzen und der unteren Beine steht in einem Zusammenhang mit subklinischer Mastitis. Durch die verschmutzte Haltungsumwelt kommen die Zitzen und die unteren Beine in vermehrten Kontakt mit Mikroorganismen, die das Auftreten einer subklinischen Mastitis begünstigen (Schreiner und Ruegg, 2003).

Wiederholbarkeit über die Zeit:

Winckler et al. (2007) untersuchten zwischen Jänner und Oktober 2003 in Deutschland in acht Holstein-Friesian-Milchviehbetrieben (29-102 Milchkühe) die Wiederholbarkeit über die Zeit der Sauberkeit von verschiedenen Körperregionen.

Für die Ermittlung der Sauberkeit wurden pro Betrieb zumindest 30 Prozent der Kühe oder mindestens 20 Kühe ausgewählt. Die Sauberkeit der Hinterbeine und des Euters wurden mit einer Fünf-Punkte-Skala fünf Mal in einem Abstand von zwei Monaten erhoben. Es wurde jeweils die Korrelation (r_s) zwischen zwei aufeinanderfolgenden Erhebungen berechnet. Für die Sauberkeit der Hinterbeine wurde eine Korrelation von 0,60 bis 0,83 ermittelt. Die Sauberkeit des Euters war weniger konstant mit einer Korrelation zwischen 0,10 und 0,61 (Winckler et al., 2007).

Auch De Rosa et al. (2003) ermittelten die Wiederholbarkeit über die Zeit der Verschmutzungsbeurteilung in vier Milchviehbetrieben mit Laufstallhaltung in Italien. Die Betriebe wurden drei Mal in einem dreiwöchigem Intervall besucht. Pro Betrieb wurden 30 bis 40 Tiere ausgewählt. Bewertet wurden fünf Körperregionen (Ano-Genitalbereich, Euter, Beine, obere Beine und hinterer Unterbauch) mit einer Skalierung von 0 bis 2 mit einer 0,5-Punkte-Abstufung. Der Kendalls Rangkorrelationskoeffizient als Maß für die Übereinstimmung der Sauberkeit in den Betrieben bewegte sich zwischen 0,51 und 0,71 (De Rosa et al., 2003).

e) Indikator Lahmheit

Lahmheit ist ein verbreitetes Problem bezüglich des Wohlergehens der Tiere. Sie verursacht Schmerzen (Whay et al., 1997) und begrenzt das normale Verhalten der Rinder (Singh et al., 1993; Hassall et al., 1993). Lahmheiten führen in der Milchviehhaltung zu erheblichen wirtschaftlichen Verlusten (Alban et al., 1996). Ein jährliches Auftreten von bis zu 54 Prozent wurde berichtet (Clarkson et al., 1996). Lahme Kühe leiden durch den Schmerz für eine lange Zeit. Es bestehen auch Hinweise, dass Lahmheiten die Schmerzempfindlichkeit der Kühe längerfristig erhöhen (Whay et al., 1997).

Inter-Observer-Reliabilität:

Leach et al. (2009) führten eine Lahmheitsbeurteilung bei 98 Milchkühen in vier Betrieben mit Anbindehaltung mit zwei Beobachtern durch. Die Beurteilung erfolgte mit einer binären Klassifikation (lahm und nicht lahm). Die Übereinstimmung zwischen Beobachter A und Beobachter B betrug 91 Prozent (Leach et al., 2009b).

147 Kühe wurden bei Winckler und Willen (2001) gleichzeitig von drei Beobachtern beurteilt. Die Milchkühe wurden anhand einer Fünf-Punkte-Skala erfasst. Die drei Beobachter hatten eine Übereinstimmung von 68 Prozent. Bei 30 Prozent der Kühe betrug die Abweichung nur ein Punkt. Eine Differenz von zwei Punkten wurde nur in zwei Prozent der erfassten Daten festgestellt (Winckler und Willen, 2001).

Intra-Observer-Reliabilität:

Leach et al. (2009) erhoben die Intra-Observer-Reliabilität der Lahmheitsbeurteilung bei 46 Kühen in zwei Betrieben. Die Daten wurden in einem zeitlichen Abstand von zwölf Stunden durch einen Beobachter aufgenommen. Es wurde eine Übereinstimmung von 70 Prozent festgestellt (Leach et al., 2009c).

Wiederholbarkeit über die Zeit:

De Rosa et al. (2003) untersuchten die Wiederholbarkeit über die Zeit des Indikators Lahmheit (Verfahrensablauf siehe Indikatoren zur Sauberkeit). Die Gangbeurteilung wurde beim Verlassen des Melkstandes mit einem Score von 0 bis 3 durchgeführt. Es zeigte sich eine Übereinstimmung nach Kendall von 0,43 bis 0,66 (De Rosa et al., 2003).

Winckler et al. (2007) ermittelten die Lahmheitsprävalenz in acht Holstein-Friesian-Betrieben in Deutschland mittels Fünf-Punkte-Skala nach Winckler und Willen (2001) (Verfahrensablauf siehe Indikator Lahmheit). Die Korrelation (r_s) zwischen zwei aufeinanderfolgenden Erhebungen schwankte zwischen 0,48 und 0,78 (Winckler et al., 2007).

f) Indikator Integumentschäden

Veränderungen der Haut haben verschiedenste Ursachen (Schulze Westerath et al., 2009). Dazu zählen Ektoparasiten (Rosenberger, 1970), agonistisches Verhalten (Menke et al., 1999), Verletzungen der Zitzen (Groth, 1984; Groth, 1985) und des Schwanzes (Schrader et al., 2001) durch Klauentritte sowie Verletzungen durch die Stallausrüstung (Groth, 1985).

Inter-Observer-Reliabilität:

Danuser und Regula (2001) erhoben im Jahr 1999 die Inter-Observer-Reliabilität von bestimmten Integumentschäden bei Milchkühen in der Schweiz. Die Übereinstimmung von zwei Beobachtern betrug bei Schwellungen von Tarsus oder Carpus 75 Prozent, bei Verletzungen an den Hinterbeinen 78 Prozent und bei Veränderungen am Tarsus 79 Prozent (Danuser und Regula, 2001).

Wiederholbarkeit über die Zeit:

Winckler et al. (2007) besuchten Holstein-Friesian-Betriebe für die Ermittlung der Wiederholbarkeit über die Zeit von Abschürfungen und Wunden an den Tarsal- und Karpalgelenken (Verfahrensablauf siehe Indikator Lahmheit). Für die Tarsalgelenke konnte für zwei aufeinanderfolgende Erhebungen eine Korrelation (r_s) zwischen 0,49 und 0,78 festgestellt werden. Für die Karpalgelenke lag die Übereinstimmung zwischen 0,05 und 0,37 (Winckler et al., 2007).

Kirchner et al. (2014) ermittelten Integumentschäden bei Maststieren. Für eine statistische Berechnung war jedoch das Auftreten von Integumentschäden zu selten.

g) Indikatoren zum Gesundheitsstatus

Krankheiten und Mortalität von Milchkühen sind ein großes Problem hinsichtlich des Wohlergehens der Tiere und wirtschaftlicher Verluste (Canali et al., 2009). Eine effektive Gesundheitsvorsorge bedarf deshalb einer geeigneten Umwelt und Haltung der Tiere. Präventive Maßnahmen wie beispielsweise eine gute Hygiene oder geeignete Impfungen können dabei helfen, Infektionen der Herde vorzubeugen. Viele Krankheiten sind multifaktoriell bedingt. Deshalb ist für eine effektive Gesundheitsvorsorge wichtig, dass die Tiere in einer passenden Umwelt gehalten werden um Stress und ein geschwächtes Immunsystem zu verhindern (SCAHAW, 2001).

Wiederholbarkeit über die Zeit:

Kirchner et al. (2014) erhoben die Wiederholbarkeit über die Zeit von gewissen Indikatoren bei Maststieren (Verfahrensablauf siehe Indikator Body Condition Score). Wie in Tabelle 3 ersichtlich, zeigte sich nur für den Indikator ‚% Tiere mit Augenausfluss‘ eine hohe Korrelation (r_s) von über 0,7 für die Betriebsbesuche mit einem Abstand von einem Monat. Eine unbedeutende Beziehung ergab sich für den Indikator ‚% Tiere mit erschwerter Atmung‘ sowohl für ein Erhebungsintervall von einem Monat aber auch für sechs Monate. Für die ‚Anzahl an verendeten Tieren während eines Jahres‘ konnte keine Korrelation berechnet werden, da das Auftreten zu gering war (Kirchner et al., 2014).

Tabelle 3: Wiederholbarkeit über die Zeit verschiedener Indikatoren zum Gesundheitsstatus bei Maststieren (Kirchner et al., 2014)

Indikator	r_s^a	n^c	r_s^b	n^d
% Tiere mit Nasenausfluss	0,48	59	0,15	49
% Tiere mit Augenausfluss	0,73	59	0,69	49
% Tiere mit erschwerter Atmung	0,03	59	0,19	49
Anzahl der Hustenereignisse pro Tier und 15 Minuten	0,55	59	0,27	49
% Tiere mit Diarrhö	0,43	59	0,63	49

^a Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von einem Monat

^b Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von sechs Monaten

^c Anzahl der Betriebe für die Berechnung von r_s^a

^d Anzahl der Betriebe für die Berechnung von r_s^b

h) Indikatoren des agonistischen Verhaltens

In Gruppen gehaltene Tiere verwenden agonistische und sozio-positive Interaktionen zum Aufbau und Erhalt einer sozialen Struktur innerhalb der Herde. Die Herdengröße (Kondo et al., 1989), das Haltungssystem beziehungsweise die Dimension des Haltungssystems (Wierenga, 1984) und das Management (Knierim, 1999) haben Einfluss auf das Auftreten und die Qualität von agonistischem Verhalten. In horntragenden Rinderherden ist die Frequenz von agonistischen Interaktionen positiv korreliert mit dem Auftreten von Hautverletzungen (Menke et al., 1999). In Herden mit enthornten Rindern äußern sich aggressive Interaktionen durch Verletzungen wie Hämatome (Laister et al., 2009).

Inter-Observer-Reliabilität:

Die Berechnung der Inter-Observer-Reliabilität erfolgte von Laister et al. (2009) mittels Erhebungen von zwei bis vier Beobachtern in sechs Milchviehbetrieben und zwei Rindermastbetrieben sowie anhand von Video-Clips. Die sechs Milchviehbetriebe und zwei Rindermastbetriebe befanden sich in Deutschland, Österreich und Italien. Erfasst wurden ‚Kopfstöße ohne Verdrängungen‘, ‚Verdrängungen‘, ‚Nachjagen‘, ‚Kämpfen‘ und ‚Aufjagen‘. Die Auswertung der Inter-Observer-Reliabilität erfolgte mit Spearmans Rangkorrelationskoeffizienten (r_s) und Kendalls Koeffizienten (W). Für ‚Nachjagen‘, ‚Kämpfen‘ und ‚Aufjagen‘ konnten jedoch keine statistischen Berechnungen erfolgen, da diese Verhaltensweisen zu selten aufgetreten waren. Bei Live-Beobachtungen variierte die Inter-Observer-Reliabilität der Verhaltensweise ‚Kopfstöße ohne Verdrängungen‘ bei zwei Beobachtern zwischen 0,61 und 1,00 (Rangkorrelation nach Spearman). Die Übereinstimmung der Korrelation von drei Beobachtern war 0,97 und von vier Beobachtern 0,84 (Rangkorrelation nach Kendall). Bei den ‚Verdrängungen‘ zeigte sich eine Korrelation (r_s) bei zwei Beobachtern von 0,81 bis 0,95. Die Übereinstimmung der Korrelation (W) von drei Beobachtern war 0,92 und von vier Beobachtern 0,85 (Laister et al., 2009).

Die Beurteilung der Inter-Observer-Reliabilität mit Videosequenzen wurde mit 55 Video-Clips und drei Beobachtern durchgeführt. ‚Nachjagen‘, ‚Kämpfen‘ und ‚Aufjagen‘ wurden selten oder nie festgestellt und keine Berechnungen durchgeführt. Der paarweise Abgleich ergab für ‚Kopfstöße ohne Verdrängungen‘ eine Korrelation zwischen 0,70 und 0,81. Der Kendalls Koeffizient (W) für alle drei Beobachter betrug 0,83. Die Inter-Observer-Reliabilität für ‚Verdrängungen‘ zwischen zwei Beobachtern ergab Korrelationen zwischen 0,71 und 0,84. Die Übereinstimmung von drei Beobachtern betrug $W=0,85$ (Laister et al., 2009).

Wiederholbarkeit über die Zeit:

Laister et al. (2009) untersuchten in ihrer Studie die Wiederholbarkeit über die Zeit von ‚Kopfstößen ohne Verdrängungen‘ und ‚Verdrängungen‘ in 43 Milchvieh- und 20 Mastrinderbetrieben in Österreich, Deutschland und Italien mit Haltungssystemen Anbindehaltung, Laufstall mit Tiefstreu und Laufstall mit Liegeboxen. In den Betrieben waren Milchkühe der Rassen Fleckvieh, Holstein Friesian und Brown Swiss vorzufinden. Bei den Mastrindern wurden Betriebe in Österreich und Deutschland mit den Rassen Fleckvieh, Limousin und Fleckvieh x Limousin Kreuzungen herangezogen. Die Haltung der Tiere war in Laufställen mit Liegeboxen oder auf Vollspalten (Laister et al., 2009).

Die Erhebung der Indikatoren erfolgte drei Mal pro Betrieb. Die zweite bzw. dritte Erhebung erfolgte 60 bzw. 180 Tage nach der ersten. In den Milchviehbetrieben mit Anbindehaltung wurde die dritte Erhebung bereits 120 Tage nach der ersten durchgeführt. Wie in Tabelle 4 ersichtlich, zeigte die Berechnung der Korrelation mit Spearmans Rangkorrelationskoeffizienten zwischen zwei Beobachtungen eine große Streuung von -0,06 bis 0,85. Beim Indikator ‚Kopfstöße ohne Verdrängungen‘ gab es unabhängig vom Zeitintervall bis auf zwei Ausnahmen mittelmäßige Korrelationen. Beim Indikator ‚Verdrängungen‘ konnte bei den Milchkühen in Laufstallhaltung für alle drei Zeitintervalle eine hohe Korrelation von über 0,7 erreicht werden. Ansonsten zeigten sich größere Schwankungen. Bezüglich der Übereinstimmung über drei Beobachtungen hinweg bewegte sich Kendall *W* im mittelmäßigen und hohen Bereich, wobei sechs von neun Haltungssystemen eine hohe Korrelation aufwiesen (Laister et al., 2009).

Tabelle 4: Spearmans Rangkorrelationskoeffizienten (r_s) für ein Zeitintervall von 60 (1_60), 120 (60_180) und 180 (1_180) Tage und Kendalls Koeffizient (*W*) für die Indikatoren ‚Kopfstöße ohne Verdrängungen‘ und ‚Verdrängungen‘ bei Milchkühen und Maststieren (Laister et al., 2009)

Indikator	Haltungssystem	r_s			<i>W</i>
		1_60	60_180	1_180	
Kopfstöße ohne Verdrängungen	Milchkühe in Anbindehaltung	0,41	0,58 ¹	-0,06 ²	0,54
	Milchkühe in Laufstallhaltung	0,65	0,60	0,57	0,74
	Maststiere mit einem Gewicht von 200-350 kg	0,76	0,63	0,58	0,79
	Maststiere mit einem Gewicht von >350 kg	0,49	0,66	0,66	0,74
Verdrängungen	Milchkühe in Anbindehaltung	0,43	0,28 ¹	0,18 ²	0,54
	Milchkühe in Laufstallhaltung	0,70	0,75	0,85	0,84
	Maststiere mit einem Gewicht von 200-350 kg	0,45	0,04	0,24	0,50
	Maststiere mit einem Gewicht von >350 kg	0,68	0,69	0,84	0,82

¹ Intervall 60 Tage (60_120) ² Intervall 120 Tage (1_120)

Winckler et al. (2007) evaluierten in der bereits näher vorgestellten Studie die Wiederholbarkeit über die Zeit des agonistischen Verhaltens bei acht deutschen Milchviehbetrieben im Jahr 2003. Bei der Erhebung wurde für die gesamte Herde die Anzahl an ‚Kopfstößen‘ und ‚Verdrängungen‘ gezählt. Die Ergebnisse der Korrelationsberechnung mit Spearman zeigten eine große Streuung zwischen -0,48 und 0,52 (Winckler et al., 2007).

Kirchner et al. (2014) berechneten bei Maststieren die ‚durchschnittliche Anzahl an Kopfstößen, Verdrängungen, Kämpfe und Nachjagen pro Stunde und Tier‘. In der bereits weiter oben ausführlich beschriebenen Studie zeigte sich für das kurze Zeitintervall (ein Monat) eine Korrelation nach Spearman von 0,74 und für den längeren Zeitabschnitt von sechs Monaten eine Korrelation von 0,48 (Kirchner et al., 2014).

i) Indikator Ausweichdistanz (‘ADF’)

Die Beziehung zwischen Mensch und Tier beziehungsweise die Wahrnehmung des Menschen durch das Tier haben einen erheblichen Einfluss auf das Wohlergehen und die Produktionsleistung (Boivin et al., 2003; Hemsworth, 2003.; Waiblinger et al., 2006). Rinder mit erhöhter Angst vor Menschen zeigen beispielsweise eine akute oder chronische Stressreaktion, haben eine geringere Milchleistung oder einen reduzierten Milchfluss (Breuer et al., 2000).

Inter-Observer-Reliabilität:

Windschnurer et al. (2008) untersuchten im Jahr 2006 die Inter-Observer-Reliabilität der Ausweichdistanz auf dem Fressplatz in 16 Milchviehbetrieben in Österreich. Die Erhebung erfolgte durch zwei Beobachter. Für die gesamte Stichprobe wurde eine hohe Übereinstimmung bei der Korrelation nach Spearman festgestellt. Die Korrelationsrechnung für ‚% Tiere ADF berührt‘ ergab 0,97 und für ‚% Tiere ADF >50cm‘ 0,92 (Windschnurer et al., 2008).

Ähnliche Untersuchungen wurden von denselben zwei Personen auch an Maststieren (in sechs Betrieben) durchgeführt. Bei dieser Erhebung zeigte sich eine Übereinstimmung von 0,52 für ‚% Tiere ADF berührt‘ und 0,66 für ‚% Tiere ADF >0,20cm‘. Im Vergleich zur Erhebung bei Milchkühen war damit das Niveau der Korrelation nach Spearman erheblich geringer (Windschnurer et al., 2009).

Wiederholbarkeit über die Zeit:

Die Wiederholbarkeit über die Zeit der Ausweichdistanz auf dem Fressplatz überprüften Winckler et al. (2007) in acht Holstein-Friesian-Betrieben in Deutschland. Die Ausweichdistanz wurde im Durchschnitt bei 80 Prozent der Milchkühe pro Betrieb

beurteilt. Es ergaben sich Korrelationen für zwei aufeinanderfolgende Erhebungen von 0,79 bis 0,91 (Winckler et al., 2007).

