



Graduate Student Literature Review: Evaluating the appropriate use of wearable accelerometers in research to monitor lying behaviors of dairy cows*

S. J. Hendriks,^{1†} C. V. C. Phyn,² J. M. Huzzey,³ K. R. Mueller,⁴ S-A. Turner,^{2‡} D. J. Donaghy,¹ and J. R. Roche^{2,5§#}

¹School of Agriculture and Environment, Massey University, Palmerston North 4410, New Zealand

²DairyNZ Ltd., Hamilton 3240, New Zealand

³Department of Animal Science, California Polytechnic State University, San Luis Obispo 93407

⁴School of Veterinary Sciences, Massey University, Palmerston North 4410, New Zealand

⁵School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

ABSTRACT

Until recently, animal behavior has been studied through close and extensive observation of individual animals and has relied on subjective assessments. Wearable technologies that allow the automation of dairy cow behavior recording currently dominate the precision dairy technology market. Wearable accelerometers provide new opportunities in animal ethology using quantitative measures of dairy cow behavior. Recent research developments indicate that quantitative measures of behavior may provide new objective on-farm measures to assist producers in predicting, diagnosing, and managing disease or injury on farms and allowing producers to monitor cow comfort and estrus behavior. These recent research developments and a large increase in the availability of wearable accelerometers have led to growing interest of both researchers and producers in this technology. This review aimed to summarize the studies that have validated lying behavior derived from accelerometers and to describe the factors that should be considered when using leg-attached accelerometers and neck-worn collars to describe lying behavior (e.g., lying time and lying bouts) in dairy cows for research purposes. Specifically, we describe accelerometer technology, including the instrument properties and methods for recording motion; the raw data output from accelerometers; and methods developed for the transformation of raw data into meaningful and

interpretable information. We highlight differences in validation study outcomes for researchers to consider when developing their own experimental methodology for the use of accelerometers to record lying behaviors in dairy cows. Finally, we discuss several factors that may influence the data recorded by accelerometers and highlight gaps in the literature. We conclude that researchers using accelerometers to record lying behaviors in dairy cattle should (1) select an accelerometer device that, based on device attachment and sampling rate, is appropriate to record the behavior of interest; (2) account for cow-, farm-, and management-related factors that could affect the lying behaviors recorded; (3) determine the appropriate editing criteria for the accurate interpretation of their data; (4) support their chosen method of recording, editing, and interpreting the data by referencing an appropriately designed and accurate validation study published in the literature; and (5) report, in detail, their methodology to ensure others can decipher how the data were captured and understand potential limitations of their methodology. We recommend that standardized protocols be developed for collecting, analyzing, and reporting lying behavior data recorded using wearable accelerometers for dairy cattle.

Key words: precision, automated, technology

INTRODUCTION

Wearable electronic monitoring technologies could change the features of intensive, large-scale dairy farming by allowing for the individualized monitoring and management of animals (Bewley et al., 2017). Since the early 2000s, there has been a substantial increase in published literature investigating the use of wearable electronic monitoring technologies for adoption in commercial farming environments (e.g., Huzzey et al., 2005; Sepúlveda-Varas et al., 2014; Rodríguez-Jimenez

Received November 10, 2019.

Accepted July 1, 2020.

*Submitted to the 2021 ADSA Foundation Graduate Student Literature Review Competition (Production, PhD) on August 6, 2020.

†Corresponding author: s.hendriks@massey.ac.nz

‡Dairy Goat Co-operative, PO Box 1398, Hamilton 3240, New Zealand.

§Advisor (john.roche@mpi.org.nz).

#Current address: Ministry for Primary Industries, Wellington 6140, New Zealand.

et al., 2018). Worldwide, intensive dairy management systems have experienced increasing herd sizes, while skilled and experienced labor has become less available, resulting in less ability and time to monitor individual animals (Bewley et al., 2017).

Dairy cows are highly motivated to lie down and will trade off other behavioral activities to maintain daily lying times (Munksgaard et al., 2005) and is an important component of cow comfort and welfare (von Keyserlingk et al., 2012). Changes in behavior—specifically, changes in lying behavior—of dairy cows are well recognized as an indicator of injury (e.g., lameness; González et al., 2008) or ill health (e.g., metabolic or infectious disease; Itle et al., 2015; Neave et al., 2018) and can be used to detect estrus (Dolecheck et al., 2015). Historically, lying behavior was monitored in experimental work through visual and video observations (Rutten et al., 2013) and on farm through visual observations and manual assessment; however, manual assessment of animal behavior is subjective (Weary et al., 2009) and is time- and resource-intensive (Mattachini et al., 2013). In contrast, wearable electronic monitoring technologies have evolved substantially with the development of more accessible hardware and software (Brown et al., 2013) and can overcome many of the limitations of traditional methods of behavior monitoring, particularly in pasture-based systems (O'Driscoll et al., 2008). The use of accelerometers allows behavior to be measured remotely, for extensive periods [e.g., HOBO (Onset Computer Corp., Pocasset, MA) and IceQube (IceRobotics Ltd., Edinburgh, UK)], and in real time [e.g., The AMS Activity Monitor (**ACT**; Lely, Maassluis, the Netherlands)], to give detailed information about lying behavior.

The increasing accessibility and rapidly growing market of technologies that are able to measure lying behavior in dairy cows (Bewley et al., 2017) and their potential practical application on farm indicate that these technologies are of high interest to both researchers and producers (Borchers et al., 2016). The technologies of interest for the present review are those that use accelerometers attached to the leg or mounted via a collar around the neck to measure lying behavior in dairy cattle (Martiskainen et al., 2009). Lying behaviors that can be derived from leg-mounted accelerometers include lying time, lying bouts (**LB**), **LB** duration, and lying laterality (Gibbons et al., 2012).