De Rosa et al. (2003) ermittelten in vier Milchviehbetrieben und zwei Büffelmilchviehbetrieben die Wiederholbarkeit über die Zeit der Ausweichdistanz. In der bereits weiter oben ausführlich beschriebenen Studie zeigte sich eine Übereinstimmung nach Kendall von 0,43 bis 0,70. (De Rosa et al., 2003).

Kirchner et al. (2014) erhoben die Ausweichdistanz in Stiermastbetrieben. Beim bereits weiter oben angeführten Versuchsablauf zeigte sich für die drei Indikatoren mit einem kürzeren Zeitintervall (ca. ein Monat) eine Korrelation nach Spearman auf einem geringen bis mittelmäßigen Niveau mit 0,54 für ‚% Tiere ADF <50cm aber nicht berührt‘, mit 0,29 für ‚% Tiere ADF 50-100cm‘ und mit 0,46 für ‚% Tiere ADF > 100 cm‘. Die Berechnung der Korrelation mit einem längeren Zeitintervall von 6 Monaten wies für alle drei Indikatoren eine mittelmäßige Korrelation auf (‚% Tiere ADF <50cm aber nicht berührt‘: 0,47, ‚% Tiere ADF 50-100cm‘: 0,67, ‚% Tiere ADF > 100 cm‘: 0,55; Kirchner et al., 2014).

j) Indikator Qualitative Verhaltenserfassung („QBA“)

Die Qualitative Verhaltenserfassung („QBA“) wurde vor einigen Jahren am Scottish Agricultural College entwickelt (Wemelsfelder et al., 2001). Das Verfahren beruht auf der Fähigkeit des menschlichen Beobachters, Details des Verhaltens der Tiere wahrzunehmen und eine Beurteilung der Körpersprache anhand von Begriffen wie ‚ruhig‘, ‚ängstlich‘ oder ‚zufrieden‘ durchzuführen. Solche Bezeichnungen werden für die Beschreibung von individuellen Unterschieden im Charakter oder Temperament von Tieren verwendet (Lanier et al., 2000; Gosling, 2001). Außerdem haben solche Bezeichnungen eine emotionale Bedeutung, da Informationen geliefert werden, die eine direkte Relevanz für das Wohlergehen der Tiere haben (Wemelsfelder und Farish, 2004; Wemelsfelder, 2007).

Inter-Observer-Reliabilität:

Die Inter-Observer-Reliabilität des Indikators ‚QBA Score‘ wurde von Andreasen et al. (2012) in 43 Milchviehbetrieben in Dänemark erhoben. Die Beobachtung der einzelnen Betriebe erfolgte zwischen 09:40 und 16:45 von zwei Personen und dauerte jeweils 20 Minuten. Die statistische Auswertung der erhobenen Daten erfolgte mit der Korrelation nach Spearman. Die Korrelation von Hauptkomponente 1 und Hauptkomponente 2 ergab Korrelationskoeffizienten von 0,72 und 0,56 (Andreasen et al., 2012).

Wiederholbarkeit über die Zeit:

Auch in der bereits ausführlich beschriebenen Studie von Kirchner et al. (2014) fand die Qualitative Verhaltenserfassung Anwendung bei Maststieren. Es konnte eine

Wiederholbarkeit über die Zeit von 0,68 für ein Zeitintervall von einem Monat festgestellt werden. Die Korrelation nach Spearman war für das längere Intervall mit sechs Monaten etwas niedriger mit 0,53 (Kirchner et al., 2014).

k) Indikatoren der Liegeplatznutzung

Für die Erhebung der Liegeplatznutzung sind in der Literatur mehrere Indizes zu finden, die das Liegeverhalten und die Stall- bzw. Liegeplatznutzung beschreiben. Zu den am häufigsten verwendeten Indizes gehören der ‚Cow Comfort Index‘ (CCI) und der ‚Stall Use Index‘ (SUI).

‚Cow Comfort Index‘ (CCI):

Der ‚Cow Comfort Index‘ beschreibt den Anteil liegender Kühe in den Liegeboxen an der Gesamtanzahl an Kühen, die sich in den Liegeboxen liegend und stehend befinden (Cook et al., 2005; Overton et al., 2002; Krawczel et al., 2008). Erstrebenswert ist ein Wert von über 85 Prozent für den ‚Cow Comfort Index‘ (Overton et al., 2002).

‚Stall Use Index‘ (SUI):

Der Stall Use Index beschreibt den Anteil der liegenden Kühe in den Liegeboxen an der Gesamtanzahl an Kühen im Stall, die nicht fressen (Cook et al., 2005; Krawczel et al., 2008).

Für die Indikatoren ‚Cow Comfort Index‘ und ‚Stall Use Index‘ gibt es bisher keine Erhebungen zur Reliabilität.

l) Indikator Aufstehvorgang

Probleme beim Aufstehen können bei Rindern zu Läsionen an den Sprunggelenken, Knien oder Zitzen führen (Mortensen, 1978). Weiters kann es zu einem erhöhten Herzschlag (Müller et al., 1989) und einer Veränderung in der Cortisolausschüttung (Ladewig und Smidt, 1989) kommen.

Inter-Observer-Reliabilität:

Chaplin und Munksgaard (2001) erhoben den Aufstehvorgang mit dem fünfstufigen Scoring-System bei 61 Holstein-Friesian-Milchkühen in Dänemark. Die Rinder wurden in zwei Ställen mit Anbindehaltung gehalten. Die Erhebung der Inter-Observer-Reliabilität erfolgte an fünf aufeinanderfolgenden Tagen jeweils um 11.30 Uhr durch zwei Personen. Die Kühe wurden von einer Person, die etwas entfernt hinter dem Schwanz der Kuh stand, zum Aufstehen animiert, indem sie angesprochen wurden. Wenn dies nicht ausreichend war, wurde die Milchkuh mit Berührungen auf dem Rücken zum Aufstehen aufgefordert. Sollte dies zu wenig sein,

wurden leichte Schläge auf den Rücken gegeben, um die Kuh zum Aufstehen aufzufordern. Die Übereinstimmung (21 bis 34 Aufstehvorgänge) zwischen zwei Beobachtern bewegte sich zwischen 53 und 81 Prozent (Chaplin und Munksgaard, 2001).

Wiederholbarkeit über die Zeit:

Plesch et al. (2010) evaluierten die Wiederholbarkeit über die Zeit für die Beurteilung des Aufstehvorganges in 35 Milchviehbetrieben in Österreich und Deutschland. Von August 2005 bis April 2006 wurden die Betriebe drei Mal besucht. Unter anderem wurde die Aufstehdauer in Sekunden erhoben. Die zweite und dritte Erhebung wurde 60 und 180 Tage nach der ersten Erhebung durchgeführt. In Betrieben mit Anbindehaltung erfolgte die dritte Erhebung bereits 120 Tage nach der ersten Erhebung. Die statistische Berechnung der Wiederholbarkeit über die Zeit erfolgte mit dem Kendall Koeffizienten (W) und brachte ein Ergebnis von 0,74 (Plesch et al., 2010).

3.4 Reliabilität von aggregierten Messgrößen des Wohlergehens

Für aggregierte Messgrößen des Wohlergehens ist zum Thema Reliabilität kaum Literatur zu finden. Vom europäischen Forschungsprojekt Welfare Quality[®] gibt es für aggregierte Messgrößen keine Studien für Milchkühe. Kirchner et al. (2014) evaluierten die Wiederholbarkeit über die Zeit des Welfare Quality[®] Erhebungsprotokolls bei Maststieren.

Wiederholbarkeit über die Zeit:

Kirchner et al. (2014) untersuchten die Wiederholbarkeit über die Zeit in Stiermastbetrieben in Österreich, Deutschland und Italien für das Welfare Quality[®] Erhebungsprotokoll. Der Versuchsablauf wurde bereits oben näher beschrieben. Aus den einzelnen Indikatoren des Welfare Quality[®] Erhebungsprotokolls werden die Welfare-Kriterien gebildet (Welfare Quality[®], 2009). Wie in Tabelle 5 zu sehen ist, wurde für zehn Welfare-Kriterien die Wiederholbarkeit über die Zeit berechnet. Bei einem Intervall von einem Monat zeigten fünf von zehn Welfare-Kriterien eine hohe Korrelation von über 0,7 und vier von zehn Welfare-Kriterien eine mittelmäßige Korrelation von über 0,4. Einzig beim Welfare-Kriterium ‚Abwesenheit von andauerndem Hunger‘ war die Korrelation mit 0,27 gering. Die Korrelation nach Spearman war bei den Welfare-Kriterien mit einem Intervall von sechs Monaten bei allen auf einem mittleren bis hohen Niveau. Auch das Welfare-Kriterium ‚Abwesenheit von andauerndem Hunger‘ konnte mit 0,68 eine beachtlich höhere Korrelation aufweisen.

In einem nächsten Schritt werden beim Welfare Quality[®] Erhebungsprotokoll aus den Welfare-Kriterien die Welfare-Prinzipien gebildet (Welfare Quality[®], 2009). Diese in

Tabelle 6 ersichtlichen Welfare-Prinzipien zeigten eine Wiederholbarkeit über die Zeit für ein Intervall von einem Monat von über 0,69. Bis auf das Welfare-Prinzip ‚Gute Fütterung‘ war die Korrelation bei den anderen Dreien mit einem längeren Zeitraum von sechs Monaten deutlich geringer.

Tabelle 5: Wiederholbarkeit über die Zeit der Welfare-Kriterien bei Maststieren (Kirchner et al., 2014)

Welfare-Kriterien	r_s^a	n^b	r_s^c	n^d
Abwesenheit von andauerndem Hunger	0,27	59	0,68	49
Abwesenheit von andauerndem Durst	0,67	59	0,73	49
Liegekomfort	0,70	59	0,43	49
Bewegungsfreiheit	0,73	59	0,70	49
Abwesenheit von Verletzungen	0,57	59	0,53	49
Abwesenheit von Krankheiten	0,76	59	0,55	49
Abwesenheit von Schmerzen	0,92	59	0,99	49
Ausdruck von Sozialverhalten	0,78	59	0,56	49
Mensch-Tier-Beziehung	0,61	59	0,72	49
Positiver emotionaler Zustand	0,68	59	0,53	49

^a Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von einem Monat

^b Anzahl der Betriebe für r_s^a

^c Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von einem Monat

^d Anzahl der Betriebe für r_s^a

Tabelle 6: Wiederholbarkeit über die Zeit der Welfare-Prinzipien bei Maststieren (Kirchner et al., 2014)

Welfare-Prinzipien	r_s^a	n^b	r_s^c	n^d
Gute Fütterung	0,72	59	0,74	49
Gute Haltung	0,79	59	0,60	49
Gute Gesundheit	0,82	59	0,52	49
Artgemäßes Verhalten	0,69	59	0,41	49

^a Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von einem Monat

^b Anzahl der Betriebe für r_s^a

^c Spearman Rangkorrelationskoeffizient (r_s) für ein Intervall von einem Monat

^d Anzahl der Betriebe für r_s^a

4 Tiere, Material und Methode

4.1 Betriebe

Zur Bearbeitung der Fragestellung wurden 30 Milchviehbetriebe an verschiedenen Standorten in Kärnten, der Steiermark und Oberösterreich besucht.

Folgende Voraussetzungen waren neben der Zusage der Betriebsleiter für die Auswahl der Betriebe notwendig:

- Laufstallhaltung
- Milchkühe der Rasse Fleckvieh und/oder Holstein-Friesian

Um einen Überblick über die landwirtschaftlichen Betriebe im jeweiligen Bundesland zu bekommen, wurde vom Landeskontrollverband eine Liste von Betrieben mit Milchviehhaltung angefordert. In einem ersten Schritt wurden die Betriebsführenden telefonisch kontaktiert und über das Prozedere der Erhebung informiert. Im Großen und Ganzen war es nicht schwierig, Betriebe für die Erhebung zu gewinnen. Nur wenige Betriebsführende hatten kein Interesse an dem Projekt teilzunehmen. In einem weiteren Telefonat wurde der konkrete Termin für die Datenerhebung vereinbart. Die Durchführung der Erhebung in den Betrieben wurde von den Landwirt/innen aufmerksam verfolgt. Viele Fragen wurden zu den einzelnen Indikatoren gestellt. Generell konnte eine sehr gute Zusammenarbeit mit den Bauernfamilien festgestellt werden.

In Kärnten wurden 15 Betriebe, in der Steiermark 13 Betriebe und in Oberösterreich 2 Betriebe besucht. 14 Betriebe hielten Milchkühe der Rasse Fleckvieh, 8 Betriebe Milchkühe der Rasse Holstein-Friesian, 7 Betriebe Milchkühe der Rasse Fleckvieh und Holstein-Friesian und ein Betrieb Milchkühe der Rasse Holstein-Friesian und Braunvieh. Auf dem Betrieb mit den Rassen Holstein-Friesian und Braunvieh wurden die Milchkühe der Rasse Braunvieh bei den herdenbezogenen Erhebungen berücksichtigt, bei den Einzeltiererhebungen jedoch nicht.

In den Betrieben waren im Durchschnitt 40,8 Milchkühe ($n=30$; $MIN=20$; $MAX=67$; $SD=11,8$) vorzufinden. Bei der Datenerhebung wurden jedoch Tiere, die nicht laktierend waren oder nicht mit laktierenden Tieren gemeinsam gehalten wurden, nicht berücksichtigt. Dies waren Kühe in der Abkalbebox, Kühe in der Krankenbox oder trockenstehende Kühe in einer eigenen Trockensteherbox. 18 Betriebe hatten zum Zeitpunkt der Datenerhebung Kühe in der Trockensteherbox. In den 12 anderen Betrieben gab es entweder keine trockenstehenden Kühe oder sie wurden mit den laktierenden Kühen gehalten. Für die Datenerhebung gab es somit am Tag 1 der Erhebung im Durchschnitt 37,6 Milchkühe ($SD=9,48$, $MIN=20$, $MAX=60$, $n=30$) und am Tag 2 der Erhebung im Durchschnitt 37,7 Milchkühe ($SD=9,33$, $MIN=20$, $MAX=60$, $n=30$) pro Betrieb.

4.2 Haltungssysteme

Das vorherrschende Haltungssystem waren Liegeboxenlaufställe mit Tiefboxen (n=28), in zwei Betrieben wurden Hochboxen verwendet. Sechs verschiedene Melksysteme waren in den Betrieben vorzufinden: Fischgrätenmelkstand (16), Tandemmelkstand (9), Swing-Over Melkstand (2), Melkroboter (1), Rohrmelkanlage (1) und Side-by-Side Melkstand (1). Im Betrieb mit der Rohrmelkanlage wurden die Milchkühe für den Melkvorgang in den Altstall mit Anbindehaltung getrieben.

4.3 Datenerhebungsmethode

Um die kurzfristige Wiederholbarkeit über die Zeit des Welfare Quality® Erhebungsprotokolls für Milchkühe, der qualitativen Beurteilung des Aufstehvorganges und für die Indizes der Liegeplatznutzung zu berechnen, wurden alle Betriebe zwei Mal im Abstand von vier Tagen besucht. Die Datenerhebung erfolgte zwischen Februar und Mai 2013.



Abbildung 3: Beispielhafte Abbildung des Ablaufs der Erhebungen zur Ermittlung der kurzfristigen Wiederholbarkeit über die Zeit in den 30 Betrieben

4.4 Beobachtende Personen

Die Datenerhebung in den Betrieben erfolgte durch zwei Beobachter. Beide Personen waren Studierende der Universität für Bodenkultur. Eine Person war männlich und die andere weiblich. Von jedem Beobachter wurden jeweils 15 Betriebe beurteilt. Vor Beginn der Datenerhebung wurden sie in einer zweitägigen Schulung für die Erhebung der Indikatoren trainiert. An beiden Tagen wurden in Theorieeinheiten am Vormittag die einzelnen Indikatoren des Welfare Quality® Erhebungsprotokolls, die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung besprochen. An beiden Nachmittagen wurde in den zwei Betrieben der Ablauf der Erhebungen praktisch durchgeführt. Im Weiteren dienten Videos und Fotos einiger Indikatoren wie ‚Body Condition Score‘, ‚Abliegezeit‘, ‚Lahmheit‘, ‚Sauberkeit von Euter, Unterbein und Hinterhand‘ und ‚Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen‘. Die Übereinstimmung der

Beobachter mit dem Goldstandard für die einzelnen Indikatoren zeigt sich in Tabelle 7.

Tabelle 7: Beobachterübereinstimmung mit dem Goldstandard für ausgewählte Indikatoren des Welfare Quality® Erhebungsprotokolls

Indikatoren	Beobachter 1		Beobachter 2	
	n	Übereinstimmung	n	Übereinstimmung
Sauberkeit des Unterbeins	46 ¹	83 %	46 ¹	89 %
Sauberkeit der Hinterhand	57 ¹	86 %	57 ¹	79 %
Sauberkeit des Euters	57 ¹	82 %	57 ¹	88 %
Body Condition Score	32 ¹	91 %	32 ¹	84 %
Lahmheit	39 ²	54 %	39 ²	77 %
Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen	79 ²	81 %	84 ²	75 %
Abliegezeit	85 ²	0,97 ³	85 ²	0,95 ³

¹ Anzahl an Fotos

² Anzahl an Videos

³ Spearman Rangkorrelationskoeffizient (r_s)

4.5 Datenerhebung

Nach Beendigung des Morgenmelkens wurde mit der Erhebung der Daten begonnen. Der Ablauf der Datenerhebung wurde in jedem Betrieb gleich gestaltet. Die Indikatoren wurden in folgender Reihenfolge erhoben: Ausweichdistanz auf dem Fressplatz, Qualitative Verhaltensbeurteilung, Verhaltenserfassung Sozial- und Ruheverhalten, Herd Scan, Aufstehvorgang, Einzeltieruntersuchung, Checkliste Ressourcen und Fragen zum Management.

a) Ausweichdistanz auf dem Fressplatz

Mit dem Indikator Ausweichdistanz auf dem Fressplatz („ADF“) (Welfare Quality®, 2009) wurde die Datenerhebung im Betrieb gestartet. Nachdem die Milchkühe den Melkstand verlassen hatten, wurden die Kühe im Fressgitter fixiert. Aus einer Entfernung von 200 cm, mit einer Geschwindigkeit von einem Schritt pro Sekunde, einer Schrittlänge von ca. 60 cm und nach vorne in einem Winkel von 45° ausgestrecktem Arm näherte sich der Beobachter der Kuh. Sollte vor den Kühen nicht 200 cm Platz sein, so wurde die Annäherung in einem Winkel von max. 45° und einer Distanz von 250 cm gewählt. Bei Auftreten einer Ausweichreaktion wurde die Distanz zwischen Hand und Flotzmaul in 10cm-Schritten geschätzt. Bei einer Berührung des Flotzmaules mit der Hand war die Distanz 0 cm. Die Anzahl an erhobenen Tieren orientierte sich an der Herdengröße wie in Tabelle 8 ersichtlich. Die Erhebung dauerte pro Kuh ca. eine Minute.

Tabelle 8: Anzahl der Milchkühe für die Erhebung der Ausweichdistanz auf dem Futterplatz (Welfare Quality[®], 2009)

Herdengröße	Anzahl zu bewertender Milchkühe
30	30
40	30
50	33
60	37
70	41

In Summe wurde in den 30 Betrieben die Ausweichdistanz 1917 Mal erhoben. Am Tag 1 der Erhebung erfolgte die Erfassung im Durchschnitt bei 31,9 Milchkühen (SD=5,57, MIN=20, MAX=50, n=30) bei insgesamt 957 Erhebungen. Am 2. Tag wurden insgesamt 960 Milchkühe und im Durchschnitt 32,0 Milchkühe pro Betrieb (SD=6,43, MIN=20, MAX=55, n=30) erfasst.

b) Qualitative Verhaltensfassung

Die Beurteilung der Verhaltensqualität (Welfare Quality[®], 2009) erfolgte anhand von 20 verschiedenen Begriffen wie ‚ruhig‘ oder ‚neugierig‘. Das Ausmaß, in dem die einzelnen expressiven Qualitäten des Verhaltens wahrgenommen werden, wurde auf einer visuellen Analog-Skala von 125 mm Länge erfasst. Umso stärker eine Eigenschaft in der Herde ersichtlich war umso weiter rechts wurde die Linie markiert. Schlussendlich wurde die Entfernung in Millimeter vom Minimum (0 mm) zum markierten Punkt gemessen.

Jede Herde wurde für 20 Minuten beobachtet. Bei den gesamten Erhebungen wurden zwischen ein bis zwei Beobachtungspunkte gewählt. Bei einem Beobachtungspunkt wurde die Herde 20 Minuten vom selben Punkt aus beobachtet und bei zwei Beobachtungspunkten jeweils für 10 Minuten. Für eine bessere Sicht wurde die Beobachtung von einer Klappleiter vom Futtertisch aus durchgeführt.

c) Verhaltensbeobachtung Sozial- und Ruheverhalten

Bei der Verhaltensbeobachtung Sozial- und Ruheverhalten (Welfare Quality[®], 2009) wurde eine Reihe an Indikatoren erhoben. Dazu zählen Indikatoren des Ruheverhaltens (‚Durchschnittliche Abliegezeit‘, ‚% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren‘, ‚% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen‘), Indikatoren des agonistischen Verhaltens (‚Durchschnittliche Anzahl an Kopfstößen pro Tier und Stunde‘, ‚Durchschnittliche Anzahl an Verdrängungen pro Tier und Stunde‘) sowie ein Indikator des Gesundheitsstatus (‚Anzahl der Hustenereignisse pro Tier und

15 Minuten'). Beim Indikator ‚Durchschnittliche Anzahl an Verdrängungen pro Tier und Stunde‘ wurden neben den ‚Verdrängungen‘ auch die agonistischen Verhaltensweisen ‚Nachjagen‘, ‚Kämpfen‘ und ‚Aufjagen‘ einberechnet.

In Summe wurden für diese Erhebung ca. 150 Minuten benötigt, wobei die reine Beobachtungszeit 120 Minuten dauerte. Für die Erhebung wurde der Stall in Segmente unterteilt. Pro Betrieb wurde die Lauf- und Liegefläche der Milchkühe in vier bis sechs Segmente aufgeteilt. In 9 Betrieben wurden vier Segmente, in 15 Betrieben fünf Segmente und in 6 Betrieben sechs Segmente gewählt. Jedes Segment wurde in der Beobachtungszeit von 120 Minuten zwei Mal erhoben, woraus sich folgende Beobachtungszeiten pro Segment ergeben: bei vier Segmenten 15 Minuten, bei fünf Segmenten 12 Minuten und bei sechs Segmenten 10 Minuten. Die Datenerhebung erfolgte erhöht auf einer Klappleiter auf dem Futtertisch. Bei Betrieben mit einem Außenbereich wurde dieser als ein Segment angesehen und die Erhebung erfolgte im Freien.

Für jedes Segment wurden vor der Beobachtung die Anzahl der stehenden Kühe, die Anzahl der liegenden Kühe sowie die Anzahl der Kühe, die teilweise oder komplett außerhalb der Liegefläche liegen, erhoben. Während der Beobachtungszeit wurde die Anzahl an agonistischen Verhaltensweisen wie ‚Kopfstöße‘, ‚Verdrängungen‘, ‚Nachjagen‘, ‚Kämpfen‘ und ‚Aufjagen‘ gezählt. Außerdem wurde die Anzahl an hörbaren Hustenereignissen gezählt, die Abliegezeit der Kühe mit einer Stoppuhr gemessen und darauf geachtet, ob die Kuh dabei mit der Haltungseinrichtung kollidiert. Nach Ablauf der Beobachtungszeit wurden für das Segment erneut die Anzahl der stehenden Kühe, der liegenden Kühe sowie der Kühe, die teilweise oder komplett außerhalb der Liegefläche lagen, erhoben.

In Summe wurde die Abliegedauer bei 554 Milchkühen ermittelt, 276 am Tag 1 und 278 am Tag 2. Am Tag 1 der Erhebung wurden pro Betrieb 9,2 Abliegevorgänge (SD=2,04, MIN=6, MAX=14, n=30) und am Tag 2 9,27 Abliegevorgänge pro Betrieb (SD=2,5, MIN=7, MAX=16, n=30) beurteilt.

d) Herd Scan

Der Herd Scan wurde parallel zu den oben erläuterten Verhaltensbeobachtungen durchgeführt. Dieser dient der Feststellung der Indizes der Liegeplatznutzung (‚Cow Comfort Index‘ und ‚Stall Use Index‘) (Cook et al., 2005; Overton et al., 2002; Krawczel et al., 2008). Bevor für ein Segment die Verhaltensbeobachtung erfolgte, wurde der Herd Scan durchgeführt. Erhoben wurde für die ganze Herde die Anzahl der Kühe, die liegen, die fressen oder trinken, die in der Liegefläche stehen und die irgendwo im Stall stehen. Bei Beobachtungen in beispielsweise fünf Segmenten wurde der Herd Scan in Summe elf Mal erhoben. Zehn Mal vor und einmal am Ende der Verhaltensbeobachtung. In Summe wurde der Herd

Scan auf 28 Betrieben jeweils am 1. Tag und am 2. Tag der Erhebung erhoben. Beobachter 1 und Beobachter 2 haben jeweils in 14 Betrieben den Herd Scan erhoben. Beide Beobachter verzichteten aus organisatorischen Gründen am jeweils ersten Betrieb auf die Erhebung des Herd Scan.

e) Aufstehvorgang

Für die qualitative Beurteilung des Aufstehvorgangs wurde im Ablauf der Erhebungen zum ersten Mal die Lauffläche der Milchkühe betreten. Die liegenden Kühe sollten mit so wenig Einsatz wie möglich (Stimme) zum Aufstehen ermutigt werden. Der Aufstehvorgang wurde mit dem Scoring-System nach Chaplin und Munksgaard (2001), verändert durch Hauschild (2012), beurteilt. Eine genauere Beschreibung des Scoring-Systems befindet sich in Tabelle 9. Die Anzahl der Aufstehvorgänge orientierte sich an der Anzahl der liegenden Kühe. In Summe wurden 693 Aufstehvorgänge beurteilt. Am Tag 1 wurden in Summe 347 Aufstehvorgänge und im Durchschnitt pro Betrieb 11,6 Aufstehvorgänge (SD=4,16, MIN=6, MAX=20, n=30) beurteilt. Am 2. Tag der Erhebung wurden insgesamt 346 Aufstehvorgänge und pro Betrieb im Durchschnitt 11,5 Aufstehvorgänge (SD=4,38, MIN=6, MAX=23, n=30) beurteilt.

Tabelle 9: Scoring-System für die Beurteilung des Aufstehvorganges (Chaplin und Munksgaard, 2001; verändert durch Hauschild, 2012)

Score	Beschreibung
1	gleichmäßige, fließende Bewegung, normale Bewegungsabfolge
2	kurze Pause (<3s) auf Karpalgelenken, normale Bewegungsabfolge
3	lange Pause (>3s) auf Karpalgelenken, normale Bewegungsabfolge
4	lange Pause auf den Karpalgelenken und/oder Schwierigkeiten beim Aufstehen
5	Abweichung von der normalen Bewegungsabfolge

f) Einzeltieruntersuchung

Bei der Einzeltieruntersuchung (Welfare Quality[®], 2009) wurden folgende Indikatoren erhoben: Body Condition Score, Sauberkeit des Euters, des Unterbeins und der Hinterhand, Lahmheit, Integumentschäden und Indikatoren zum Gesundheitsstatus (Nasenausfluss, Augenausfluss, erschwerte Atmung, Diarrhö, Vulvausfluss).

Die Erhebung der verschiedenen Indikatoren erfolgte auf der Lauffläche der Rinder. Die Milchkühe wurden dazu im Fressgitter fixiert. Eine gewisse Anzahl an Milchkühen befand sich meistens im Fressgitter, weitere Kühe konnten durch das Anschieben des Futters zum Futtertisch gelockt werden. Da vorab

der Indikator Aufstehvorgang erhoben wurde, standen die Kühe bereits und konnten leichter zum Futtertisch bewegt werden.

Die Anzahl an erhobenen Tieren orientierte sich an der Herdengröße wie in Tabelle 8 ersichtlich.

Die Erhebung der Daten dauerte pro Kuh zirka drei Minuten. Die gesamte Einzeltieruntersuchung wurde für jede Kuh separat durchgeführt. Begonnen wurde damit, die Ohrmarken- oder die Transpondernummer jeder Kuh sowie die Rasse (Milchrasse oder Doppelnutzungsrasse) zu notieren. Dann wurde der Indikator Body Condition Score erhoben. Verschiedene vorgegebene Körperregionen wurde nur mit dem Auge einer Gesamtbewertung von ‚normal‘, ‚sehr mager‘ oder ‚sehr dick‘ unterzogen. Bei den Indikatoren der Sauberkeit wurde nur entweder die rechte oder linke Körperseite der Kuh begutachtet. In Summe wurde darauf geachtet, pro Betrieb eine gleiche Anzahl an Erhebungen von linken und rechten Körperseiten zu haben. Die Sauberkeit des Euters, der Hinterhand und des Unterbeins wurde jeweils mit ‚verschmutzt‘ oder ‚sauber‘ bewertet. Bei den Integumentschäden wurde die gleiche Körperseite wie bei der Sauberkeit bewertet. Erhoben wurden haarlose Stellen, Läsionen und Schwellungen von sechs verschiedenen Körperregionen. Die Indikatoren des Gesundheitsstatus (Nasenausfluss, Augenausfluss, erschwerte Atmung, Diarrhö, Vulvausfluss) wurden mit ‚Ja‘ oder ‚Nein‘ bewertet. Mit dem Indikator Lahmheit wurde die klinische Bewertung abgeschlossen. Jede Kuh wurde nach den Kriterien ‚nicht lahm‘, ‚geringgradig lahm‘ und ‚hochgradig lahm‘ beurteilt. Dazu wurde die Kuh aus dem Fressgitter freigelassen und in Bewegung benotet.

Die klinische Bewertung wurde in den 30 Betrieben in Summe 1857 Mal durchgeführt. Am Tag 1 wurden im Durchschnitt 30,9 Milchkühe pro Betrieb (SD=3,99, MIN=20, MAX=41, n=30) und in Summe 926 Milchkühe erhoben. Am Tag 2 erfolgte eine Beurteilung von im Durchschnitt 31,0 Milchkühen pro Betrieb (SD=3,93, MIN=20, MAX=41, n=30); in Summe wurden am Tag 2 931 Milchkühe bewertet.

g) Checkliste Ressourcen

Bei der Checkliste Ressourcen (Welfare Quality[®], 2009) wurden alle im Stall für die Milchkühe zugänglichen Wasserquellen erhoben und bewertet. Für jede Tränke wurde Folgendes erhoben: Art der Tränke (Tränkebecken oder Wassertrog), Sauberkeit der Tränke (sauber, teilweise verschmutzt, verschmutzt), Funktionstüchtigkeit der Tränke (Ja/Nein) und die Wasserdurchflussmenge.

h) Fragen zum Management

Die Fragen zum Management (Welfare Quality[®], 2009) wurden an den Betriebsführer/an die Betriebsführerin gestellt.

Sie umfassten:

- Wie viele Kühe werden durchschnittlich im Jahr im Betrieb gehalten?
- Wie viele Tage im Jahr und wie viele Stunden am Tag haben die Kühe Zugang zu einer Auslauffläche?
- Wie viele Tage im Jahr und wie viele Stunden am Tag haben die Kühe Zugang zur Weide?
- Wie viele festliegende Kühe gab es während der letzten 12 Monate?
- Wie viele Kühe hatten während der letzten 12 Monate eine Schweregeburt?
- Wie viele Kühe verstarben während der letzten 12 Monate im Betrieb aufgrund einer Krankheit, eines Unfalles oder Euthanasie?
- Gehalt an somatischen Zellen in der Milch: Wie viel Kühe hatten in den letzten drei Monaten einen somatischen Milchzellgehalt von über 400.000?
- Enthornung: Werden die Rinder im Betrieb enthornt? Wie hoch ist der prozentuelle Anteil an enthornten Kälbern? Mit welchem Alter werden die Kälber enthornt? Wird die Enthornung mit einem Brenneisen oder mit einer Ätzpaste durchgeführt? Werden die Kälber vor dem Enthornen betäubt (Lokalanästhesie)? Erhalten die Kälber eine postoperative Schmerzbehandlung?

4.6 Statistische Auswertung

Die Aufbereitung der Daten erfolgte in Microsoft Excel, Version 2010. Dabei wurden alle Indikatoren, Welfare-Kriterien, Welfare-Prinzipien und die Gesamtbeurteilung für jeden Betrieb wie im Welfare Quality[®] Erhebungsprotokoll beschrieben berechnet (Welfare Quality[®], 2009). Bei der qualitativen Beurteilung des Aufstehvorganges wurde der prozentuelle Anteil der jeweiligen Scores ausgewiesen (Chaplin und Munksgaard, 2001). Die Berechnung für die Indizes der Liegeplatznutzung ‚Cow Comfort Index‘ und ‚Stall Use Index‘ erfolgte wie folgt:

‚Cow Comfort Index‘ (CCI) = Anzahl liegender Kühe in Liegeboxen / Anzahl liegender und stehender Kühe in Liegeboxen (Cook et al., 2005)

‚Stall Use Index‘ (SUI) = Anzahl liegender Kühe in Liegeboxen / Gesamtanzahl der Kühe im Stall, die nicht fressen oder trinken (Cook et al., 2005)

In einer ersten statistischen Analyse wurden die einzelnen Indikatoren und aggregierten Messgrößen auf Normalverteilung mit SAS 9.2 für Windows überprüft. Der Großteil der Indikatoren und aggregierten Messgrößen wies keine Normalverteilung auf, daher wurde für die Korrelationsberechnung die Methode nach Spearman mit SAS 9.2 für Windows gewählt.

Das Kapitel *Status des Wohlergehens in den Untersuchungsbetrieben* gibt einen Überblick über Mittelwert, Standardabweichung, Minimum und Maximum der einzelnen Indikatoren und aggregierten Messgrößen an den beiden Tagen der Erhebung auf Herdenebene. Für binäre Merkmale, die als Fragen zum Management erhoben wurden, wird der prozentuelle Anteil ausgewiesen.

Für die kurzfristige Wiederholbarkeit über die Zeit wurden die Daten der Indikatoren und aggregierten Messgrößen auf Herdenebene vom 1. und 2. Tag der Erhebung miteinander sowohl für die gesamte Stichprobe als auch für die Betriebe von Beobachter 1 und Beobachter 2 separat korreliert. Einige Indikatoren wurden in der Korrelationsberechnung nicht berücksichtigt, da es entweder zu keinen Veränderungen gekommen war oder das Auftreten bei der Erhebung zu selten war. Bei den Fragen zum Management blieben aufgrund des Erhebungsintervalls folgende Daten unverändert: ‚Zugang zu einer Freilauffläche‘, ‚Anzahl an Tagen im Jahr mit Zugang zu einer Freilauffläche‘, ‚Anzahl an Stunden pro Tag mit Zugang zu einer Freilauffläche‘, ‚Zugang zu einer Weide‘, ‚Anzahl an Tagen im Jahr mit Zugang zu einer Weide‘, ‚Anzahl an Stunden pro Tag mit Zugang zu einer Weide‘, ‚% Tiere mit subklinischer Mastitis‘, ‚% verendeter Tiere während eines Jahres‘, ‚% Tiere mit Schweregeburt‘, ‚% Tiere mit Festliegen‘, ‚Enthornung‘, ‚Einsatz von Betäubungsmitteln‘, ‚Einsatz von Schmerzmitteln‘ und ‚Kupieren der Schwänze‘. Auch die Indikatoren ‚Anzahl Tränkebecken‘, ‚Anzahl Wassertröge‘, ‚Länge der Wassertröge‘, ‚Anzahl Wassertröge und Tränkebecken mit ausreichendem Wasserdurchfluss‘ und ‚Länge der Wassertröge mit ausreichendem Wasserdurchfluss‘ blieben unverändert. Für die Indikatoren ‚Anzahl der Hustenereignisse pro Tier und 15 Minuten‘, ‚% Tiere mit Nasenausfluss‘, ‚% Tiere mit Augenausfluss‘ und ‚% Tiere mit erschwerter Atmung‘ konnte aufgrund des seltenen Auftretens bei der Erhebung keine Korrelationsrechnung durchgeführt werden.

In einem weiteren Schritt wurden zwei kleinere Datensätze geschaffen, die sich hinsichtlich der Übereinstimmung der Kühe in der Stichprobe bei der Einzeltieruntersuchung an Tag 1 und Tag 2 unterscheiden. Eine Untergruppe bestand aus Betrieben mit einer 100%igen Übereinstimmung der Kühe in der Stichprobe bei der Einzeltieruntersuchung (6 Betriebe), die andere aus Betrieben mit einer <70%igen Übereinstimmung bei der Einzeltieruntersuchung (6 Betriebe). Im Weiteren wird für die Betriebe mit einer 100%igen Übereinstimmung der Kühe der Begriff ‚identische Stichprobe‘ und für die Betriebe mit einer <70%igen Übereinstimmung der Begriff ‚abweichende Stichprobe‘ verwendet.

In einem ersten Schritt wurden bei den sechs Betrieben mit einer identischen Stichprobe der Frage nachgegangen, wie hoch der prozentuelle Anteil an gleichen

Bewertungen an den beiden Erhebungstagen bei Betrieben mit einer identischen Stichprobe bei der Einzeltieruntersuchung ist. Dazu wurde bei jeder einzelnen Milchkuh die Übereinstimmung der Indikatoren der Einzeltieruntersuchung (Sauberkeit, Body Condition Score, Integumentschäden, Diarrhö, Lahmheit) beider Beurteilungen analysiert. Pro Betrieb wurden Mittelwert (MW) und Standardabweichung (SD) berechnet, sowie das Minimum (MIN) und Maximum (MAX) für die sechs Betriebe.

Anschließend wurden die Indikatoren der Einzeltieruntersuchung beider Gruppen miteinander verglichen. Dazu wurde eine Korrelationsberechnung nach Spearman durchgeführt. Bei den Betrieben mit einer identischen Stichprobe wurden im Durchschnitt 27,2 Milchkühe pro Betrieb (n=6; MIN=20; MAX=33; SD=5,3) gehalten. In den Betrieben mit einer abweichenden Stichprobe konnten im Durchschnitt 50,8 Milchkühe pro Betrieb (n=6; MIN=39; MAX=60; STABW=7,2) beobachtet werden.