When reviewing studies using the same type of accelerometer under similar farming systems, a lack of user understanding of the appropriate use of accelerometer devices is apparent (e.g., discrepancies in the protocols used for collecting, analyzing, and reporting data between studies using the same type of accelerometer in cows under similar systems; Supplemental Materials

and Supplemental Table S1; <https://doi.org/10.3168/jds.2019-17887>). To obtain accurate measures of lying behavior using accelerometers in dairy cows, thought should be given to developing standardized protocols for collecting, analyzing, and reporting data across studies (Anderson et al., 2013), including understanding the management, editing, and analysis of large data sets. Several methodological issues affect the accuracy of accelerometer-derived lying behavior and should be considered by all users.

In this review, we describe the instrument properties and methods for recording lying behavior using accelerometers to improve the understanding of the importance of sampling frequency and sampling interval selection to record the behavior of interest. Our review highlights several factors for researchers to consider when deploying the devices in the field in dairy cows and when editing large data sets derived from accelerometers. We discuss potential limitations in the validation of accelerometer-derived lying behavior that researchers need to evaluate when developing their own experimental methodology. Finally, we will discuss the need to create a robust set of criteria and an adequate explanation of those criteria in the methodology of studies using accelerometers in dairy cows.

LITERATURE SEARCH

The web-based literature databases searched were Web of Science (<http://wokinfo.com/>) and Scopus (<https://www.scopus.com>). Scientific articles (both peer-reviewed and non-peer-reviewed), conference proceedings, and abstracts written in English were considered. Search terms were accelerometer(s), automatic, automatic monitoring, behavior (behavior), behavior monitoring (behavior-monitoring), cattle, cow(s), dairy, dairy cow(s), data logger(s), logger(s), lying behavior, recording lying, sampling interval(s), technology, technical note, 3-dimensional, validation. Inclusion criteria were that the paper must report on lying behavior measured by either omnidirectional, 2- or 3-dimensional (**2D** or **3D**) accelerometers attached by collars or leg bracelets in dairy cows.

This review focuses on validation studies that have used identical loggers on opposite legs, direct observation, or video recording to validate accelerometer-derived lying behaviors. Studies comparing the performance of 2 different technologies for validation are not considered in this review or reported in tables because of the different sampling frequencies and sampling intervals between devices that could confound the data. The accelerometer-derived lying behaviors that were the focus of this review are lying time, **LB**, and **LB** duration. Lying time can be defined as a durational

activity and is typically measured as the time an animal spends engaged in lying within a set time (e.g., daily lying time, 24-h period; h/d); LB can be defined as a transitional activity and includes instances where animals transition from standing to lying positions and then back to standing (Ledgerwood et al., 2010). Lying bouts are measured as the number of times an animal transitions (e.g., number of bouts/d). Further, the time of a single bout of lying can be defined as the LB duration (e.g., min/bout). Although the total duration of an activity is valuable for understanding the time budgets of animals, transitional behaviors can provide valuable information that total durational activities cannot explain (Kok et al., 2015).

From our literature search, 13 studies have validated a range of accelerometers for durational (daily lying and standing time) and transitional activities (e.g., LB) under a range of conditions, sampling frequencies and intervals, and editing criteria. The purpose of this review is not to give an overall recommendation for the use of accelerometers. Instead, we provide a concise summary of the validation studies that have been undertaken in adult cattle for others to reference to discern the appropriate experimental methodology based on their study design and objectives. General information about these studies are described and summarized in Tables A1, A2, and A3 in the Appendix of this review, including information about the validation study design, the accelerometers used, and behaviors validated.

ACCELEROMETER TECHNOLOGY

Wearable accelerometers can remotely and efficiently collect detailed data related to measures of animal behavior (Brown et al., 2013). The accelerometer measures changes in velocity over time and is attached to the animal in a specific location so that the orientation of the device can provide detailed information about movement and body position relating to the behavior of interest (Scheibe and Gromann, 2006). A summary of the devices referred to throughout this review, manufacturer details, behaviors recorded, attachment of the device, sampling intervals available for selection to record the behavior of interest, data retrieval, and on-board data storage is provided in Table 1.

Device Attachment

Accelerometers require attachment on a specific anatomical location and orientation on the body through the use of either collars (e.g., attached around the neck of the cow; Martiskainen et al., 2009), leg bands (e.g., attached to the lateral or medial side of the front or hind limb; Müller and Schrader, 2003), and user-

designed device housing (e.g., attached using vet wrap; Rodríguez-Jimenez et al., 2018), depending on the behavior of interest. User-designed device housings are typically used for research purposes if the accelerometer is not contained within an easily attachable housing [e.g., The Actiwatch Activity Monitoring System (Actiwatch; Cambridge Neurotechnology, Fenstanton, Cambridgeshire, UK) and HOB0].

The specific location of accelerometer attachment determines the behavior that can be distinguished from the raw data and can affect the accuracy of the lying behaviors that can be recorded. For example, attachment in the same location on opposite legs (e.g., left and right back legs) has little influence on the accuracy of lying time and LB information (Shepley et al., 2017). Although accelerometer attachment on the front or back leg appears to have little influence on the accuracy of lying time information (Nielsen et al., 2018), it appears the accuracy of LB information is reduced when the device is attached to the front compared with the back leg (Charlton et al., 2017); however, research in this area is limited.

Accelerometer attachment around the neck can also affect the accuracy of lying time recorded. Elischer et al. (2013) reported that lying is not distinguishable from standing when using the ACT device, which is a neck-worn collar where activity units are reported in 2-h blocks (Table A2). In agreement, Martiskainen et al. (2009) reported that several misclassifications of lying occurred (sensitivity = 0.80) when using a neck-worn accelerometer (ADXL330, Analog Devices Inc., Wilmington, MA) at a higher sampling frequency (10 Hz) and reported that this could be explained by a lack of difference in the position of the neck when the cow is lying compared with standing. It may also be due, in part, to movement of the accelerometer on the animal obscuring measurements when attached around the neck (Brown et al., 2013). Therefore, although neck-worn collars can classify lying to a greater degree at higher sampling frequencies, the neck is not an appropriate location to attach an accelerometer when the interest is in accurately recording lying time; fixed accelerometer attachment to a hind limb appears to be more suited to measure lying behaviors. These examples emphasize the importance of considering the accelerometer device and options for device attachment.