5 Ergebnisse

5.1 Status des Wohlergehens in den Untersuchungsbetrieben

Der Indikator ‚Sauberkeit der Wasserstellen‘ wird mit ‚sauber‘, ‚teilweise verschmutzt‘ und ‚verschmutzt‘ ausgewiesen. Am 1. Tag der Erhebung wurden im Mittel 30 Prozent der Wasserstellen der Betriebe als ‚sauber‘ und 70 Prozent der Wasserstellen als ‚teilweise verschmutzt‘ bewertet. Am 2. Tag der Erhebung waren 23,3 Prozent der Wasserstellen ‚sauber‘ und 76,7 Prozent der Wasserstellen in den Betrieben ‚teilweise verschmutzt‘. Es gab keinen Betrieb mit ‚verschmutzten‘ Wasserstellen. Tabelle 10 enthält die Ergebnisse der im Welfare Quality® Erhebungsprotokoll enthaltenen Managementfaktoren. Während mehr als die Hälfte der Betriebe einen Laufhof anboten, wurde Weidegang nur in einem Drittel der Betriebe durchgeführt. Ein Einsatz von Betäubungsmitteln bei der Kälberenthornung fand in mehr als 75% der Betriebe statt; postoperative Schmerzbekämpfung dagegen nur bei etwa einem Viertel der Betriebe.

Tabelle 10: Managementbezogene Indikatoren des Welfare Quality® Erhebungsprotokolls in den 30 Milchviehbetrieben

Welfare-Kriterien	Indikatoren	Ja	Nein
Bewegungsfreiheit	Zugang zu einer Freilaufläche	56,7 %	43,3 %
	Zugang zu einer Weide	33,3 %	67,7 %
Abwesenheit von Schmerzen	Enthornung	100 %	0 %
	Einsatz von Betäubungsmitteln	76,7 %	23,3 %
	Einsatz von Schmerzmitteln	23,3 %	76,7 %
	Kupieren der Schwänze	0 %	100 %

Tabellen 11 bis 14 geben einen Überblick über Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) der einzelnen Indikatoren und aggregierten Messgrößen jeweils für die 1. und 2. Datenerhebung. Im Mittel ergaben sich für die Mehrzahl der Indikatoren des Welfare Quality® Erhebungsprotokolls nur geringe Unterschiede. Beispielsweise wies der Indikator ‚% Tiere hochgradig lahm‘ mit 7,2% an beiden Tagen den gleichen Mittelwert auf. Die größte Abweichung lag für die Indikatoren ‚% Tiere mit verschmutztem Euter‘, ‚% Tiere ADF <50cm aber nicht berührt‘ und ‚% Tiere ADF 50-100 cm‘ vor.

Bei den Indikatoren zum Aufstehvorgang waren die Ergebnisse beider Erhebungen sehr ähnlich. Der größte Unterschied beim Mittelwert zeigte sich beim ‚% Aufstehvorgang Score 2‘. Auch für die Indizes der Liegeplatznutzung waren die Mittelwerte mit 88,8 und 87,4 beim ‚Cow Comfort Index‘ sowie mit 71,8 und 69,2 beim ‚Stall Use Index‘ auf einem ähnlichen Niveau.

Ein ähnliches Bild zeigte sich bei den aggregierten Welfare Quality[®] Messgrößen. Der größte Unterschied offenbarte sich beim Kriterium ‚Mensch-Tier-Beziehung‘. Bezüglich der Welfare-Prinzipien konnten ebenfalls geringe Unterschiede im Mittelwert festgestellt werden. Große Unterschiede lagen hingegen bei den Minimum- und Maximumwerten auf Herdenebene sowohl bei den Indikatoren als auch bei den aggregierten Messgrößen vor.

Tabelle 11: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Indikatoren des Welfare Quality[®] Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben

Welfare-Kriterien	Indikatoren	1. Beurteilung			2. Beurteilung		
		MW	SD	MIN - MAX	MW	SD	MIN - MAX
Abwesenheit von andauerndem Hunger	% sehr magere Kühe	3,9	5,4	0,0 - 18,8	4,2	4,9	0,0 - 15,6
Abwesenheit von andauerndem Durst	Anzahl Tränkebecken	1,1	1,4	0,0 - 5,0	1,1	1,4	0,0 - 5,0
	Anzahl Wassertröge	1,8	1,1	0,0 - 4,0	1,8	1,1	0,0 - 4,0
	Länge der Wassertröge	208,5	140,6	0,0 - 500,0	208,5	140,6	0,0 - 500,0
	Anzahl Wassertröge und Tränkebecken mit ausreichendem Wasserdurchfluss	2,0	1,6	0,0 - 6,0	2,0	1,6	0,0 - 6,0
	Länge der Wassertröge mit ausreichendem Wasserdurchfluss	202,8	125,0	0,0 - 500,0	211,5	136,9	0,0 - 500,0
Liegekomfort	Durchschnittliche Abliegezeit	5,0	0,9	3,8 - 8,5	5,1	0,7	3,9 - 6,3
	% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren	16,5	17,2	0,0 - 50,0	15,5	14,9	0,0 - 57,1
	% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen	4,6	5,2	0,0 - 23,1	6,0	8,3	0,0 - 30,3
	% Tiere mit verschmutzten Unterbeinen	66,9	29,7	11,4 - 100,0	68,6	31,8	11,4 - 100,0
	% Tiere mit verschmutzter Hinterhand	64,3	23,3	20,0 - 100,0	65,5	22,6	29,4 - 100,0
	% Tiere mit verschmutztem Euter	40,7	19,0	3,0 - 78,1	45,2	21,0	3,0 - 78,0
Bewegungsfreiheit	Anzahl an Tagen im Jahr mit Zugang zu einer Freilauffläche	153,0	164,8	0,0 - 365,0	153,0	164,8	0,0 - 365,0
	Anzahl an Stunden pro Tag mit Zugang zu einer Freilauffläche	10,8	11,1	0,0 - 24,0	10,8	11,1	0,0 - 24,0
	Anzahl an Tagen im Jahr mit Zugang zu einer Weide	41,7	65,6	0,0 - 180,0	41,7	65,6	0,0 - 180,0
	Anzahl an Stunden pro Tag mit Zugang zu einer Weide	2,8	4,7	0,0 - 18,0	2,8	4,7	0,0 - 18,0
Abwesenheit von Verletzungen	% Tiere nicht lahm	76,0	13,5	50,0 - 96,7	74,5	13,4	43,8 - 90,9
	% Tiere geringgradig lahm	16,8	10,8	0,0 - 38,7	18,4	11,0	4,6 - 53,1
	% Tiere hochgradig lahm	7,2	10,8	0,0 - 39,0	7,2	9,1	0,0 - 29,0
	% ohne Integumentschäden	69,1	17,9	23,0 - 100,0	65,5	16,0	17,0 - 83,3

	% Tiere mit geringgradigen Integumentschäden	22,4	12,7	0,0 - 57,0	24,5	9,1	10,0 - 46,0
	% Tiere mit hochgradigen Integumentschäden	8,9	9,5	0,0 - 43,0	10,0	10,0	0,0 - 50,0
Abwesenheit von Krankheiten	Anzahl der Hustenereignisse pro Tier und 15 Minuten	0,0	0,0	0,0 - 0,0	0,0	0,0	0,0 - 0,0
	% Tiere mit Nasenausfluss	0,8	2,5	0,0 - 12,5	0,3	1,3	0,0 - 6,7
	% Tiere mit Augenausfluss	0,0	0,0	0,0 - 0,0	0,0	0,0	0,0 - 0,0
	% Tiere mit erschwerter Atmung	0,0	0,0	0,0 - 0,0	0,0	0,0	0,0 - 0,0
	% Tiere mit Diarrhö	4,3	9,6	0,0 - 46,7	4,7	8,3	0,0 - 28,1
	% Tiere mit Vulvausfluss	0,0	0,0	0,0 - 0,0	0,0	0,0	0,0 - 0,0
	% Tiere mit SCC>400.000	12,7	7,3	2,1 - 36,4	12,7	7,3	2,1 - 36,4
	% verendeter Tiere während eines Jahres	2,8	3,3	0,0 - 13,2	2,8	3,3	0,0 - 13,2
	% Tiere mit Schwereburten	5,1	5,3	0,0 - 21,1	5,1	5,3	0,0 - 21,1
	% festliegende Kühe	5,6	3,0	0,0 - 10,0	5,6	3,0	0,0 - 10,0
Ausdruck von Sozialverhalten	Durchschnittliche Anzahl an Kopfstößen pro Tier und Stunde	1,1	0,8	0,2 - 3,4	1,1	0,8	0,1 - 4,2
	Durchschnittliche Anzahl an Verdrängungen pro Tier und Stunde	0,5	0,4	0,0 - 1,9	0,6	0,5	0,0 - 2,0
Ausdruck anderer Verhaltensweisen	Anzahl an Tagen im Jahr mit Zugang zu einer Weide	41,7	65,6	0,0 - 180,0	41,7	65,6	0,0 - 180,0
	Anzahl an Stunden pro Tag mit Zugang zu einer Weide	2,8	4,7	0,0 - 18,0	2,8	4,7	0,0 - 18,0
Mensch-Tier-Beziehung	% Tiere ADF berührt	43,1	22,9	3,0 - 80,0	40,6	24,8	0,0 - 80,0
	% Tiere ADF <50cm aber nicht berührt	45,5	21,7	17,1 - 88,0	39,9	19,7	15,6 - 81,8
	% Tiere ADF 50-100 cm	10,7	9,5	0,0 - 43,0	18,3	15,5	0,0 - 79,0
	% Tiere ADF >100cm	0,8	1,5	0,0 - 5,6	1,3	2,2	0,0 - 7,0
Positiver emotionaler Zustand	Qualitative Verhaltensbeurteilung	-1,1	2,2	-5,6 - 2,9	-0,7	2,1	-3,9 - 2,4

Tabelle 12: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung bei der 1. und 2. Beurteilung

Indikatoren	n	1. Beurteilung			2. Beurteilung		
		MW	SD	MIN - MAX	MW	SD	MIN - MAX
% Aufstehvorgang Score 1	30	57,5	25,9	15,5 - 100,0	57,4	19,7	21,4 - 100,0
% Aufstehvorgang Score 2	30	34,5	22,5	0,0 - 87,5	32,6	17,5	0,0 - 64,3
% Aufstehvorgang Score 3	30	6,6	8,2	0,0 - 25,0	7,4	7,6	0,0 - 20,0
% Aufstehvorgang Score 4	30	1,2	3,3	0,0 - 12,5	1,9	4,6	0,0 - 20,0
% Aufstehvorgang Score 5	30	0,2	1,1	0,0 - 5,9	0,6	2,3	0,0 - 10,0
Cow Comfort Index (CCI)	28	88,8	5,2	75,3 - 99,1	87,4	10,2	58,6 - 99,5
Stall Use Index (SUI)	28	71,8	9,7	44,3 - 96,1	69,2	12,2	38,2 - 86,6

Tabelle 13: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Welfare-Kriterien des Welfare Quality® Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben

		1. Beurteilung			2. Beurteilung		
Welfare-Prinzipien	Welfare-Kriterien	MW	SD	MIN - MAX	MW	SD	MIN - MAX
Gute Fütterung	Abwesenheit von andauerndem Hunger	79,7	23,7	31,5 - 100,0	76,8	24,4	35,1 - 100,0
	Abwesenheit von andauerndem Durst	77,5	38,1	3,0 - 100,0	82,9	33,4	3,0 - 100,0
Gute Haltung	Liegekomfort	39,0	15,4	8,6 - 70,9	41,4	11,9	11,8 - 63,5
	Bewegungsfreiheit	100,0	0,0	100,0 - 100,0	100,0	0,0	100,0 - 100,0
Gute Gesundheit	Abwesenheit von Verletzungen	57,9	18,6	20,5 - 95,0	55,2	16,9	18,5 - 83,0
	Abwesenheit von Krankheiten	49,9	21,1	22,3 - 100,0	48,6	16,9	24,7 - 86,0
	Abwesenheit von Schmerzen	50,4	15,9	28,0 - 75,0	50,4	15,9	28,0 - 75,0
Artgemäßes Verhalten	Ausdruck von Sozialverhalten	56,8	21,8	12,9 - 95,3	57,0	22,4	4,1 - 98,7
	Ausdruck anderer Verhaltensweisen	15,7	26,2	0,0 - 72,6	15,7	26,2	0,0 - 72,6
	Mensch-Tier Beziehung	68,3	10,4	43,9 - 91,3	62,5	12,3	33,3 - 91,3
	Positiver emotionaler Zustand	42,9	19,0	10,8 - 79,1	46,7	18,4	20,8 - 74,3

Tabelle 14: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Welfare-Prinzipien des Welfare Quality® Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben

		1. Beurteilung			2. Beurteilung		
Welfare-Prinzipien		MW	SD	MIN - MAX	MW	SD	MIN - MAX
Gute Fütterung		68,9	31,8	14,6 - 100,0	71,6	28,5	14,6 - 100,0
Gute Haltung		61,6	9,7	42,4 - 81,7	63,1	7,5	44,4 - 77,0
Gute Gesundheit		44,1	12,0	24,3 - 66,6	43,8	11,0	22,9 - 63,4
Artgemäßes Verhalten		31,2	11,5	18,5 - 67,9	31,4	11,7	16,1 - 61,0

Tabelle 15 zeigt, dass alle Betriebe mit ‚gehobener Standard‘ oder mit ‚akzeptabel‘ bewertet wurden. 66,7% bzw. 80% der Betriebe wurden bei der 1. bzw. 2. Beurteilung als ‚gehobener Standard‘ beurteilt.

Tabelle 15: Anteil an Betrieben in den Kategorien der Welfare Quality® Gesamtbeurteilung bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben

Gesamtbeurteilung	1. Beurteilung	2. Beurteilung
% Betriebe ‚ausgezeichnet‘	0,0	0,0
% Betriebe ‚gehobener Standard‘	66,7	80,0
% Betriebe ‚akzeptabel‘	33,3	20,0
% Betriebe ‚nicht klassifiziert‘	0,0	0,0

5.2 Kurzfristige Wiederholbarkeit über die Zeit von Einzelindikatoren und aggregierten Messgrößen des Wohlergehens

5.2.1 Alle Untersuchungsbetriebe

Tabelle 16 enthält die Korrelationen auf Herdenebene für die einzelnen Indikatoren des Welfare Quality[®] Erhebungsprotokolls für Milchkühe jeweils für die gesamte Stichprobe bzw. für die jeweils von Beobachter 1 und Beobachter 2 erhobenen Betriebe. In Anlehnung an Martin und Bateson (1993), die eine Korrelation ab 0,7 als ausreichend hoch bezeichnen, erreichten über alle Betriebe hinweg 9 von 22 Indikatoren eine Korrelation von mindestens 0,7. Getrennt nach Beobachter 1 und Beobachter 2 betrug dieser Anteil 4 von 22 bzw. 8 von 22. Nur zwei Indikatoren zeigten eine Korrelation $>0,7$ über alle drei Stichproben hinweg. Dies waren ‚% sehr magere Kühe‘ und ‚% Tiere nicht lahm‘. Vergleichbare, aber auf einem niedrigeren Niveau liegende Korrelationen bestanden nur bei wenigen Indikatoren wie beispielsweise ‚% Tiere mit verschmutztem Euter‘ oder ‚Durchschnittliche Anzahl an Kopfstößen pro Tier und Stunde‘. Teilweise lagen erhebliche Unterschiede zwischen den einzelnen Stichproben vor, wie bei der ‚Sauberkeit der Wasserstellen‘: gesamte Stichprobe 0,15, Beobachter 1 0,58 und Beobachter 2 -0,18. Die für Beobachter 1 und Beobachter 2 errechneten Korrelationen wiesen bei 8 von 22 einen Unterschied von $\leq 0,1$ auf, dagegen lagen ausgeprägte Unterschiede für Indikatoren wie ‚% Tiere ADF <50 cm aber nicht berührt‘ oder ‚% Tiere geringgradig lahm‘ vor.

Tabelle 17 enthält die Korrelationen für die gesamte Stichprobe sowie für Beobachter 1 und 2 für die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung. Die Scores 1 bis 3 der qualitativen Beurteilung des Aufstehvorganges wiesen bezüglich der gesamten Stichprobe und der von Beobachter 1 erhobenen Stichprobe bis auf eine Ausnahme eine hohe Korrelation auf. Bei Beobachter 2 konnte ein geringeres Niveau der Korrelation festgestellt werden. Bei ‚% Aufstehvorgang Score 4‘ und bei ‚% Aufstehvorgang Score 5‘ lagen niedrige Korrelationen vor, jedoch lag der Anteil für beide Scores unter zwei Prozent. Der ‚Cow Comfort Index‘ zeigte für alle drei Berechnungen nach Martin und Bateson (1993) eine mittelmäßige Korrelation. Dieses Niveau der Korrelation wurde auch beim ‚Stall Use Index‘ erreicht. Lediglich die von Beobachter 2 erhobene Stichprobe wies eine geringere Korrelation auf.