Data Recording

An accelerometer is a spring-like, piezoelectric sensor that generates a wave-like voltage signal that is proportional to the acceleration (change in velocity) it experiences over time (Dow et al., 2009). The sensor measures either a summed acceleration or the accelera-

Table 1. Description of the devices described this review including the name of the technology, manufacturer, behavior recorded, device attachment, sampling interval(s), data retrieval, and on-board storage

Technology ¹	Manufacturer	Behavior recorded ²	Attachment ³	Sampling interval(s)	Data retrieval ⁴	On-board storage ⁵
Actiwatch	Cambridge Neurotechnology	Activity	Leg band	2 s to 15 min	Manual	≥1.5 d
AfiAct	Afimilk	Activity, LB, and LT	Leg band	1 min to 2 h	Manual	—
AfiTagII	Afimilk	LB and LT	Leg band	15 min	UHF	22 h
ACT	Lely	Activity, EAT, LT, and RM	Collar	2 h	UHF	24 h
CowScout Leg Sensor	GEA Farm Technologies	Activity, LB, and LT	Leg band	15 min	UHF	24 h
Halter	Halter USA Inc.	Activity	Collar	—	Cellular	—
HOBO	Onset Computer Corp.	<i>g</i> force or degrees of tilt on the 3 axes (user classification required)	User-designed attachment	Fast mode: 0.01 to 0.99 s Normal mode: 1 s to 18 h	Manual	≥3.6 min ≥6 h
IceQube	IceRobotics Ltd.	Activity, LB, and LT	Leg band	15 min, 1 h, 2 h, 1 d, and 1 wk	Manual	9 d
IceTag 2D	IceRobotics Ltd.	Activity, LB, and LT	Leg band	1 s, 1 min, 15 min, 1 h, 2 h, 1 d, and 1 wk	Manual	60 d
IceTag 3D	IceRobotics Ltd.	Activity, LB, and LT	Leg band	1 s, 1 min, 15 min, 1 h, 2 h, 1 d, and 1 wk	Manual	60 d
IDA Tracker	Connecterra	Activity, EAT, LT, and RM	Collar	—	UHF	24 h
The Track a Cow	ENG Systems	LB, LT, and visits to feed bunk	Leg band	1 min	Cellular	24 h

¹Technology: Actiwatch = The Actiwatch Activity Monitoring System (Cambridge Neurotechnology, Fenstanton, Cambridgeshire, UK), AfiAct = AfiAct Pedometer Plus (ceased manufacture) and AfiTagII = AfiTag II Pedometer (AfiMilk, Kibbutz, Afikim, Israel), ACT = Automatic Milking System Activity Monitor (Lely, Maassluis, the Netherlands), CowScout = CowScout Leg Sensor (GEA Farm Technologies, Bönen, Germany), Halter = Halter Activity Monitor (Halter USA Inc., Auckland, New Zealand), HOB0 = HOB0 Pendant G Acceleration Data Logger (Onset Computer Corp., Pocasset, MA), IceQube = IceQube Activity Monitor (IceRobotics Ltd., Edinburgh, UK), IceTag 2D = IceTag 2.004 Activity Monitor (ceased manufacture 2008; IceRobotics Ltd.), IceTag 3D = IceTag 3D Activity Monitor (IceRobotics Ltd.), IDA Tracker = IDA Tracker (Connecterra B.V., Amsterdam, the Netherlands), TAC = The Track a Cow (ENGs, Rosh Pina, Israel).

²LB = lying bouts; LT = lying time; EAT = eating time; RM = rumination time.

³Attachment refers to the way the device is attached to the cow. Collar refers to those worn around the neck of the cow. Leg band refers to those worn on the leg (typically the hind leg is preferred).

⁴Cellular = data are transferred automatically through the use of satellites; Manual = manual download is required using product-specific software; UHF = data are transferred automatically via ultra-high-frequency data download technology using either fixed or portable wireless transceivers.

⁵On-board storage is the total recording time of data storage available on the device based on the shortest possible sampling interval. For devices that allow the selection of longer sampling intervals, the number of recording days increases with the selection of a longer sampling interval (i.e., fewer data points recorded per day).

tion at defined intervals and is derived when the sensor is deformed by gravitational (due to changes in animal posture) as well as inertial acceleration (due to animal movement) (Shepard et al., 2008). Acceleration is measured in the direction of a single plane of movement. In a 2D accelerometer, the acceleration is measured along the X- and Y-axes; however, accelerometers have been developed to allow the acceleration to be recorded across 3 axes [e.g., IceTag 2.004 Activity Monitor (IceTag 2D; ceased manufacture 2008, IceRobotics Ltd.) vs. IceTag 3D Activity Monitor (IceTag 3D; IceRobotics Ltd.)]; Table A1). In a 3D accelerometer, acceleration is recorded across the X-, Y-, and Z-axes, which are aligned orthogonally to signal vertical, forward, and lateral movement (Tamura et al., 2019). Omnidirectional accelerometers, where acceleration is measured along multiple planes in all directions are rarely used in research involving cattle, with only one early study in dairy cows appearing in our literature search (e.g., Actiwatch, developed for medical research in humans; Müller and Schrader, 2003).

To record lying behavior using devices fixed to the leg of the cow, the 3D accelerometer allows specific leg orientations associated with lying and standing behavior (X- and Y-axes) to be recorded as well as lying laterality (Z-axis; e.g., lying on the left or right side; Robert et al., 2009). The X-axis is perpendicular to the ground during standing events (Figure 1a), the Y-axis is perpendicular to the ground during lying events (Figure 1b; Robert et al., 2009), and the Z-axis runs parallel to the ground pointing away from the sagittal plane (Darr and Epperson, 2009). To record lying behavior using neck-worn collars, the 3D accelerometer is attached at the top or the side of the cow's neck. Acceleration along the x-axis is associated with vertical movement of the cow's head (e.g., up and down), acceleration along the y-axis is associated with horizontal movement of the cow's head (e.g., forward and backward), and acceleration along the Z-axis is associated with lateral movement of the cow's head (Figure 2; Tamura et al., 2019).