Tabelle 16: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit ausgewählter Indikatoren des Welfare Quality® Erhebungsprotokolls für Milchkühe für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2

Welfare-Kriterien	Indikatoren	Korrelation zwischen 1. und 2. Beurteilung					
		Gesamte Stichprobe		Beobachter 1		Beobachter 2	
		r_s	n	r_s	n	r_s	n
Abwesenheit von andauerndem Hunger	% sehr magere Kühe	0,91 ¹	30	0,79 ¹	15	0,89 ¹	15
Abwesenheit von andauerndem Durst	Sauberkeit der Wasserstellen	0,15 ⁴	30	0,58 ³	15	-0,18 ⁴	15
Liegekomfort	Durchschnittliche Abliegezeit	0,57 ²	30	0,43 ⁴	15	0,21 ⁴	15
	% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren	0,48 ²	30	0,67 ²	15	0,32 ⁴	15
	% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen	0,51 ²	30	0,48 ⁴	15	0,48 ⁴	15
	% Tiere mit verschmutztem Unterbein	0,89 ¹	30	0,54 ³	15	0,56 ³	15
	% Tiere mit verschmutzter Hinterhand	0,88 ¹	30	0,61 ³	15	0,91 ¹	15
	% Tiere mit verschmutztem Euter	0,66 ¹	30	0,56 ³	15	0,60 ³	15
Abwesenheit von Verletzungen	% Tiere nicht lahm	0,86 ¹	30	0,89 ¹	15	0,87 ¹	15
	% Tiere geringgradig lahm	0,72 ¹	30	0,89 ¹	15	0,27 ⁴	15
	% Tiere hochgradig lahm	0,84 ¹	30	0,65 ²	15	0,87 ¹	15
	% Tiere mit ohne Integumentschäden	0,68 ¹	30	0,53 ³	15	0,79 ¹	15
	% Tiere mit geringgradigen Integumentschäden	0,51 ²	30	0,34 ⁴	15	0,58 ³	15
	% Tiere mit hochgradigen Integumentschäden	0,65 ¹	30	0,64 ³	15	0,59 ³	15
Abwesenheit von Krankheiten	% Tiere mit Diarrhö	0,23 ⁴	30	0,12 ⁴	15	0,32 ⁴	15
Ausdruck von Sozialverhalten	Durchschnittliche Anzahl an Kopfstößen pro Tier und Stunde	0,54 ²	30	0,49 ⁴	15	0,49 ⁴	15
	Durchschnittliche Anzahl an Verdrängungen pro Tier und Stunde	0,34 ⁴	30	0,54 ³	15	0,25 ⁴	15
Mensch-Tier-Beziehung	% Tiere ADF berührt	0,87 ¹	30	0,28 ⁴	15	0,76 ²	15
	% Tiere ADF <50cm aber nicht berührt	0,80 ¹	30	0,07 ⁴	15	0,72 ²	15
	% Tiere ADF 50-100cm	0,54 ²	30	0,64 ³	15	0,72 ²	15
	% Tiere ADF >100cm	0,63 ¹	30	0,58 ³	15	-	-
Positiver emotionaler Zustand	Qualitative Verhaltenserfassung	0,89 ¹	30	0,70 ²	15	0,48 ⁴	15

n: Anzahl an Betrieben

¹: p<0,001, ²: p<0,01, ³: p<0,05, ⁴: p>0,05

Abbildung 4 enthält Streuungsdiagramme für die Indikatoren des Ruheverhaltens, die beispielhaft die unterschiedlichen Ergebnisse erklären. Beim Indikator ‚% Tiere nicht lahm‘ lagen die Korrelationen für alle drei Stichproben über dem Grenzwert von 0,7 (Diagramm links). Für beide Beobachter liegt eine ähnliche Verteilung und Übereinstimmung hinsichtlich nicht lahmer Kühe vor. Das mittlere und rechte Diagramm verdeutlichen dagegen, dass die Einschätzung des Lahmheitsgrads uneinheitlich erfolgte. Während Beobachter 1 bei vergleichbarem mittlerem Gesamtanteil an lahmen Tieren nur geringe Anteile an hochgradig lahmen Tieren erhob, war dieses Verhältnis bei Beobachter 2 fast umgekehrt. Bei ‚% Tiere hochgradig lahm‘ weisen die gesamte Stichprobe und Beobachter 1 eine hohe Korrelation auf, hingegen Beobachter 2 mit 0,27 eine deutlich niedrigere. Beim Indikator ‚% Tiere hochgradig lahm‘ sind die Korrelationen allerdings wieder höher für beide Beobachter.

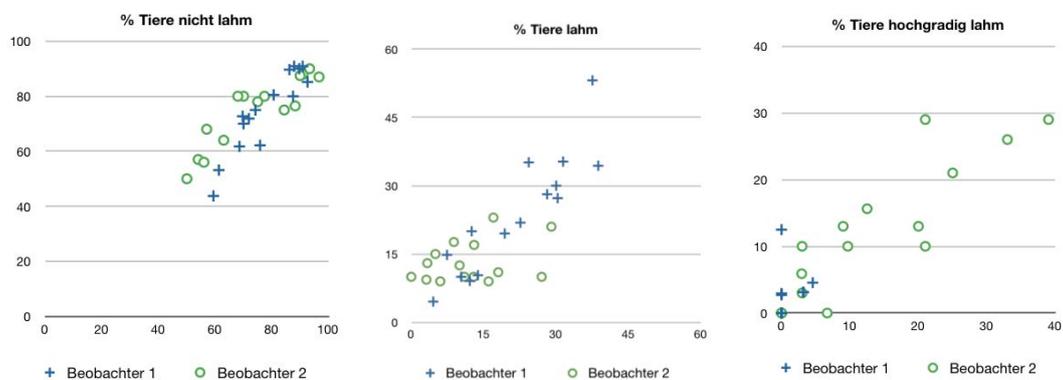


Abbildung 4: Streuungsdiagramme für die Indikatoren ‚% Tiere nicht lahm‘, ‚% Tiere geringgradig lahm‘ und ‚% Tiere hochgradig lahm‘ für die Betriebe von Beobachter 1 und 2

Tabelle 17: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit für die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2

Indikatoren	Korrelation zwischen 1. und 2. Beurteilung					
	Gesamte Stichprobe		Beobachter 1		Beobachter 2	
	r_s	n	r_s	n	r_s	n
% Aufstehvorgang Score 1	0,78 ¹	30	0,74 ²	15	0,54 ³	15
% Aufstehvorgang Score 2	0,70 ¹	30	0,71 ²	15	0,27 ⁴	15
% Aufstehvorgang Score 3	0,69 ¹	30	0,77 ¹	15	0,36 ⁴	15
% Aufstehvorgang Score 4	0,32 ⁴	30	0,10 ⁴	15	0,63 ³	15
% Aufstehvorgang Score 5	-0,05 ⁴	30	-0,07 ⁴	15	-	-
Cow Comfort Index (CCI)	0,47 ³	28	0,56 ³	14	0,56 ³	14
Stall Use Index (SUI)	0,42 ³	28	0,59 ³	14	0,25 ⁴	14

n: Anzahl an Betrieben

¹: $p < 0,001$, ²: $p < 0,01$, ³: $p < 0,05$, ⁴: $p > 0,05$

Die Welfare-Kriterien des Welfare Quality[®] Erhebungsprotokolls wiesen, wie in Tabelle 18 ersichtlich, im Vergleich zu den Einzelindikatoren höhere Korrelationen auf. Für sieben von elf Welfare-Kriterien lagen für die gesamte Stichprobe sowie bei Beobachter 1 und Beobachter 2 Korrelationen >0,7 vor. Bei den Welfare-Kriterien ‚Mensch-Tier Beziehung‘ und ‚Positiver emotionaler Zustand‘ wurde der Grenzwert nur für die gesamte Stichprobe und Beobachter 1 überschritten. Die Wiederholbarkeit über die Zeit für das Welfare-Kriterium ‚Liegekomfort‘ schwankte stark für die verschiedenen Stichproben, blieb aber immer unter dem Grenzwert von 0,7. Auch die Wiederholbarkeit für das Welfare-Kriterium ‚Ausdruck von Sozialverhalten‘ erreichte nur ein mittelmäßiges Niveau.

Wie in Tabelle 19 ersichtlich, zeigten drei von vier Welfare-Prinzipien des Welfare Quality[®] Erhebungsprotokolls für Milchkühe eine Korrelation >0,7 für die gesamte Stichprobe sowie für Beobachter 1 und Beobachter 2. Beim Welfare-Prinzip ‚Gute Haltung‘ lagen die Korrelationen deutlich unter dem Grenzwert von 0,7.

Tabelle 18: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit der Welfare-Kriterien des Welfare Quality[®] Erhebungsprotokolls für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2

Welfare-Prinzipien	Welfare-Kriterien	Korrelation zwischen 1. und 2. Beurteilung					
		Gesamte Stichprobe		Beobachter 1		Beobachter 2	
		r_s	n	r_s	n	r_s	n
Gute Fütterung	Abwesenheit von andauerndem Hunger	0,91 ¹	30	0,79 ¹	15	0,89 ¹	15
	Abwesenheit von andauerndem Durst	0,92 ¹	30	1,00 ¹	15	0,90 ¹	15
Gute Haltung	Liegekomfort	0,41 ³	30	0,66 ²	15	0,10 ⁴	15
	Bewegungsfreiheit	1,00 ¹	30	1,00 ¹	15	1,00 ¹	15
Gute Gesundheit	Abwesenheit von Verletzungen	0,93 ¹	30	0,96 ¹	15	0,90 ¹	15
	Abwesenheit von Krankheiten	0,95 ¹	30	0,96 ¹	15	0,94 ¹	15
	Abwesenheit von Schmerzen	1,00 ¹	30	1,00 ¹	15	1,00 ¹	15
Artgemäßes Verhalten	Ausdruck von Sozialverhalten	0,58 ¹	30	0,70 ²	15	0,51 ⁴	15
	Ausdruck anderer Verhaltensweisen	1,00 ¹	30	1,00 ¹	15	1,00 ¹	15
	Mensch-Tier-Beziehung	0,79 ¹	30	0,70 ²	15	0,51 ¹	15
	Positiver emotionaler Zustand	0,89 ¹	30	0,70 ²	15	0,48 ⁴	15

n: Anzahl an Betrieben

¹: p<0,001, ²: p<0,01, ³: p<0,05, ⁴: p>0,05

Tabelle 19: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit der Welfare-Prinzipien des Welfare Quality[®] Erhebungsprotokolls für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2

Welfare-Prinzipien	Korrelation zwischen 1. und 2. Beurteilung					
	Gesamte Stichprobe		Beobachter 1		Beobachter 2	
	r_s	n	r_s	n	r_s	n
Gute Fütterung	0,92 ¹	30	0,90 ¹	15	0,97 ¹	15
Gute Haltung	0,41 ³	30	0,66 ²	15	0,10 ⁴	15
Gute Gesundheit	0,93 ¹	30	0,93 ¹	15	0,92 ¹	15
Artgemäßes Verhalten	0,87 ¹	30	0,94 ¹	15	0,81 ¹	15

n: Anzahl an Betrieben

¹: $p < 0,001$, ²: $p < 0,01$, ³: $p < 0,05$, ⁴: $p > 0,05$

Bei der Gesamtbeurteilung der Betriebe anhand des Welfare Quality[®] Erhebungsprotokolls für Milchkühe wurden 26 von 30 Betrieben derselben Kategorie zugeordnet. Bei Beobachter 1 lag die Übereinstimmung bei 100 Prozent. Bei Beobachter 2 erhielten 4 von 15 Betrieben eine unterschiedliche Beurteilung.

5.2.2 Untersuchungsbetriebe mit identischer bzw. abweichender Stichprobe bei der Einzeltieruntersuchung

Wie in Tabelle 20 ersichtlich wurde in den Betrieben mit identischer Stichprobe bei der wiederholten Erhebung in Abhängigkeit vom Indikator ein sehr unterschiedliches Maß an Übereinstimmung erreicht. Bei den Integumentschäden schwankte die mittlere Übereinstimmung zwischen 85,2 und 99,2 %, bei den Indikatoren der Sauberkeit hingegen nur zwischen 95,5 und 79,2 %. Letztere Indikatoren wiesen auch die größte Spreizung der Werte auf. Bei den übrigen Indikatoren lag mit im Durchschnitt über 90 Prozent eine hohe Übereinstimmung vor.

Tabelle 20: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für ausgewählte Indikatoren des Welfare Quality® Erhebungsprotokolls in Betrieben mit identischer Stichprobe

Indikatoren	n	MW ¹	SD	MIN – MAX ¹
Sauberkeit des Unterbeins	6	70,9	13,2	50,0 – 90,6
Sauberkeit der Hinterhand	6	79,2	8,0	65,0 – 88,9
Sauberkeit des Euters	6	65,5	13,4	50,0 – 82,8
Body Condition Score	6	92,0	10,1	75,0 – 100,0
Integumentschäden ² am Tarsalgelenk	6	85,2	7,2	75,0 – 97,0
Integumentschäden ² an den Oberbeinen	6	92,1	8,4	78,8 – 100,0
Integumentschäden ² an Nacken, Schulter und Rücken	6	94,0	7,6	78,8 – 100,0
Integumentschäden ² am Karpalgelenk	6	98,4	2,6	93,9 – 100,0
Integumentschäden ² an Flanke, Seite und Euter	6	92,7	8,8	75,8 – 100,0
Andere Integumentschäden ²	6	99,2	1,9	95,5 – 100,0
Diarrhö	6	98,7	2,1	95,5 – 100,0
Lahmheit	6	93,4	7,2	81,2 – 100,0

n: Anzahl an Betrieben

¹ Angabe in Prozent (%)

² haarlose Stellen, Läsionen und Schwellungen

Tabelle 21 enthält die Korrelationen von ausgewählten Indikatoren des Welfare Quality® Erhebungsprotokolls, die aus der Einzeltieruntersuchung hervorgehen, für die gesamte Stichprobe bzw. für die Betriebe mit einer identischen und abweichenden Stichprobe. Der Vergleich zeigt, dass Betriebe mit einer identischen Stichprobe keine wesentlich höheren Korrelationen aufweisen konnten als jene Betriebe mit einer abweichenden Stichprobe. Bei 8 von 11 lag bei den Betrieben mit einer identischen Stichprobe eine niedrigere Korrelation vor. Eine höhere Korrelation bestand bei den Indikatoren ‚% sehr magere Kühe‘, ‚% Tiere geringgradig lahm‘ und bei ‚% Tiere hochgradig lahm‘.

Tabelle 21: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit ausgewählter Indikatoren des Welfare Quality® Erhebungsprotokolls für die gesamte Stichprobe sowie für Betriebe mit einer identischen bzw. abweichenden Stichprobe

Welfare-Kriterien	Indikatoren	Korrelation zwischen 1. und 2. Beurteilung					
		Gesamte Stichprobe		Identische Stichprobe		Abweichende Stichprobe	
		r_s	n	r_s	n	r_s	n
Abwesenheit von andauerndem Hunger	% sehr magere Kühe	0,91 ¹	30	0,96 ¹	6	0,88 ²	6
	% Tiere mit verschmutztem Unterbein	0,89 ¹	30	0,45 ⁴	6	0,91 ²	6
	% Tiere mit verschmutzter Hinterhand	0,88 ¹	30	0,68 ⁴	6	0,86 ³	6
Abwesenheit von Verletzungen	% Tiere mit verschmutztem Euter	0,66 ¹	30	0,50 ⁴	6	0,86 ³	6
	% Tiere nicht lahm	0,86 ¹	30	0,75 ⁴	6	0,79 ³	6
	% Tiere geringgradig lahm	0,72 ¹	30	0,79 ³	6	0,29 ⁴	6
	% Tiere hochgradig lahm	0,84 ¹	30	1,00 ¹	6	0,91 ²	6
	% Tiere ohne Integumentschäden	0,68 ¹	30	0,50 ⁴	6	0,89 ²	6
	% Tiere mit geringgradigen Integumentschäden	0,51 ²	30	0,36 ⁴	6	0,79 ³	6
Abwesenheit von Krankheiten	% Tiere mit hochgradigen Integumentschäden	0,65 ¹	30	0,29 ⁴	6	0,50 ⁴	6
	% Tiere mit Diarrhö	0,23 ⁴	30	-0,39 ⁴	6	0,18 ⁴	6

n: Anzahl an Betrieben

¹: p<0,001, ²: p<0,01, ³: p<0,05, ⁴: p>0,05

6 Diskussion

Die Berechnung der Wiederholbarkeit über die Zeit erfolgte für alle Einzelindikatoren und aggregierten Messgrößen mit Spearmans Rangkorrelationskoeffizienten (r_s). Für die Beurteilung der Wiederholbarkeit über die Zeit kann Martin und Bateson (1993) herangezogen werden. Die in Tabelle 22 ersichtliche Interpretation wird in der Diskussion als Interpretationsgrundlage verwendet. Dabei wird als Grenzwert für eine ausreichende Wiederholbarkeit eine Korrelation von 0,7 angesehen.

Tabelle 22: Interpretation des Spearman Rangkorrelationskoeffizienten (r_s) (Martin und Bateson, 1993)

Wert von r_s	Interpretation
<0,2	Fast unbedeutende Beziehung
0,2 - 0,4	Geringe Korrelation; definitive aber geringe Beziehung
0,4 - 0,7	Mittelmäßige Korrelation; beachtliche Beziehung
0,7 - 0,9	Hohe Korrelation; deutliche Beziehung
0,9 - 1,0	Sehr hohe Korrelation; sehr zuverlässige Beziehung

6.1 Indikatoren des Wohlergehens

Lediglich bei 9 von 22 Indikatoren des Welfare Quality[®] Erhebungsprotokolls lag bei der gesamten Stichprobe eine Korrelation über dem Grenzwert von 0,7 vor. Dies waren Indikatoren zu Sauberkeit, Lahmheit, Ausweichdistanz, Qualitativer Verhaltenserfassung und Body Condition Score. Über alle drei Stichproben hinweg betrachtet (gesamt bzw. getrennt für die beiden Beobachter) erreichte die Korrelation nur bei zwei Indikatoren den Grenzwert. Die Ergebnisse liegen damit in der Größenordnung der von Kirchner et al. (2014) ermittelten Wiederholbarkeit in Stiermastbetrieben über einen Zeitraum von einem Monat (7 von 18 Indikatoren) Jedoch waren dies Indikatoren zur Ressource Wasser, des Ruheverhaltens, zum Gesundheitsstatus, zur Bewegungsfreiheit, der Sauberkeit und des agonistischen Verhaltens. Lediglich Indikatoren der Sauberkeit weisen also eine ähnliche Wiederholbarkeit in beiden Untersuchungen auf. Mit einem längeren Zeitintervall fiel in der Untersuchung von Kirchner et al. (2014) das Niveau der Korrelation tendenziell ab. So ergab sich bei einem Zeitintervall von sechs Monaten nur mehr bei 2 von 18 Indikatoren eine Korrelation $>0,7$. Dies traf auf zwei ressourcenbezogene Indikatoren zum Wasser zu (Kirchner et al., 2014).

Für eine Reihe von Indikatoren erfolgte aufgrund des seltenen Auftretens in der Stichprobe keine Korrelationsberechnung. Die Indikatoren „Anzahl der

Hustenereignisse pro Tier und 15 Minuten', ,% Tiere mit Nasenausfluss', ,% Tiere mit Augenausfluss' und ,% Tiere mit erschwerter Atmung' waren bei der Erhebung nahezu nicht vorzufinden. Auch bei Kirchner et al. (2014) konnte aufgrund des seltenen Auftretens in der Stichprobe bei gewissen Indikatoren keine Korrelationsberechnung erfolgen. Dies waren jedoch nicht nur Indikatoren zum Gesundheitsstatus, sondern auch Indikatoren zur Ressource Wasser, der Bewegungsfreiheit, der Integumentschäden und der Abwesenheit von Schmerzen.

Die Indikatoren ,% sehr magere Kühe' sowie ,% Tiere nicht lahm' wiesen sowohl bei der gesamten Stichprobe als auch bei Beobachter 1 und 2 eine hohe Korrelation auf. Das hohe Niveau der Korrelation der vorliegenden Ergebnisse war auch bei den Betrieben mit einer identischen und einer abweichenden Stichprobe zu sehen. Eine gute Intra-Observer-Reliabilität und repräsentative Stichproben vorausgesetzt, ist sowohl für die Körperkondition als auch die Gangbeurteilung nicht davon auszugehen, dass sich eine deutliche Änderung im Anteil der betroffenen Tiere innerhalb von wenigen Tagen ergibt. Bei identischer Stichprobe hat die Wahl der Körperseite ebenfalls keinen Einfluss auf die beiden Indikatoren.

Die Wiederholbarkeit über die Zeit war für die Mehrheit der untersuchten Indikatoren jedoch nicht zufriedenstellend. Das Zeitintervall bei der Erhebung war mit vier Tagen kurz. Es gab kaum Veränderungen in der Zusammensetzung der Herde. Die Beobachter hatten nicht den Eindruck, dass es zu Veränderungen im Management gekommen war. Auch der Tageszeitpunkt der Erhebung gestaltete sich gleich. Für die Unterschiede in der Wiederholbarkeit über die Zeit werden daher im Folgenden mögliche Ursachen diskutiert.