Accelerometers record the change in velocity along each of the axes, separately, per unit time. Although every device is different, the 3D accelerometer will capture information about the acceleration along the 3 axes at a predetermined sampling frequency (Brown et al., 2013). Sampling frequencies for technologies that have been validated in the literature range from 1 to 100 Hz and are either fixed and predetermined by the product manufacturer [e.g., IceTag 3D (16 Hz); IceQube (4 Hz); (IceRobotics Ltd., 2017)] or can be programmed by the user [HOBO; range: 0.017 to 100 Hz (Onset, 2019)]. A sampling frequency of 1 Hz corresponds to the capture of 1 acceleration measurement per second on each axis

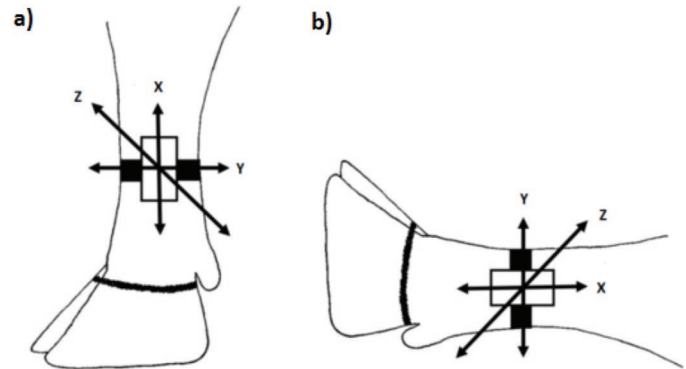


Figure 1. Illustration of the position of the 3-dimensional accelerometer and X-, Y-, and Z-axes on the lateral aspect of the rear limb in a standing (a) and lying (b) positions.

and, therefore, would correspond to 3 measurements per second, which rapidly accumulates into millions of logged measurements over time (Brown et al., 2013). Consideration of a suitable sampling frequency when choosing a device is important because the sampling frequency is often specific to the device and typically cannot be altered (Chen and Bassett, 2005).

Unnecessarily high sampling frequencies should be avoided when digital storage space is limited (Halsey

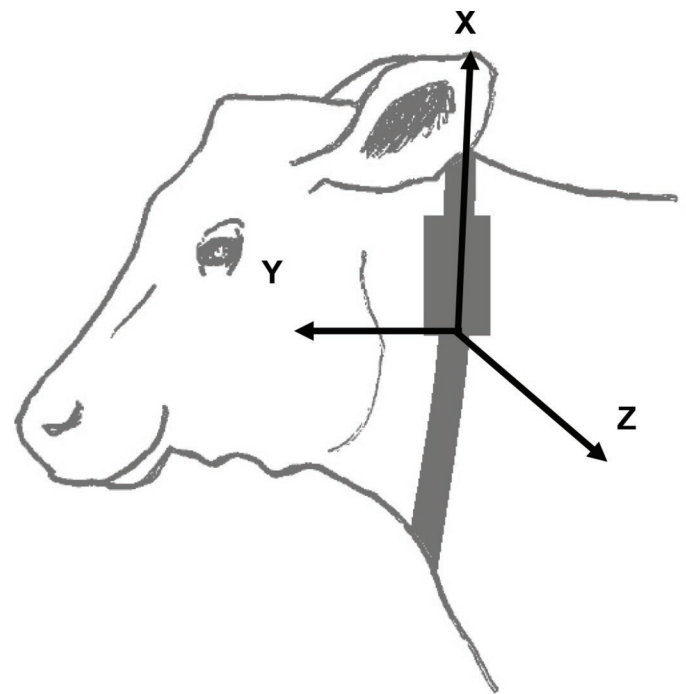


Figure 2. Position of the 3-dimensional accelerometer and illustration of X-, Y-, and Z-axes at the side of the cow's neck when the cow is standing.

et al., 2009). For example, the HOBO has 64 KB of internal memory (approximately 21.8K combined X-, Y-, and Z-axis readings; Onset, 2019) and this allows data storage for ~6 h using a 1-s sampling interval or for 15 d using a 1-min sampling interval. Typically, devices used for research purposes allow shorter sampling intervals [e.g., HOBO (≥ 0.01 s); IceTag 2D and IceTag 3D (≥ 1 s)], where the device may be removed more regularly, whereas devices used commercially often summarize the data within a longer sampling interval ≥ 15 min (e.g., ACT and IceQube) to maximize memory capacity and the battery life of the device (Table 1). However, devices developed for research and commercial purposes are increasingly incorporating automated data download features to reduce the need for such large on-board storage capacity.

Behavior Classification

Accelerometer raw data output contains a wave-like signal with units of voltage that are recorded during

acceleration of the sensors (Brown et al., 2013), and these signals can be used to determine the behavior occurring at each time point (Robert et al., 2009). In devices without software containing proprietary algorithms, before analysis, these wave-like signals must be classified manually into specific behavioral categories through validation studies, which involves synchronizing the raw data output with observed behaviors to determine the specific signals corresponding to specific behaviors (Martiskainen et al., 2009). A graphic representation of the wave-like signals generated on each of the 3-axes (X-, Y-, and Z-axes) is depicted in Figure 3; this identifies different levels of voltage for lying and standing behaviors and the distinctive patterns that correspond with these behaviors.

To classify the waveform patterns for a specific behavior, they can be categorized either using a custom-designed classification system or automatically using software developed by the manufacturer (Brown et al., 2013). Many commercially available devices [e.g., ACT, AfiAct Pedometer Plus (AfiAct; Kibbutz, Afikim, Is-

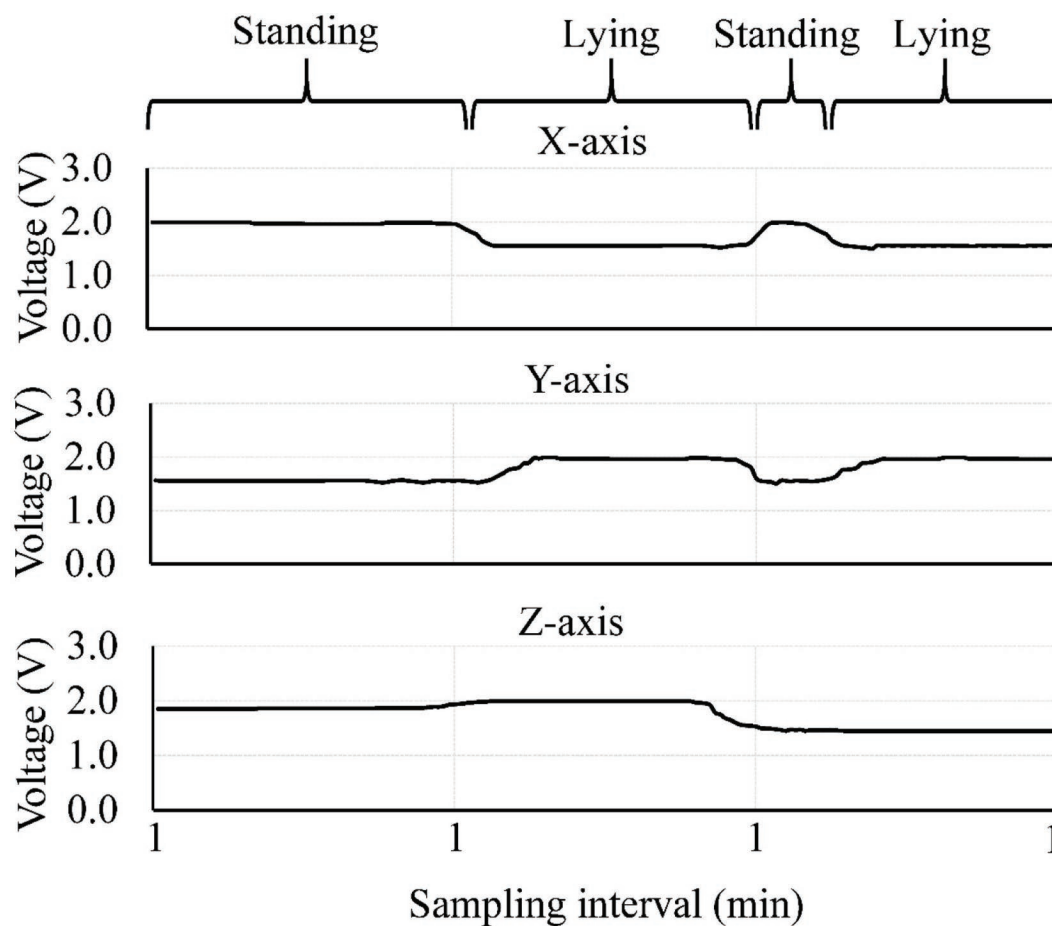


Figure 3. A graphical representation of the wave-like signals generated by different levels of voltage recorded on each of the 3 axes (X-, Y-, and Z-axes) of a 3-dimensional accelerometer.

rael), CowScout (GEA Farm Technologies, Bönen, Germany), IceQube, and The Track a Cow (TAC; ENGS, Rosh Pina, Israel)] and some research devices (e.g., IceTag 2D and IceTag 3D) include software provided by the manufacturer (e.g., IceManager 2010, IceRobotics Ltd.) that categorizes the waveform patterns to time (or count) of a specific behavior (e.g., steps, standing, lying) using a proprietary algorithm, allowing the user to download summary files; however, much of the early research in the field of ethology required development of a classification system and the Actiwatch and HOBO devices are both examples of devices that rely on the user to interpret the data (Müller and Schrader, 2003; Ledgerwood et al., 2010).

Data Retrieval

Raw data are recorded on the on-board memory of the device. Data retrieval includes either manual download using product-specific software after retrieval of the device or automatically using ultra-high-frequency data download technology (Brown et al., 2013; Table 1). Automated data retrieval allows the data to be downloaded through the use of either fixed (e.g., IceQube) or portable (e.g., IDA Tracker, Connecterra B. V., Amsterdam, the Netherlands) wireless transceivers (range: 5 to >1,000 m) or anywhere with cellular coverage through the use of satellites (Halter, Halter USA Inc., Auckland, New Zealand; Brown et al., 2013); however, the ability to manually or automatically download data is device-specific. The downloaded data are summarized according to the manufacturer or user-defined sampling interval (Robert et al., 2009) and can typically be further summarized using the product software (e.g., the IceTag 3D data can be summarized into 1-s, 1-min, 15-min, 1-h, 2-h, 1-d, and 1-wk intervals; Table 1; IceRobotics Ltd., 2017). Although there are several accelerometers on the market with a range of sampling interval options and data retrieval and download options, when choosing a device, it is important to consider the data granularity and sampling duration required, ease of device retrieval, and the storage capacity or data off-loading abilities of the device (Brown et al., 2013).

TRANSFORMATION OF DATA INTO LYING BEHAVIOR

Classifying the waveform patterns into distinct behaviors allows the data to be further summarized into interpretable information in a summary file for each device. Lying behaviors are often classified into durational (e.g., daily lying time) and transitional behaviors

(e.g., daily LB, LB duration, and lying laterality). The summary file contains information about the behavior recorded within each sampling interval and orders the data by a timestamp, which indicates the sequence in which the behaviors occurred (Ungar et al., 2018). For example, for the IceTag, the IceManager 2010 software generates 1 summary file per cow containing lying time (s), standing time (s), and number of steps taken for 1-min sampling intervals. These summary files are typically further summarized to generate single behavioral values across a specific period of time (e.g., over a 24-h period to give daily lying time). Further, this summary output can then be used to calculate transitional behavior using a method similar to that published online by the University of British Columbia (e.g., HOBO devices; Ledgerwood et al., 2010; UBC, 2013) or may be obtained from the device software (e.g., IceTag and IceQube devices; IceManager 2010) to generate a second summary file containing all recorded LB with a start date, start time (hh:mm:ss), and duration (s). These additional files can be further summarized to generate single behavioral values across a specific period (e.g., LB are often summarized over a 24-h period to give total daily LB and mean LB duration).