Bei den vorliegenden Ergebnissen wurden die managementbasierten Indikatoren bei der Datenerhebung als Fragen zum Management gestellt. Bei einem Zeitintervall von vier Tagen konnte es zu keinen Veränderungen kommen. Demzufolge wurde keine Korrelationsberechnung durchgeführt. Bei der Ressource Wasser zeigten sich bei 5 von 6 Indikatoren keine Veränderungen. Die Beurteilung der ,Sauberkeit der Tränken' erfolgte in ,sauber', ,teilweise schmutzig' und ,schmutzig'. Mit einer Korrelation für die gesamte Stichprobe von 0,15 ist dieser Indikator praktisch nicht wiederholbar. Für die Beurteilung der Tränken als ,sauber' müssen sowohl Tränke als auch Wasser zum Zeitpunkt der Erhebung sauber sein. ,Teilweise verschmutzte' Tränke bedeutet, dass das Wasser sauber sein muss jedoch die Tränke verschmutzt sein darf. Bei einer ,verschmutzten' Tränke sind sowohl Tränke als auch Wasser verschmutzt (Welfare Quality[®], 2009). Mit dieser Beurteilung war es bei den vorliegenden Ergebnissen nicht möglich, bei den drei verschiedenen Stichproben eine hohe Korrelation zu erreichen.

Um das Ziel einer hohen Korrelation für die Wiederholbarkeit über die Zeit erreichen, wäre es ausreichend, eine Beurteilung in ,sauber' und ,verschmutzt' vorzunehmen. Das entscheidende Kriterium ist die Sauberkeit des Wassers und dies wird sowohl

bei der Beurteilung als ‚sauber‘ und ‚teilweise verschmutzt‘ erfüllt. Außerdem kann man davon ausgehen, dass sich der Betriebsführende bei einer Beurteilung als ‚sauber‘ oder ‚teilweise verschmutzt‘ mit der Sauberkeit der Wasserstellen auseinandergesetzt hat. Ein weiterer Grund, der für diese Beurteilung spricht, ist, dass bei der Berechnung des Welfare-Kriteriums ‚Abwesenheit von andauerndem Durst‘ nur mehr der Frage nachgegangen wird: Sind die Wasserstellen ‚sauber‘? Die Antwort auf diese Frage lautet entweder ‚Ja‘ oder ‚Nein‘ (Welfare Quality®, 2009). Wäre bei den vorliegenden Ergebnissen die vorgeschlagene Beurteilung erfolgt, hätten alle 30 Betriebe an beiden Tagen die gleiche Beurteilung erhalten. Demzufolge wäre auch die Wiederholbarkeit über die Zeit mit $r_S=1,00$ sehr hoch gewesen.

Die Ursache für wenig repräsentative Ergebnisse hinsichtlich der Wiederholbarkeit über die Zeit bei den Indikatoren zum Ruheverhalten liegt vermutlich in der zu geringen Stichprobengröße. Bei allen drei Indikatoren offenbarte sich für die gesamte Stichprobe eine Korrelation auf einem mittelmäßigen Niveau. Der Korrelationskoeffizient für den Indikator ‚Durchschnittliche Abliegezeit‘ war bei Beobachter 1 mit 0,43 höher als jener von Beobachter 2 mit 0,21, jedoch im Vergleich zur gesamten Stichprobe mit 0,57 niedriger.

Kirchner et al. (2014) erreichten bei Maststieren mit Erhebungen im Abstand von einem Monat mit 0,70 eine höhere Korrelation. Auch Plesch et al. (2010) ermittelten bei Milchkühen ein bedeutend höheres Niveau der Korrelation. Mit einer Korrelation von 0,78 für Milchkühe zeigte sich eine zufriedenstellende Wiederholbarkeit über die Zeit mit einem Zeitintervall von 60 und 180 Tagen nach der ersten Erhebung. Auch für die Indikatoren ‚% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren‘ mit 0,95 und ‚% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen‘ mit 0,87 lag ein hohes Niveau der Wiederholbarkeit über die Zeit vor. Im Vergleich dazu lagen die vorliegenden Ergebnisse auf einem bedeutend niedrigeren Niveau.

Das Welfare Quality® Erhebungsprotokoll für Milchkühe erachtet die Beurteilung von zumindest sechs Abliegevorgängen als ausreichend (Welfare Quality®, 2009). Dies basiert auf Berechnungen von Plesch et al. (2010), die ausgehend von größeren Stichproben mit sechs Abliegevorgängen eine repräsentative Abbildung des Gesamtgeschehens ermittelten. Möglicherweise war die Variabilität innerhalb der von Plesch et al. (2010) untersuchten Betriebe geringer als in denen der vorliegenden Untersuchung. Tatsächlich beurteilte Beobachter 1 an beiden Tagen der Erhebung im Durchschnitt mehr Abliegevorgänge (MW 10,3) als Beobachter 2 (MW 8,1 bzw. 8,2) und erzielte auch höhere Wiederholbarkeiten. Bei nur sechs beurteilten Kühen spiegelt sich das Ergebnis einer einzelnen ‚positiven‘ Kuh bei den Indikatoren ‚% Tiere, die beim Abliegen mit der Haltungseinrichtung kollidieren‘ und ‚% Tiere, die teilweise oder komplett außerhalb der Liegefläche liegen‘ mit 16,7 % im Ergebnis wieder. Auch bei der ‚durchschnittlichen Abliegezeit‘ kann eine einzelne Kuh zu einer deutlichen Verzerrung des Betriebsergebnisses führen. Es sollte daher das Ziel sein,

den Einfluss der einzelnen Kuh auf das Betriebsergebnis zu minimieren. Es wäre deshalb notwendig, die Anzahl an zu beobachtenden Abliegevorgängen pro Betrieb zu erhöhen, um besser das tatsächliche Geschehen in den Betrieben abzubilden.

Beim Indikator ‚% Tiere mit verschmutztem Euter‘ ergab sich im Vergleich zu den beiden anderen Indikatoren der Tierverschmutzung eine niedrigere Korrelation (z.B. $r_s=0,66$ für die gesamte Stichprobe). Auch bei Winckler et al. (2007) zeigte sich bei der Sauberkeit des Euters im Vergleich zur Sauberkeit der Hinterbeine eine geringere Wiederholbarkeit über die Zeit ($r_s=0,10-0,61$). Im Welfare Quality® Erhebungsprotokoll für Milchkühe wird eine Beurteilung des Euters als ‚verschmutzt‘ folgendermaßen beschrieben: deutliche Verschmutzung (Schichten von Schmutz) der Euterhaut oder jeglicher Schmutz auf den Zitzen (Welfare Quality®, 2009). Geringe Verschmutzungen an den Zitzen führen daher bereits zu einer Beurteilung als ‚verschmutzt‘. Verschmutzungen an den Zitzen erhöhen das Risiko von Mastitis (Winckler et al., 2007). Dieser Zustand kann schnell auftreten und so können große Unterschiede zwischen den zwei Erhebungsterminen entstehen. Es wäre deshalb notwendig, eine eindeutigere Definition für die Verschmutzung der Zitzen vorzunehmen.

Eine weitere mögliche Ursache für die geringe Wiederholbarkeit über die Zeit könnte ein zu geringes Auftreten (Prävalenz) in der Stichprobe sein. Dies betrifft Indikatoren zum Gesundheitsstatus, der Ausweichdistanz und der qualitativen Beurteilung des Aufstehvorganges. Gepaart mit relativ geringen Stichprobengrößen können einzelne ‚positive‘ Tiere die Prävalenz relativ stark ansteigen lassen.

Ähnlich wie bei Kirchner et al. (2014) für Maststiere ließ sich für den Indikator ‚% Tiere mit Diarrhö‘ nur eine geringe Korrelation mit 0,23 für die gesamte Stichprobe feststellen. Trotz einer hohen Übereinstimmung von 98,7 % bei den sechs Betrieben mit einer identischen Stichprobe war auch bei diesen Betrieben keine hohe Korrelation möglich. An den beiden Tagen der Erhebung wurden bei 4,3 bzw. 4,7 % der Kühe Anzeichen von Diarrhö festgestellt; dies entspricht bei Stichprobengrößen von 30,9 bzw. 31,0 Milchkühen 1,3 bzw. 1,4 Tieren mit Anzeichen von Diarrhö. Auf Einzelbetriebsebene zeigten sich jedoch Unterschiede von bis zu 26,7 Prozent. Mögliche Ursachen für eine geringe Wiederholbarkeit über die Zeit ergeben sich daraus, dass einerseits das Auftreten in der Stichprobe gering ist und andererseits die Stichprobenauswahl einen beträchtlichen Einfluss auf das Ergebnis hat.

Bei den Indikatoren zur Ausweichdistanz war die Wiederholbarkeit in den Klassen mit hoher Ausweichdistanz, die gleichzeitig die geringsten Tieranteile auswiesen, unzureichend ‚% Tiere ADF 50-100cm‘: 10,7 bzw. 18,3 % der Tiere, $r_s=0,54$; ‚% Tiere ADF >100cm‘: 0,8 bzw. 1,3 % der Tiere, $r_s=0,63$. Jedoch offenbarte sich ein kontantes Niveau für alle drei Stichproben. Auch bei Kirchner et al. (2014) wurden nur geringe Anteile der Tiere in die beiden Klassen eingestuft und die Wiederholbarkeit war ebenfalls gering bis mittelmäßig. Ein Vergleich mit Winckler et

al. (2007) ist nicht zulässig, da bei der Korrelationsberechnung nur jene Kühe berücksichtigt wurden, die eine Berührung zuließen.

Ähnliche Zusammenhänge liegen vermutlich auch für die qualitative Beurteilung des Aufstehvorganges vor. So wurden mit ‚Aufstehvorgang Score 4‘ lediglich 1,2 bzw. 1,9 % und mit ‚Aufstehvorgang Score 5‘ 0,2 bzw. 0,6 % der getesteten Kühe eingestuft. Im Vergleich zu den drei anderen Scores der qualitativen Beurteilung des Aufstehvorganges zeigte sich bei Score 4 und 5 eine niedrige Korrelation. Ähnlich wie bei den Indikatoren zum Ruheverhalten wurden pro Betrieb 11,6 bzw. 11,5 Milchkühe an den beiden Tagen der Erhebung beurteilt. Folglich spiegelte sich die Beurteilung der einzelnen Kuh mit ca. 10 % im Ergebnis wider. Bei der qualitativen Beurteilung des Aufstehvorganges sind somit das niedrige Auftreten in der Stichprobe, der Stichprobenumfang und die Stichprobenauswahl mögliche Ursachen für die niedrige Wiederholbarkeit über die Zeit bei Score 4 und Score 5.

Bei gewissen Indikatoren der Lahmheit, der Ausweichdistanz, der qualitativen Beurteilung des Aufstehvorganges und der qualitativen Verhaltenserfassung lag die Korrelation für die gesamte Stichprobe und für jeweils einen Beobachter auf einem hohen Niveau, jedoch zeigte sich jeweils bei einem Beobachter eine deutlich geringere Korrelation. Eine mögliche Ursache könnte eine geringe Intra-Observer-Reliabilität des jeweiligen Beobachters sein. Die Intra-Observer-Reliabilität bezieht sich auf die Übereinstimmung von aufeinanderfolgenden Messungen des gleichen Beobachters (Scott et al., 2011).

Bei den drei Indikatoren zur Lahmheit lag die Korrelation für die gesamte Stichprobe über dem Grenzwert von 0,7. Ähnliche Erhebungen zur Wiederholbarkeit über die Zeit konnten das Niveau der vorliegenden Ergebnisse nur teilweise bestätigen, wobei der Abstand der Erhebungen länger war. Bei der von Winckler et al. (2007) ermittelten Wiederholbarkeit über die Zeit mit einem Zeitabstand von zwei Monaten und insgesamt fünf Erhebungen erreichten zumindest zwei Berechnungen den Grenzwert von 0,7. Die Schwankungen in der Korrelation zwischen zwei aufeinanderfolgenden Erhebungen lagen trotzdem zwischen 0,48 und 0,78. De Rosa et al. (2003) untersuchte die Wiederholbarkeit über die Zeit bei einem dreiwöchigen Intervall drei Mal und erzielte eine Korrelation nach Kendall von 0,43 bis 0,66. In der vorliegenden Untersuchung war die niedrige Korrelation bei Beobachter 2 für den Indikator ‚% Tiere geringgradig lahm‘ und bei Beobachter 1 mit einer hochgradigen Lahmheit auffallend. Abbildung 4 zeigt, dass beide Beobachter weitgehend ähnliche Anteile an nicht lahmen Tieren ermittelten, sich aber deutlich in den Anteilen ‚geringgradig lahmer‘ bzw. ‚hochgradig lahmer‘ Tiere unterscheiden. Dies legt unter der Annahme, dass sich die beiden Betriebsgruppen nicht grundlegend unterscheiden, nahe, dass hier eine geringe Beobachterübereinstimmung bei der Differenzierung der Lahmheitsgrade bestand. Gleichzeitig streuen als Hinweis für eine geringe Intra-Observer-Reliabilität die Werte für die Anteile geringgradig lahmer Tiere stärker bei Beobachter 2. Eine mögliche Ursache für die niedrige Korrelation

bei Beobachter 2 für den Indikator ‚% Tiere geringgradig lahm‘ und bei Beobachter 1 für den Indikator ‚% Tiere hochgradig lahm‘ könnte auch das geringe Auftreten in der Stichprobe sein. Eine Erhebung zur Intra-Observer-Reliabilität von Leach et al. (2009c) stellte eine Übereinstimmung von 70 Prozent bei einem zeitlichen Abstand der Erhebungen von 12 Stunden fest. Aufgrund dieser Studie und der prozentuellen Unterschiede bei der Beurteilung der geringgradigen und hochgradigen Lahmheit könnte die geringe Intra-Observer-Reliabilität eine mögliche Ursache sein.

Bei der Ausweichdistanz wurde überwiegend eine Einstufung in die Klassen ‚% Tiere ADF berührt‘ und ‚% Tiere ADF <50cm aber nicht berührt‘ mit in Summe über 80 Prozent an den beiden Tagen der Erhebung vorgenommen. Mit einer Korrelation von 0,80 bzw. 0,87 für die gesamte Stichprobe ergab sich ein ähnliches Niveau der Korrelation wie bei Winckler et al. (2007). Bei der von De Rosa et al. (2003) und Kirchner et al. (2014) ermittelten Wiederholbarkeit über die Zeit offenbarte sich ein geringeres Niveau. Die ermittelte Korrelation von Beobachter 1 lag im Vergleich zu den zwei anderen Stichproben bei beiden Indikatoren auf einem deutlich niedrigeren Niveau. Eine mögliche Ursache für die geringe Korrelation bei Beobachter 1 könnte eine niedrige Intra-Observer-Reliabilität sein.

Auch bei der qualitativen Beurteilung des Aufstehvorganges gibt es Hinweise auf eine niedrige Intra-Observer-Reliabilität. Die Beurteilung mit Score 1, 2 und 3 war mit 98,6 bzw. 97,4 % in den Betrieben vorherrschend. Die Berechnungen für die gesamte Stichprobe und für Beobachter 1 wiesen eine hohe Korrelation auf, jedoch für Beobachter 2 nicht.

Bei der qualitativen Verhaltenserfassung zeigte sich sowohl bei der gesamten Stichprobe wie auch bei Beobachter 1 eine hohe Korrelation. Bei den Erhebungen von Kirchner et al. (2014) lag die Wiederholbarkeit über die Zeit bei einem zeitlichen Abstand von einem Monat mit 0,68 nahezu über dem Grenzwert von 0,7. Bei einem längeren Zeitintervall von sechs Monaten fiel die Korrelation auf 0,53 ab. Auch bei diesem Indikator könnte eine niedrige Intra-Observer-Reliabilität die Ursache für die niedrige Korrelation sein.

Bei der Sauberkeit verschiedener Körperregionen erzielten De Rosa et al. (2003) eine mittelmäßige Wiederholbarkeit ($r_S=0,51-0,71$). Bei dieser Beurteilung wurde bei der Berechnung der Wiederholbarkeit über die Zeit jedoch nicht nach Körperregionen unterschieden. Bei Winckler et al. (2007) zeigten sich bei vier Berechnungen zur Wiederholbarkeit über die Zeit in einem Abstand von zwei Monaten bei der Sauberkeit der Hinterbeine eine mittelmäßige bis hohe Korrelation ($r_S=0,60-0,83$). Bei den vorliegenden Ergebnissen lag für die Indikatoren ‚% Tiere mit verschmutztem Unterbein‘ und ‚% Tiere mit verschmutzter Hinterhand‘ bei der gesamten Stichprobe eine hohe Korrelation mit 0,89 bzw. 0,88 vor. Jedoch fiel die Wiederholbarkeit bei Beobachter 1 und 2 stark ab. Betrachtet man die sechs Betriebe mit einer identischen Stichprobe, so lag die Übereinstimmung der Sauberkeit des Unterbeins bei 70,9 % und der Sauberkeit der Hinterhand bei 79,2 %. Dies ist im Vergleich zu den anderen Indikatoren der Einzeltieruntersuchung deutlich geringer. Es muss

jedoch bedacht werden, dass nicht darauf geachtet wurde, zwei Mal dieselbe Körperseite der einzelnen Kuh zu bewerten. Auch bei der Sauberkeit des Unterbeins und der Hinterhand können eine geringe Intra-Observer-Reliabilität, aber auch die Bewertung der unterschiedlichen Körperseiten eine mögliche Ursache für die niedrige Korrelation sein.

Bei den Integumentschäden zeigte sich bei den drei Indikatoren für die gesamte Stichprobe eine Korrelation von 0,51 bis 0,68. Auch bei Winckler et al. (2007) lag eine ähnliche Wiederholbarkeit über die Zeit bei der Erhebung von Abschürfungen und Wunden an den Tarsalgelenken zwischen 0,48 und 0,78 vor. Bei den Carpalgelenken hingegen war die Korrelation mit 0,05 bis 0,37 geringer. Aufgrund der hohen Übereinstimmung der einzelnen Körperregionen bei den Betrieben mit einer identischen Stichprobe und der Annahme, dass sich innerhalb weniger Tage kaum Tiere verletzen oder eine Liegestelle entwickeln, war mit einer höheren Korrelation gerechnet worden. Jedoch war das Niveau der Korrelation bei den Betrieben mit einer identischen Stichprobe niedriger als bei jenen mit einer abweichenden Stichprobe. Alle sechs Betriebe mit einer identischen Stichprobe wurden von Beobachter 1 beurteilt. Auch die Stichprobe der 15 Betriebe von Beobachter 1 wies eine niedrigere Korrelation auf als jene von Beobachter 2. Bei einer Erhebung zur Inter-Observer-Reliabilität von Danuser und Regula (2001) lag die Übereinstimmung von verschiedenen Körperregionen und einer Beurteilung durch zwei Beobachter zwischen 75 und 79 Prozent. Demzufolge war jede vierte bis fünfte Beurteilung unterschiedlich. Bei den Indikatoren zu den Integumentschäden könnte ähnlich wie bei der Sauberkeit verschiedener Körperregionen eine geringe Intra-Observer-Reliabilität und die Bewertung der unterschiedlichen Körperseiten mögliche Ursachen für eine niedrige Korrelation sein.