VALIDATION OF ACCELEROMETERS

Validation of behavior data recorded by accelerometers with a known gold standard can provide useful information about measurement variability and measurement error that may exist and, therefore, the reliability of the data (Watson and Petrie, 2010). There are discrepancies in the protocols used for collecting, analyzing, and reporting data between studies using the same accelerometers in cows under similar systems, and these protocols are not always supported by validation studies, indicating a lack of understanding by some users of the factors that should be considered when using accelerometers (Supplemental Materials and Supplemental Table S1; <https://doi.org/10.3168/jds.2019-17887>). For example, several studies using IceTag devices in lactating dairy cows in freestall barns have either failed to edit data (e.g., false LB were not discarded from the original data) or report whether the data were edited before interpretation (e.g., Telezhenko et al., 2012; Kokin et al., 2014), or referenced validation studies in their methodologies that did not fit with the technology of interest (e.g., Gibbons et al., 2012). Therefore, the purpose of the following section is to disentangle some of the information surrounding the factors that affect validation study outcomes and recommendations such as the gold standards chosen, causes of measurement error and variability, and, finally, the accelerometer recording settings.

Gold Standards for Validation

Validation is important to determine the accuracy of an accelerometer-derived behavior (whether an observed value agrees with the true value; Watson and Petrie, 2010). Validation studies involve taking 2 methods to assess the same variable or outcome and to evaluate how well they agree (Watson and Petrie, 2010). One method is regarded as the “gold standard” and the other method is commonly a quicker, cheaper, or otherwise more efficient method that may replace the gold standard (Watson and Petrie, 2010). Validation of accelerometers compares the behavioral data summarized from the device with visual observations through direct observation (McGowan et al., 2007), video recording (Müller and Schrader, 2003), or by attaching 2 loggers per animal on opposite legs to identify agreement between the 2 devices (Kok et al., 2015). Some studies use a method whereby the performance of 2 different technologies is compared (a validated device is compared with technologies not previously validated; Bewley et al., 2017); however, studies validating devices by comparing 2 different technologies were not considered in this review.

Researchers have used different gold standards, or similar gold standards but different experimental methods, for making comparisons to the gold standard for the validation of accelerometers, which complicates comparison between studies because they are rarely equivalent (Rutten et al., 2013). This issue can result in similar validation studies making different recommendations. For example, a gross overestimation of LB from the IceTag 3D compared with the gold standard (direct observations recorded using a 5-min scan-sampling approach) was reported and it was concluded that this was due to a large majority of false short bouts; therefore, the authors recommended caution when interpreting data from IceTag 3D devices (Rutter et al., 2014). It is worth noting, however, that a LB duration of <5 min is plausible (Kok et al., 2015), and, therefore, the 5-min scan sampling approach used as the gold standard may have resulted in missed LB recordings and an underestimation of LB in the gold standard. In contrast, McGowan et al. (2007) reported 100% agreement when validating the accuracy of LB recorded between the IceTag 3D and the gold standard (continuous direct observations); therefore, the authors recommended the IceTag 3D as both a research and industry tool.

These differences in study recommendations across the same accelerometers under similar experimental conditions may add confusion for researchers interested in using accelerometers. Although direct observation is typically a preferred gold standard, the data set in the

study by McGowan et al. (2007) may have been limited because of a small variation in the LB recorded due to the short data collection period (e.g., across 3 d; ~9.3 h of data was recorded). In particular, if the short data collection period results in a lack of short LB in the test data set, this would make it difficult to validate the device for recording true short LB and as a result may be, in part, responsible for the different findings reported by the 2 aforementioned studies. Therefore, it is important to choose an appropriate gold standard method for comparison with measured data to allow true short LB to be distinguished from false LB. For example, we propose continuous or frequent scan sample (≤ 1 min) observation as the best-practice gold standard. Future validation studies should consider the choice of the gold standard and the variation and size of the test data set (e.g., number of data records) required to validate the device of interest, as these factors can affect the performance reported.

Measures of Behavioral Data Agreement

Several measures of agreement have been used in validation studies to evaluate the accuracy and reliability of behavior data recorded by accelerometers compared with a gold standard. The most common measures of agreement include accuracy, positive and negative predictive values, sensitivity (**Se**) and specificity (**Sp**) measures, Lin's concordance correlation coefficients (ρ_c), Pearson correlation coefficients (r_p), Spearman rank correlation (r_s), and coefficients of determination (R^2 ; Appendix Tables A2 and A3). A range of agreement statistics have been presented in validation studies; however, in some studies inappropriate measures of repeatability are reported (e.g., r_p ; Watson and Petrie, 2010). When comparing studies, the accuracy, Se, Sp, and ρ_c values are the preferred measures of reliability; however, where these measures are not reported, other measures, as outlined above, are presented. To improve the interpretation of reported R^2 , these values are interpreted as recommended by Ledgerwood et al. (2010), where $R^2 \geq 0.90$ and a slope and intercept not statistically different ($P < 0.05$) from 1 and 0, respectively, indicate a high level of accuracy.

Future work should consider whether the agreement statistics presented in validation studies are sufficiently accurate and appropriate to support a reliable assessment of the measure(s) presented.

Measurement Error

Measurement error exists when the observed values differ from the true values due to random and systematic error (Watson and Petrie, 2010). Random error is

considered to be normally distributed and, therefore, these errors tend to balance out on average (Hibbert, 2007). Systematic error can be caused by incorrect interpretation of data, which influences the overall accuracy of the measured data (Watson and Petrie, 2010). Where systematic error and its causes are known, the error can be eliminated or minimized, a correction can be applied to compensate, or an algorithm built into the software to make the adjustment (Hibbert, 2007). The editing criteria, sampling frequency, sampling interval (Charlton et al., 2017), and software used for interpretation of the raw data are major determinants of the accuracy of behaviors recorded by accelerometers and should be considered when making comparisons between validation studies and different devices (Kok et al., 2015), and when determining an appropriate experimental method when using the device in a research setting.