Beide Indikatoren des agonistischen Verhaltens wiesen eine geringe bis mittelmäßige Korrelation auf. Für die Wiederholbarkeit des agonistischen Verhaltens über die Zeit gibt es einige Studien. Bei den von Laister et al. (2009) durchgeführten Erhebungen in einem Abstand von 60 und 180 Tagen zeigte sich sowohl bei Milchkühen und Masttieren und auch im Laufstall und in Anbindehaltung eine große Streuung der Korrelationskoeffizienten ($r_S = -0,06-0,85$). Ein ähnliches Bild ergab sich bei Winckler et al. (2007). Bei einem zeitlichen Abstand von zwei Monaten lag die Korrelation bei vier Berechnungen zwischen -0,48 und 0,52. Bei Kirchner et al. (2014) zeigte sich für das kurze Zeitintervall eine Korrelation von 0,74 und für das längere Intervall eine Korrelation von 0,48. Es ist bekannt, dass das Sozialverhalten von vielen Faktoren wie der Besatzdichte, der Eingliederung von Tieren in die Herde, der Anwesenheit von brünstigen Tieren und einer von-Tag-zu-Tag Variation des Verhaltens beeinflusst wird (Winckler et al., 2002). Ein mögliche Ursache für die geringe Wiederholbarkeit über die Zeit könnten relativ geringe Änderungen in der Umwelt sein.

Erhebungen zum ‚Cow Comfort Index‘ und ‚Stall Use Index‘ von Ito et al. (2009) stellten keinen Zusammenhang zwischen zwei Messungen mit einem zeitlichen Abstand von fünf Tagen fest. Bei dieser Berechnung wurde jedoch die Erhebung zu verschiedenen Tageszeiten durchgeführt (vor dem morgendlichen bzw. vor dem abendlichen Melken). Bei den vorliegenden Ergebnissen lag für den ‚Cow Comfort Index‘ und den ‚Stall Use Index‘ für die gesamte Stichprobe eine mittelmäßige Korrelation vor. Diese bestätigte sich auch bei Beobachter 1 und Beobachter 2 bis auf eine Ausnahme. Hauschild (2012) ermittelte beide Indikatoren in 17 Milchviehbetrieben in Österreich. Der Mittelwert lag für den ‚Cow Comfort Index‘ bei 88 % und für den ‚Stall Use Index‘ bei 69 %. Die vorliegenden Ergebnisse erreichten an den beiden Tagen der Erhebung einen ähnlichen Mittelwert: ‚Cow Comfort Index‘ 88,8 bzw. 87,4 % und ‚Stall Use Index‘ 71,8 bzw. 69,2 %. Auf Ebene der Einzelbetriebe offenbarten sich jedoch Unterschiede zwischen Tag 1 und 2 für den ‚Cow Comfort Index‘ von bis zu 23,6 % und für den ‚Stall Use Index‘ von bis zu 31,7 %. Relativ geringe Veränderungen in der Umwelt und eine nicht ausreichend oft wiederholte Erhebung könnten mögliche Ursachen für die konstant geringe Wiederholbarkeit über die Zeit sein.

6.2 Aggregierte Messgrößen des Wohlergehens

a) Welfare-Kriterien des Welfare Quality® Erhebungsprotokolls

9 von 11 Welfare-Kriterien wiesen eine hohe (‚Mensch-Tier Beziehung‘, ‚Positiv emotionaler Zustand‘) bis sehr hohe Korrelation (‚Abwesenheit von andauerndem Hunger‘, ‚Abwesenheit von andauerndem Durst‘, ‚Bewegungsfreiheit‘, ‚Abwesenheit von Verletzungen‘, ‚Abwesenheit von Krankheiten‘, ‚Abwesenheit von Schmerzen‘, ‚Ausdruck anderer Verhaltensweisen‘) für die gesamte Stichprobe auf. Diese hohe Wiederholbarkeit zeigte sich nicht nur bei der gesamten Stichprobe, sondern auch bei jener von Beobachter 1, Beobachter 2 und bei den Betrieben mit einer identischen und abweichenden Stichprobe. Auf einem niedrigeren Niveau befand sich hingegen die Korrelation für die Welfare-Kriterien ‚Liegekomfort‘ mit 0,41 und ‚Ausdruck von Sozialverhalten‘ mit 0,58 für die gesamte Stichprobe.

Im Vergleich zu Kirchner et al. (2014), die die Wiederholbarkeit des Welfare Quality® Erhebungsprotokolls bei Masttieren über ein Intervall von einem Monat ermittelten, lag damit für mehr Welfare-Kriterien eine ausreichende Korrelation vor (Kirchner et al.: 5 von 10). Im Vergleich zu den vorliegenden Ergebnissen war auch das Niveau der Korrelation niedriger; eine Korrelation von über 0,9 konnte nur für ein Welfare-Kriterium erreicht werden.

Die Wiederholbarkeit über die Zeit war bei den aggregierten Messgrößen im Vergleich zu den Einzelindikatoren höher. Dieser Effekt hat mehrere potentielle Ursachen. Die Kombination von zwei oder mehreren Indikatoren kann Schwankungen ausgleichen. Die Scores für einige Welfare-Kriterien wie

beispielsweise ‚Abwesenheit von Krankheiten‘ oder ‚Liegekomfort‘ werden mit Schwellenwerten berechnet. Befindet sich die Prävalenz des einzelnen Indikators unter der Schwelle, kann die Korrelation des einzelnen Indikators niedrig sein dies muss jedoch keine Auswirkungen auf die Korrelation der Welfare-Kriterien haben. Auch Entscheidungsbäume werden für die Berechnung von Welfare-Kriterien-Scores verwendet, wie beispielsweise ‚Abwesenheit von andauerndem Durst‘. Auf der Ebene der Einzelindikatoren wird bei der ‚Sauberkeit der Wasserstellen‘ zwischen ‚sauber‘, ‚teilweise verschmutzt‘ und ‚verschmutzt‘ unterschieden. Beim Entscheidungsbaum gibt es jedoch nur noch eine Unterscheidung in ‚sauber‘ und ‚verschmutzt‘ (Kirchner et al., 2014; Welfare Quality[®], 2009). Die Unterscheidung auf Ebene der Einzelindikatoren führte bei den vorliegenden Ergebnissen zu einer fast unbedeutenden Beziehung. Nachdem beim Welfare-Kriterium ‚Abwesenheit von andauerndem Durst‘ ‚teilweise verschmutzte‘ Tränken als ‚sauber‘ gelten und es praktisch keine ‚verschmutzten‘ Tränken gab, war die Korrelation beim Welfare-Kriterium auf einem deutlich höheren Niveau vorzufinden.

Das Welfare-Kriterium ‚Ausdruck von Sozialverhalten‘ bildet sich aus zwei Indikatoren des agonistischen Verhaltens. Beide Indikatoren ‚Durchschnittliche Anzahl an Kopfstößen pro Tier und Stunde‘ und ‚Durchschnittliche Anzahl an Verdrängungen pro Tier und Stunde‘ konnten nur eine Korrelation von 0,54 bzw. 0,34 aufweisen. Daher war die niedrige Korrelation des Welfare-Kriteriums nachvollziehbar.

b) Welfare-Prinzipien des Welfare Quality[®] Assessment Protocols

3 von 4 Welfare-Prinzipien zeigten eine hohe (‚Artgemäßes Verhalten‘) bis sehr hohe Korrelation (‚Gute Fütterung‘, ‚Gute Gesundheit‘). Dieses Niveau der Korrelation war auch jeweils bei Beobachter 1 und 2 zu sehen. Die Wiederholbarkeit des Welfare-Prinzips ‚Gute Haltung‘ war mit 0,41 für die gesamte Stichprobe dagegen deutlich geringer. Auch bei Kirchner et al. (2014) lag bei Masttieren für ein Intervall von einem Monat für drei Welfare-Prinzipien eine hohe Korrelation vor; allerdings war bei Kirchner et al. (2014) auch das Prinzip ‚Gute Haltung‘ ausreichend wiederholbar und das Prinzip ‚Artgemäßes Verhalten‘ unterschritt nur knapp den Grenzwert.

c) Gesamtbeurteilung des Welfare Quality[®] Assessment Protocols

Mit einer Übereinstimmung von 86,7 % erzielten 26 von 30 Betrieben bei der Gesamtbeurteilung eine gleiche Bewertung an beiden Tagen der Erhebung. Im Vergleich dazu wurde von Kirchner et al. (2014) eine ähnlich hohe Übereinstimmung erzielt mit 79 Prozent in einem Intervall von einem Monat und mit 75 Prozent in einem Intervall von sechs Monaten.

7 Fazit

Bei 9 von 22 ausgewählten Indikatoren des Welfare Quality® Erhebungsprotokolls lag bei der gesamten Stichprobe eine Korrelation über dem Grenzwert von 0,7 vor, jedoch bei nur zwei Indikatoren sowohl bei der gesamten Stichprobe als auch bei Beobachter 1 und 2. Für die Mehrheit der untersuchten Einzelindikatoren war die Wiederholbarkeit über die Zeit nicht zufriedenstellend, obwohl das Zeitintervall mit vier Tagen kurz war, es kaum zu Änderungen in der Zusammensetzung der Herden gekommen war und der Tageszeitpunkt der Erhebungen unverändert blieb. Für eine unzureichende Wiederholbarkeit gibt es verschiedene mögliche Ursachen. Dazu zählen eine geringe Stichprobengröße (Indikatoren zum Ruheverhalten), geringe Prävalenzen in einzelnen Kategorien (Gesundheitsstatus, Lahmheit, Ausweichdistanz, qualitative Beurteilung des Aufstehvorganges) und eine geringe Intra-Observer-Reliabilität (Lahmheit, Ausweichdistanz, qualitative Beurteilung des Aufstehvorganges, qualitative Verhaltensbeurteilung, Sauberkeit, Integumentschäden). Für die Indikatoren des agonistischen Verhaltens und für die Indizes der Liegeplatznutzung waren geringe Veränderungen der Umwelt möglicherweise ausschlaggebend für eine niedrige Korrelation.

Bei den aggregierten Messgrößen des Wohlergehens wurde bei 9 von 11 Welfare-Kriterien und bei 3 von 4 Welfare Prinzipien eine hohe Wiederholbarkeit über die Zeit festgestellt. Potentielle Gründe für die im Vergleich zu den Einzelindikatoren höheren Korrelationen sind die Kombination von zwei oder mehreren Indikatoren, die Schwankungen ausgleichen können, sowie die Berechnung mit Schwellenwerten oder Entscheidungsbäumen. Die Ursache für die niedrige Korrelation beim Welfare-Kriterium ‚Liegekomfort‘ und beim Welfare-Prinzip ‚Gute Haltung‘ waren die Indikatoren des Ruheverhaltens und bei ‚Ausdruck von Sozialverhalten‘ die Indikatoren des agonistischen Verhaltens. Die Tatsache, dass innerhalb von wenigen Tagen aber bei mehr als 10 % der Betriebe eine unterschiedliche Gesamtbeurteilung erfolgte, lässt jedoch Zweifel an der Gesamtzuverlässigkeit (einschließlich der Beobachterübereinstimmung und des Evaluierungssystems) aufkommen.

Um eine höhere Wiederholbarkeit über die Zeit bei den Einzelindikatoren und aggregierten Messgrößen zu erhalten, ist es vermutlich notwendig, die Definitionen einzelner Indikatoren zu überarbeiten und ein intensiveres Training mit den Beobachtern durchzuführen.

8 Zusammenfassung

Methoden der Beurteilung des Wohlergehens auf Betriebsebene müssen aussagekräftig, einfach durchführbar und reliabel (zuverlässig) sein. Bei der Reliabilität wird unterschieden zwischen der Übereinstimmung der Ergebnisse verschiedener Beobachter (Inter-Observer-Reliabilität), der Übereinstimmung bei aufeinanderfolgenden Messungen des gleichen Beobachters (Intra-Observer-Reliabilität) und der Übereinstimmung zwischen Beobachtungen von zumindest zwei verschiedenen Erhebungen desselben Beobachters (Wiederholbarkeit über die Zeit).

Das Ziel dieser Arbeit war es, die Wiederholbarkeit über die Zeit von verschiedenen Einzelindikatoren und aggregierten Messgrößen des Wohlergehens zu evaluieren. Die Datenerhebung erfolgte in 30 Milchviehbetrieben an verschiedenen Standorten in Kärnten, der Steiermark und Oberösterreich. Erfasst wurde das Welfare Quality[®] Erhebungsprotokoll für Milchkühe; zusätzlich erfolgte die qualitative Beurteilung des Aufstehvorganges und die Erhebung von Indizes der Liegeplatznutzung. Zwei Beobachter beurteilten jeweils 15 Betrieben zwei Mal in einem zeitlichen Abstand von vier Tagen. Die Berechnung der Übereinstimmung zwischen den zwei Erhebungen erfolgte mittels Korrelation nach Spearman.

Bei den Indikatoren des Welfare Quality[®] Erhebungsprotokolls lag bei 9 von 22 Einzelindikatoren eine Korrelation über dem Grenzwert von 0,7 vor. Andererseits zeigte sich bei einer Reihe von Indikatoren eine geringe Wiederholbarkeit für das gewählte Intervall von nur vier Tagen. Bei der qualitativen Beurteilung des Aufstehvorganges wiesen drei von fünf Indikatoren eine fast zufriedenstellende Korrelation ($r_s \geq 0,69$) auf. Für die Indizes der Liegeplatznutzung ‚Cow Comfort Index‘ und ‚Stall Use Index‘ lag eine Korrelation von 0,47 bzw. 0,42 vor. Als Ursachen für die geringe Wiederholbarkeit über die Zeit ließ sich Folgendes feststellen: geringe Stichprobengröße, geringes Auftreten in der Stichprobe, Reaktion auf geringfügige Veränderungen der Umwelt und geringe Intra-Observer-Reliabilität.

Bei den aggregierten Messgrößen lag die Korrelation bei 9 von 11 Welfare-Kriterien und bei 3 von 4 Welfare-Kriterien über dem Grenzwert von 0,7. Die Kombination von zwei oder mehreren Indikatoren sowie die Berechnung anhand von Schwellenwerten und Entscheidungsbäumen sind potentielle Gründe für das Vorliegen einer besseren Wiederholbarkeit im Vergleich zu den Einzelindikatoren.

Um die Wiederholbarkeit über die Zeit zu verbessern, sind möglicherweise vor allem ein intensiveres Training der Beobachter und eine Überarbeitung der Definition einzelner Indikatoren erforderlich.

9 Abstract

Methods of assessing on-farm animal welfare must be valid, feasible and reliable. Special forms of reliability are the agreement of the results of different observers (inter-observer reliability), the agreement of successive measurements of the same observer (intra-observer reliability) and the agreement between observations of at least two different assessments by the same observer (consistency over time).

The aim of the present study was to evaluate the consistency over time for several measures and aggregated scores. Data were gathered on 30 dairy farms at different locations in Carinthia, Styria and Upper Austria. The Welfare Quality[®] Assessment protocol for dairy cows, as well as the rising behaviour and selected indices of stall use were assessed. Each of the two observers assessed 15 farms twice at an interval of four days.

9 out of 22 measures of the Welfare Quality[®] Assessment protocol showed a Spearman rank correlation coefficient higher than 0,7. Some other measures revealed a low consistency over time. For the qualitative assessment of rising behaviour three out of five scores correlated almost satisfactorily ($r_s \geq 0,69$). The different indices of stall use showed only low correlations ('cow comfort index' 0,47 and 'stall use index' 0,42). Reasons for the low consistency over time could be: the small sample size, a low incidence in the sample, the response to small changes in the environment and a low intra-observer reliability.

Consistency improved with the aggregation of single parameters. At criterion level, scores for 9 out of 11 criteria were considered consistent and at principle level this was the case for three out of four principles. The combination of two or more measures, the calculation using threshold values and decision trees are potential reasons for achieving a higher correlation compared to the single measures.

In order to obtain a higher consistency over time, a more intensive observer training as well as a redefinition of selected measures might be advisable.

10 Abbildungsverzeichnis

Abbildung 1: Wege zur Verbesserung des Wohlergehens der Tiere des Welfare Quality® Projektes (Blokhus, 2005).....	11
Abbildung 2: Schritte von links nach rechts von den Indikatoren („Measures“), Welfare-Kriterien („Criteria“), Welfare-Prinzipien („Principles“) zur Gesamtbeurteilung („Overall Assessment“) (Welfare Quality®, 2009)	12
Abbildung 3: Beispielhafte Abbildung des Ablaufs der Erhebungen zur Ermittlung der kurzfristigen Wiederholbarkeit über die Zeit in den 30 Betrieben.....	28
Abbildung 4: Streudiagramme für die Indikatoren ‚% Tiere nicht lahm‘, ‚% Tiere geringgradig lahm‘ und ‚% Tiere hochgradig lahm‘ für die Betriebe von Beobachter 1 und 2.....	44

11 Tabellenverzeichnis

Tabelle 1: Indikatoren, Welfare-Kriterien und Welfare-Prinzipien des Welfare Quality® Erhebungsprotokolls für Milchkühe (Welfare Quality®, 2009).....	12
Tabelle 2: Wiederholbarkeit über die Zeit von zwei Indikatoren zur Ressource Wasser bei Maststieren (Kirchner et al., 2014).....	15
Tabelle 3: Wiederholbarkeit über die Zeit verschiedener Indikatoren zum Gesundheitsstatus bei Maststieren (Kirchner et al., 2014)	19
Tabelle 4: Spearmans Rangkorrelationskoeffizienten (r_s) für ein Zeitintervall von 60 (1_60), 120 (60_180) und 180 (1_180) Tage und Kendalls Koeffizient (W) für die Indikatoren ‚Kopfstöße ohne Verdrängungen‘ und ‚Verdrängungen‘ bei Milchkühen und Maststieren (Laister et al., 2009)	21
Tabelle 5: Wiederholbarkeit über die Zeit der Welfare-Kriterien bei Maststieren (Kirchner et al., 2014).....	26
Tabelle 6: Wiederholbarkeit über die Zeit der Welfare-Prinzipien bei Maststieren (Kirchner et al., 2014).....	26
Tabelle 7: Beobachterübereinstimmung mit dem Goldstandard für ausgewählte Indikatoren des Welfare Quality® Erhebungsprotokolls	29
Tabelle 8: Anzahl der Milchkühe für die Erhebung der Ausweichdistanz auf dem Futterplatz (Welfare Quality®, 2009).....	30
Tabelle 9: Scoring-System für die Beurteilung des Aufstehvorganges (Chaplin und Munksgaard, 2001; verändert durch Hauschild, 2012).....	32
Tabelle 10: Managementbezogene Indikatoren des Welfare Quality® Erhebungsprotokolls in den 30 Milchviehbetrieben	37
Tabelle 11: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Indikatoren des Welfare Quality® Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben	38
Tabelle 12: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung bei der 1. und 2. Beurteilung.....	39

Tabelle 13: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Welfare-Kriterien des Welfare Quality® Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben.....	40
Tabelle 14: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für die einzelnen Welfare-Prinzipien des Welfare Quality® Erhebungsprotokolls für Milchkühe bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben.....	40
Tabelle 15: Anteil an Betrieben in den Kategorien der Welfare Quality® Gesamtbeurteilung bei der 1. und 2. Beurteilung in den 30 Milchviehbetrieben	41
Tabelle 16: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit ausgewählter Indikatoren des Welfare Quality® Erhebungsprotokolls für Milchkühe für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2	43
Tabelle 17: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit für die Beurteilung des Aufstehvorganges und die Indizes der Liegeplatznutzung für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2	44
Tabelle 18: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit der Welfare-Kriterien des Welfare Quality® Erhebungsprotokolls für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2	45
Tabelle 19: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit der Welfare-Prinzipien des Welfare Quality® Erhebungsprotokolls für die gesamte Stichprobe bzw. für Beobachter 1 und Beobachter 2	46
Tabelle 20: Mittelwert (MW), Standardabweichung (SD), Minimum (MIN) und Maximum (MAX) für ausgewählte Indikatoren des Welfare Quality® Erhebungsprotokolls in Betrieben mit identischer Stichprobe.....	47
Tabelle 21: Spearman Rangkorrelationskoeffizienten (r_s) für die kurzfristige Wiederholbarkeit über die Zeit ausgewählter Indikatoren des Welfare Quality® Erhebungsprotokolls für die gesamte Stichprobe sowie für Betriebe mit einer identischen bzw. abweichenden Stichprobe	48
Tabelle 22: Interpretation des Spearman Rangkorrelationskoeffizienten (r_s) (Martin und Bateson, 1993)	49

12 Literaturverzeichnis

ALBAN, L., AGGER, J.F., LAWSON, L.G. (1996): Lameness in tied Danish dairy cattle: the possible influence of housing systems, management, milk yield and prior incidents of lameness. *Preventive Veterinary Medicine* 29, 135-149.