Systematic Error I: False LB

Short LB occur more frequently when an animal is experiencing physical discomfort; for example, during parturition (Huzzey et al., 2005), and these rapid movements must be captured using a higher sampling frequency (e.g., ≤ 2 min; Chen and Bassett, 2005) than is needed for behaviors that animals spend a considerable amount of time engaged in (e.g., lying time; Ledgerwood et al., 2010). A rule of thumb has been proposed that the sampling interval should be at least half that of the highest frequency movement being classified (Chen and Bassett, 2005); therefore, if the interest is in both lying time and LB, a shorter sampling interval (higher resolution) should be selected (e.g., if it is plausible for the animal to engage in a 4-min bout of lying, the sampling interval required to record this behavior accurately would be ≤ 2 min; Chen and Bassett, 2005; Ledgerwood et al., 2010).

A common systematic error that occurs in accelerometers is the generation of false, short lying events in the data (e.g., the accelerometer records short lying events when the cow is standing) and can occur when the animal temporarily shifts its leg position during activities such as grooming or grazing (O'Driscoll et al., 2008). Lying bouts of very short duration should be treated with suspicion and a validated minimum LB criterion applied to remove such errors from the data set that are not biologically meaningful (Mattachini et al., 2013) without removing valuable information that correctly describes the behavior of the animal.

A LB criterion is defined as a set minimum LB duration that the user deems is indicative of true lying behavior and is used to adjust the data set so that erroneous short LB are removed. A LB criterion is based

on ethological considerations of lying behavior that reflects true behavior. In validation studies, the device of interest is tested by applying different LB criteria to the data to discard false LB, across a range of sampling intervals under controlled experimental conditions, to test and identify the optimum sampling interval and editing criteria that will result in the highest level of accuracy for recording lying time and LB behaviors (Ledgerwood et al., 2010). To our knowledge, most accelerometers used for research purposes (e.g., HOBO and IceTag 3D devices) require the user to remove false LB data through data editing rather than relying on algorithms built into software to remove these false LB. Therefore, users should carefully consider applying appropriate LB criterion to their data where the interest is in recording both lying time and LB (Mattachini et al., 2013).

Several studies have evaluated sampling intervals ranging from 1 s to 60 min with either no data editing or editing using a range of LB criteria (e.g., LB ≤ 6 s to ≤ 30 min discarded) to determine the upper and lower limits of accelerometers, where the interest was in accurately recording LB in dairy cows (Table A3). Where the sampling interval was ≥ 10 min, poor agreement has been reported when no LB criteria were applied or when LB < 1 min were discarded from Actiwatch (10-min sampling interval; Müller and Schrader, 2003) and IceTag 2D data (range: 10- to 60-min sampling intervals; Mattachini et al., 2013; Table A3). A range of true LB durations from 4 s to 4 h 20 min have been reported in lactating dairy cows under freestall housing (Kok et al., 2015). Therefore, it is not surprising that low-resolution sampling intervals ranging from 10 to 60 min and those that do not recommend the use of a LB criterion (e.g., McGowan et al., 2007; Felton et al., 2012; Borchers et al., 2016; Nielsen et al., 2018) appear to be inaccurate for recording LB behavior, assuming a true LB duration of 4 s is plausible (Table A3).

Where the sampling interval was 5 min, moderate to poor agreement was reported when no LB criteria were applied or when LB < 5 min were discarded from HOBO and IceTag2D data under a 5-min sampling interval (Ledgerwood et al., 2010; Mattachini et al., 2013; Table A3). Short LB (< 5 min) that do not last the duration of longer sampling intervals (e.g., ≥ 5 min) may provide important information about cow comfort and health (Ledgerwood et al., 2010) but are likely to be missed and the total number of LB substantially underestimated when a 5-min sampling interval is used to record LB (Mattachini et al., 2013). Therefore, the sampling interval and LB criteria commonly recommended tends to be ≤ 4 min to accurately detect LB (Ledgerwood et al., 2010; Tolkamp et al., 2010; Mattachini et al., 2013; Kok et al., 2015; Henriksen and Munksgaard, 2019),

as relatively few LB records have a duration less than 4 min (Table A3); however, true LB durations <4 min are plausible in dairy cows (Kok et al., 2015).

Sampling intervals ranging from 1 s to 2 min have been investigated in dairy cows for recording daily LB and, with the appropriate LB criterion applied to the data, result in the highest levels of accuracy when the device is attached to the hind leg (Table A3). When the sampling interval was 1 min, substantial agreement was reported for the HOBO and AfiAct devices without the removal of false LB (Ledgerwood et al., 2010 and Felton et al., 2012, respectively).

Substantial agreement with improved accuracy was reported across validation studies, when the sampling interval was shorter (e.g., 1 s, 1 min, or 2 min) and false LB were removed from the data using an appropriate LB criterion. Three devices have been validated, with a range of appropriate LB criteria reported, including the HOBO (1-min sampling interval; LB ≤ 1 or 2 min discarded), IceTag 2D (1- or 2-min sampling interval; LB ≤ 2 min discarded), and IceQube devices (1-s sampling interval; LB $\leq 33, 55, 90$, or 148 s discarded; Table A3). Overall, selecting a higher-resolution sampling interval and applying an appropriate LB criterion improves the accuracy of LB recorded.

Systematic Error II: Sampling Frequency and Sampling Interval

A sampling interval contains a summary of the counts of behavior recorded at the device's predetermined sampling frequency (Robert et al., 2009). For example, a sampling frequency of 1 Hz with a sampling interval of 1 min would produce 60 records within each interval (Dow et al., 2009); this indicates the resolution of sampling. Although the sampling interval of a device is usually specific to the product, the sampling interval is typically user-defined in research devices, where a range of sampling interval options (range = 1 s to 1 wk) are available for selection (e.g., IceTag 2D and IceTag 3D), and often fixed in commercial devices (e.g., 2 h, ACT; Lely). In some cases, it is not possible to program the device to further summarize the data into a defined sampling interval, and therefore, the data are summarized according to the selected sampling frequency (e.g., HOBO). The sampling frequency and sampling interval both influence the resolution and amount of data obtained and the accuracy of recorded lying behavior (Mattachini et al., 2013). Therefore, the ability to select the sampling interval should be taken into consideration when choosing a device, depending on the intention of use, as the level of data granularity required will depend on the research question and objectives of the study (Mitlöhner et al., 2001).