ANDREASEN S.N., WEMELSFELDER F., SANDOE P., FORKMAN B. (2012): The correlation of Qualitative Behavior Assessments with Welfare Quality[®] protocol outcomes in on-farm welfare assessment of dairy cattle. *Applied Animal Behaviour Science* 143, 9-17.

ANONYM (2009): Five Freedoms.

URL: <http://www.fawc.org.uk/freedoms.htm> (21.01.2014)

BLOKHUIS, H.J., HOPSTER, H., GEVERINK, N.A., KORTE, S.M., VAN REENEN, C.G. (1998): Studies of stress in farm animals. *Comparative Haematology International* 8, 94-101.

BLOKHUIS, H.J. (2005): Introducing the Welfare Quality[®] Project. In: Science and society improving animal welfare. Welfare Quality Consortium: Lelystad, Netherlands.

BLOKHUIS, H.J. (2008): International cooperation in animal welfare: The Welfare Quality[®]. *Acta Veterinaria Scandinavica* 2008, 50 (Suppl 1):S10.

BOIVIN, X., LENSINK, J., TALLET, C., VEISSIER, I. (2003): Stockmanship and farm animal welfare. *Animal Welfare* 12, 479-492.

BREUER, K., HEMSWORTH, P.H., BARNETT, J.L., MATTHEWS, L.R., COLEMAN, G.J. (2000): Behavioural response to humans and the productivity of commercial dairy cows. *Applied Animal Behaviour Science* 66, 273-288.

BRÖRKENS N., PLESCH G., LAISTER S., ZUCCA D., WINCKLER C., MINERO M., KNIERIM U. (2009): Reliability testing concerning behaviour around resting in cattle in dairy cows and beef bulls. In: Forkman, B., Keeling, L. (ed): *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality[®] Reports No. 11*, 7-24.

BUTLER, W.R. (2003): Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. *Livestock Production Science* 83, 211-218.

CANALI, E., KEELING, L. (2009): Welfare Quality[®] project: from scientific research to on farm assessment of animal welfare. *Italian Journal of Animal Science* 8: S2, 900-903.

CANALI, E., WHAY, H.R., LEACH, K.A. (2009): Cattle health status. In: Forkman, B., Keeling, L. (ed): *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality[®] Reports No. 11*, 77-88.

CHAPLIN, S., MUNKSGAARD, L. (2001): Evaluation of a simple method for assessment of rising behaviour in tethered dairy cows. *Animal Science* 72, 191-197.

CLARKSON, M.J., DOWNHAM, D.Y., FAAULL, W.B., HUGHES, J.W., MANSON, F.J., MERRITT, J.B., MURRAY, R.D., RUSSELL, W.B., SUTHERST, J.E., WARD, W.R. (1996): Incidence and prevalence of lameness in dairy cattle. *Veterinary Record*, 138, 563-567.

COOK, N.B., BENNETT, T.B., NORDLUND, K.V. (2005): Monitoring Indices of Cow Comfort in Free-Stall-Housed Dairy Herds. *Journal of Dairy Science* 88, 3876-3885.

DANUSER, J., REGULA, G. (2001): Evaluation der Ökomassnahmen und Tierhaltungsprogramme: Bereich Artgerechte Tierhaltung, Vierter Zwischenbericht. Bundesamt für Veterinärwesen, Bern.

DE ROSA, G., TRIPALDI, C., NAPOLITANO, F., SALTALAMACCIA, F., GRASSO, F., BISEGNA, V., BORDI, A. (2003): Repeatability of some animal-related variables in dairy cows and buffaloes. *Animal Welfare* 12, 625-629.

DÖRFLER, R. L. (2007): The assessment of animal welfare in diverse livestock production systems: ethical and scientific issues. Verlag Dr. Köster Berlin, 1. Auflage, 1-38.

DYER, T.G. (2012): Water requirements and quality issues for cattle. The university of Georgia, cooperative extension.
URL: http://www.caes.uga.edu/publications/pubDetail.cfm?pk_id=7719 (11.02.2014)

FRASER, D., SORENSEN, J.T. (2010): On-farm welfare assessment for regulatory purposes: Issues and possible solutions. *Livestock Production Science* 131, 1-7.

GOSLING, S.D. (2001): From mice to men: what can we learn about personality from animal research?. *Psychological Bulletin* 127, 45-86.

GROTH, W. (1984): Mängel im Haltungssystem als Ursache von Gliedmassenschäden bei Rindern. *Tierärztliche Umschau* 39, 196-201.

GROTH, W. (1985): Kriterien für die Beurteilung von Haltungssystemen für Milchkühe und Mastbullen aus klinischer Sicht. *Tierärztliche Umschau* 40, 739-750.

HASSALL, S.A., WARD, W.R., MURRAY, R.D. (1993): The effect of lameness on the behaviour of cows during summer. *Veterinary Record* 132, 578-580.

HAUSCHILD, B.K. (2012): Evaluierung von Indikatoren des Ruheverhaltens von Milchkühen in Liegeboxen-Laufställen für die Anwendung im On-Farm Welfare Assessment. Diplomarbeit, Universität für Bodenkultur, Wien, 29.

HEMSWORTH, P.H. (2003): Human-animal interactions in livestock production. *Applied Animal Behaviour Science* 81, 185-198.

HÖRNING, B. (2001): The Assessment of Housing Conditions of Dairy Cows in Littered Loose Housing Systems using Three Scoring Methods. *Acta Agriculturae Scandinavica, Section A - Animal Science*, S30, 42-47.

ITO, K., WEARY, D.M., VON KEYSERLINGK, M.A.G. (2009): Lying behaviour: Assessing within- and between-herd variation in free-stall-housed dairy cows. *Journal of Dairy Science* 92, 4412-4420.

JOHNSEN, P.F., JOHANNESSON, T., SANDØE, P. (2001): Assessment of Farm Animal Welfare at Herd Level: Many Goals, Many Methods. *Acta Agriculturae Scandinavica, Section A - Animal Science*, S30, 26-33.

KIRCHNER M.K., SCHULZE WESTERATH H., KNIERIM U., TESSITORE E., COZZI G., PFEIFFER C., WINCKLER C. (2014): On-farm animal welfare assessment in beef bulls: consistency over time of single measures and aggregated Welfare Quality® scores. *The Animal Consortium* 2014, 8 (3): 461-469.

KNIERIM, U. (1999): Das Verhalten von Färsen bei der Einzel- oder Gruppeneinführung in die Milchviehherde. *KTBL-Schrift*, 382, 115-120.

KNIERIM, U., WINCKLER, C. (2009): On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare* 18, 451-458.

KONDO, S., SEKINE, J., OKUBO, M., ASAHIDA, Y. (1989): The effect of group size and space allowance on the agonistic and spacing behaviour of cattle. *Applied Animal Behaviour Science* 24, 127-135.

KORTE, S.M., OLIVIER, B., KOOLHAAS, J.M. (2006): A new animal welfare concept based on allostasis. *Physiology and Behaviour* 92, 422-428.

KRAWCZEL, P.D., HILL, C.T., DANN, H.M., GRANT, R.J. (2008): Short Communication: Effect of Stocking Density on Indices of Cow Comfort. *Journal of Dairy Science* 91, 1903-1907.

LADEWIG, J., SMIDT, D. (1989): Behaviour, episodic secretion of cortisol, and adrenocortical reactivity in bulls subjected to tethering. *Hormones and Behaviour* 23, 344-360.

LAISTER, S., BRÖRKENS, N., LOLLI, S., ZUCCA, D., KNIERIM, U., MINERO, M., CANALI, E., WINCKLER, C. (2009): Reliability of measures of agonistic behavior in dairy and beef cattle. In: Forkman, B., Keeling, L. (ed): *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality® Reports No. 11*, 95-112.

LANIER, J.L., GRANDIN, T., GREEN, R.D., AVERY, D., MCGEE, K. (2000): The relationship between reaction to sudden, intermittent movements and sounds and temperament. *Animal Science* 78, 1467-1474.

LEACH, K.A., KNIERIM, U., WHAY, H.R. (2009a): Condition scoring for dairy and beef cattle and veal calves. In: Forkman, B., Keeling, L. (ed): Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality® Reports No. 11, 1-6.

LEACH, K.A., DIPPEL, S., HUBER, J., MARCH, S., WINCKLER, C., WHAY, H.R. (2009b): Assessing lameness in cow kept in tie-stalls. Journal of Dairy Science 92, 1567-1574.

LEACH, K.A., WINCKLER, C., WHAY, H.R. (2009c): Lameness in dairy and beef cattle and veal calves. In: Forkman, B., Keeling, L. (ed): Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality® Reports No. 11, 35-41.

LINN, J., RAETH-KNIGHT, M. (o.J.): Water quality and quantity for dairy cattle. University of Minnesota.

URL: <http://sheboygan.uwex.edu/files/2010/08/UMWaterQuality.pdf> (11.02.2014)

MENKE, C., WAIBLINGER, S., FÖLSCH, D.W., WIEPKEMA, P.R. (1999): Social behaviour and injuries of horned dairy cows in loose housing systems. Animal Welfare 8, 243-258.

MIELE, M., ROEX, J. (2005): Farm Animal Welfare Concerns – Consumers, Retailers and Producers. Welfare Quality® Reports No. 1.

MORTENSEN, B. (1978): Stanchion barn systems for dairy cows: an ethological evaluation. In: Proceedings of the first world congress on ethology applied to zootechnics: Madrid, Spain, 181-194.

MÜLLER, C., LADEWIG, J., THIELSCHER, H. H., SMIDT, D. (1989): Behaviour and heart rate of heifers housed in tether stanchions without straw. Physiology and Behaviour 46, 751-754.

MUNKSGAARD, L., SIMONSEN, H.B. (1996): Behavioral and pituitary adrenal-axis responses of dairy cows to social isolation and deprivation of lying down. Animal Science 74, 769-778.

MUNKSGAARD, L., JENSEN, M.B., PEDERSEN, L.J., HANSEN, S.W., MATTHEWS, L. (2005): Quantifying behavioural priorities: effects of time constraints on behavior of dairy cows, Bos Taurus. Applied Animal Behaviour Science 92, 3-14.

MURPHY, M.R. (1992): Water metabolism of dairy cattle. Journal of Dairy Science 75, 326-333.

OVERTON, M.W., SISCHO, W.M., TEMPLE, G.D., MOORE, D.A. (2002): Using Time-Lapse Video Photography to Assess Dairy Cattle Lying Behaviour in a Free-Stall Barn. Journal of Dairy Science 85, 2407-2413.

PHILLIPS, C.J., MORRIS, I.D. (2002): The ability of cattle to distinguish between, and their preference for, floors with different levels of friction, and their avoidance of floors contaminated with excreta. *Animal Welfare* 11, 21-29.

PLESCH, G., BROERKENS, N., LAISTER, S., WINCKLER, C., KNIERIM, U. (2010): Reliability and feasibility of selected measures concerning resting behaviour for the on-farm welfare assessment in dairy cows. *Applied Animal Behaviour Science* 126, 19-26.

REID, I.M., ROBERTS, C.J., TREACHER, R.J., WILLIAMS, L.A. (1986): Effect of body condition at calving on tissue mobilization, development of fatty liver and blood chemistry of dairy cows. *Animal Production* 43, 7-15.

RODRIGUEZ-LAINZ, A., HIRD, D.W., CARPENTER, T.E., READ, D.H. (1996): Case-control study of papillomatous digital dermatitis in southern California dairy farms. *Preventive Veterinary Medicine* 28, 117-131.

ROSENBERGER, G. (1970): *Krankheiten des Rindes*. Paul Parey, Berlin.

SCAHAW (2001): *The welfare of Cattle Kept for Beef Production*. European Commission, Brussels.

SCOTT, E.M., NOLAN, A.M., FITZPATRICK, J.L. (2001): Conceptual and Methodological Issues Related to Welfare Assessment: A Framework for Measurement. *Acta Agriculturae Scandinavica, Section A - Animal Science*, S30, 5-10.

SCHRADER, D.A., ROTH, H.R., WINTERLING, C., BRODMANN, N., LANGHANS, W., GEYER, H., GRAF, B. (2001): The occurrence of tail tip alterations in fattening bulls kept under different husbandry conditions. *Animal Welfare* 10, 119-130.

SCHREINER, D.A., RUEGG, P.L. (2003): Relationship between udder and leg hygiene scores and subclinical mastitis. *Journal of Dairy Science* 86, 3460-3465.

SCHUKKEN, Y.H., GROMMERS, F.J., VAN DER GEER, D., ERB, H.N., BRAND, A. (1990): Risk factors for clinical mastitis in herds with low bulk milk somatic cell count: 1. Data and risk factors for all cases. *Journal of Dairy Science* 73, 3463-3471.

SCHULZE WESTERATH, H., LAISTER, S., WINCKLER, C., KNIERIM, U. (2009): Exploration as an indicator of good welfare in beef bulls: an attempt to develop a test for on-farm assessment. *Applied Animal Behaviour Science* 116, 126-133.

SINGH, S.S., WARD, W.R., LAUTENBACH, K., MURRAY, R.D. (1993): Behaviour of lame and normal dairy cows in cubicles and a straw yard. *Veterinary Record* 133, 204-208.

SINGH, S.S., WARD, W.R., HUGHES, J.W., LAUTENBACH, K., MURRAY, R.D. (1994): Behaviour of dairy cows in a straw yard in relation to lameness. *Veterinary Record* 135, 204-208.

- WAGNER, J.J., LUSBY, K.S., OLTJEN, J.W., RAKESTRAW, J., WETTEMANN, R.P., WALTERS, L.E. (1988): Carcass composition in mature Hereford cows: estimation and effect on daily metabolizable energy intake during winter. *Animal Science* 66, 603-612.
- WAIBLINGER, S., KNIERIM, U., WINCKLER, C. (2001): The Development of an Epidemiologically Based On-Farm Welfare Assessment System for use with Dairy Cows. *Acta Agriculturae Scandinavica, Section A - Animal Science*, S30, 73-74.
- WAIBLINGER, S., BOIVIN, X., PEDERSEN, V., TOSI, M., JANCZAK, A.M., VISSER, E.K., JONES, R.B. (2006): Assessing the human-animal relationship in farmed species: a critical review. *Applied Animal Behaviour Science* 101, 185-242.
- WEBSTER, A.J.F. (2001): Review – Farm Animal Welfare: The Five Freedoms and the Free Market. *The veterinary Journal* 2001, 229-237.
- WEBSTER, J. (1994): *Animal welfare: a cool eye towards Eden*. Blackwell, Oxford, 11.
- WEBSTER, J. (2005): *Animal welfare: limping toward Eden*. 2nd ed., Blackwell, Oxford, 12.
- WECHSLER, B., SCHAUB, J., FRIEDLI, K., HAUSER, R. (2000): Behaviour and leg injuries in dairy cows kept in cubicle systems with straw bedding or soft lying mats. *Applied Animal Behaviour Science* 69, 189-197.
- WELFARE QUALITY[®] (2009): *Welfare Quality[®] assessment protocol for cattle*. Welfare Quality[®] Consortium: Lelystad, Netherlands.
- WEMELSFELDER, F., HUNTER, E.A., MENDEL, M.T., LAWRENCE, A.B. (2001): Assessing the whole animal: a Free-Choice-Profiling approach. *Applied Animal Behaviour Science* 62, 209-220.
- WEMELSFELDER, F., FARISH, M. (2004): Qualitative categories for the interpretation of sheep welfare: a review. *Animal Welfare* 13, 261-268.
- WEMELSFELDER, F. (2007): How animals communicate quality of life: the qualitative assessment of animal behaviour. *Animal Welfare* 16, 25-31.
- WHAY, H.R., WATERMAN, A.E., WEBSTER, A.J.F. (1997): Associations between locomotion, claw lesions and nociceptive threshold in dairy heifers during the peri-partum period. *Veterinary Journal* 154, 155-161.
- WHAY, H.R. (2007): The journey to animal welfare improvement. *Animal Welfare* 16, 117-122.
- WIERENGA, H.K. (1984): The social behaviour of dairy cows: some differences between pasture and cubicle system. In: *Proceedings of the International Congress of Applied Ethology in Farm Animals*. KTBL, Darmstadt, 135-138.

WINCKLER, C., BRINKMANN, J., GLATZ, J. (2007): Long-term consistency of selected animal-related welfare parameters in dairy farms. *Animal Welfare* 16, 197-199.

WINCKLER, C. (2008): The use of animal-based health and welfare parameters – what is it all about? URL: <http://orgprints.org/13405/1/13405.pdf> (24.01.2014)

WINCKLER, C., WILLEN, S. (2001): The reliability and repeatability of a lameness scoring system for use as an indicator of welfare in dairy cattle. *Acta Agriculturae Scandinavica, Section A - Animal Science*, S30, 103-107.

WINCKLER, C., BÜHNEMANN, A., SEIDEL, K. (2002): Social behaviour of commercial dairy herds as a parameter for on-farm welfare assessment. In: Koene, P. (ed): *Proceedings of the 36th International Congress of the ISAE*. ISAE: Wageningen, The Netherlands, 86.

WINDSCHNURER, I., SCHMIED, C., BOIVIN, X., WAIBLINGER, S. (2008): Reliability and inter-test relationship of tests for on-farm assessment of dairy cows' relationship to humans. *Applied Animal Behaviour Science* 114, 37-53.

WINDSCHNURER, I., BOIVIN, X., WAIBLINGER, S. (2009): Reliability of an avoidance distance test for the assessment of animals' responsiveness to humans and a preliminary investigation of its association with farmers' attitudes on bull fattening farms. *Applied Animal Behaviour Science* 117, 117-127.