At lower sampling frequencies, the device must remain in a vertical or horizontal position for a longer period to register a standing or lying event; however, lower sampling frequencies may be suitable if the interest is in recording durational behavior (e.g., daily lying time; Borchers et al., 2016). Validation studies investigating the accuracy of accelerometers for measuring daily lying time when the sampling interval was ≥ 30 min have indicated poor agreement for the ACT (2-h sampling interval; Elischer et al., 2013) and IceTag 2D devices (30- and 60-min sampling intervals; Mattachini et al., 2013) and moderate agreement in the HOBO (60-min sampling interval; Borchers et al., 2016; Table A2). Using a lower sampling frequency results in fewer data points to estimate total daily lying times, leading to either overestimation or underestimation of lying time (Borchers et al., 2016).

In general, 1- to 15-min sampling intervals are suitable for recording daily lying time, with higher-resolution sampling intervals producing improved results (Table A2). Based on the literature reviewed, 5 devices accurately recorded daily lying time at a 15-min sampling interval including AfiTag II Pedometer (AfiTagII, Afimilk; Henriksen and Munksgaard, 2019), IceTag 2D (Mattachini et al., 2013), CowScout, IceTag 3D (Nielsen et al., 2018), and IceQube devices (Borchers et al., 2016). At a 1-min sampling interval, 4 devices accurately recorded daily lying time, including the IceTag 2D (Mattachini et al., 2013), HOBO (Ledgerwood et al., 2010), TAC, and AfiAct devices (Borchers et al., 2016). All of these validation studies reported these agreement statistics without discarding LB before the analysis, except in the study validating AfiTagII (15-min sampling interval) and IceTag 2D (1-min sampling interval) devices, where LB <10 s and ≤ 2 min, respectively, were discarded when recording daily lying time.

In the aforementioned validation studies (e.g., Ledgerwood et al., 2010; Mattachini et al., 2013; Borchers et al., 2016; Nielsen et al., 2018), the removal of false LB had little effect on total daily lying time, and therefore, these studies did not recommend that a LB criterion be applied to the raw data when the interest is in recording daily lying time only. For example, at a 1-min sampling interval, the results reported by Ledgerwood et al. (2010) indicated that when LB ≤ 2 min were discarded from the HOBO data set, the Se and Sp values improved slightly compared with no editing criterion applied; however, the device was still able to accurately recording lying time irrespective of the editing methodology used. In support of this, Ungar et al. (2018) reported that where IceTag 3D devices (1-s sampling interval) were tested in beef cattle on hilly rangeland and with a large proportion (42%) of false LB recorded ≤ 1 min, the false LB only accounted

for 3% of total lying time. Therefore, it is possible to achieve high levels of agreement, when the interest is in lying time, without applying a LB criterion to the data for sampling intervals from 1 to 15 min.

Overall, the results indicate that sampling intervals ≥ 30 min would not be advisable to accurately record lying time in dairy cows, particularly when the durational totals are summations of infrequent scan samples (Elischer et al., 2013). Unbiased estimates of lying time can be obtained using sampling intervals ≤ 15 min with or without discarding false LB from the data set because the intervals are short enough relative to the duration of the behavior to allow accurate and precise predictions.

MEASUREMENT VARIABILITY

Although variation in behavioral measurements is inevitable, differences in experimental conditions between studies using accelerometers can create additional variation in the data and affect its distribution (Tolkamp et al., 2010). This variation can result in different outcomes and recommendations for the use of the accelerometer, which may cause confusion for researchers when deciding on an appropriate LB criterion (Bewley et al., 2017). Measurement variability exists because behavioral variables are repeated measures within individuals over time and there will always be intra-individual (within individual) as well as inter-individual (between individuals) variability. Variability in lying behavior and the proportion of short LB in data are influenced by cow-, farm-, and management-related factors (Munksgaard et al., 2006; Mattachini et al., 2013).

Different recommendations have been reported even though the same or similar type of accelerometers have been validated in cows of differing ages (e.g., calves and cows; Trénel et al., 2009; Ledgerwood et al., 2010), under different management (e.g., grazing pasture, McGowan et al., 2007; tiestall, Felton et al., 2012; and freestall housing, Kok et al., 2015), different underfoot conditions (e.g., slatted floors and deep bedded; Ledgerwood et al., 2010; Henriksen and Munksgaard, 2019), and different physiological states (e.g., nonlactating and lactating; Ledgerwood et al., 2010). For example, when the interest was in LB, Ledgerwood et al. (2010) reported that at a 30-s sampling interval, a higher LB criterion (≤ 60 s) was recommended for their second experiment (Exp. 2) to achieve a similar level of Se and Sp compared with the ≤ 30 s LB criterion recommended for their first experiment (Exp. 1). It was speculated that different experimental conditions caused these differences due to a greater proportion of short LB in dry cows housed individually on bedded pack in Exp. 2

(23.8% of LB < 5 min) compared with the group of lactating cows housed in a freestall barn in Exp. 1 (1.5% of LB < 5 min). Although it is difficult to disentangle whether a single factor was responsible for influencing the outcome of the study by Ledgerwood et al. (2010), other studies have reported different outcomes in cows housed under different conditions.

Underfoot conditions can affect lying behavior because of the bedding material affecting the angle of the leg when the cow is standing and lying, which could lead to variation in the number of minutes recorded as lying and standing time or result in a larger number of false LB records compared with what was observed (Henriksen and Munksgaard, 2019). The study by Henriksen and Munksgaard (2019) reported that the accuracy of the AfiTagII for recording LB was improved in cows kept on a slatted floor compared with dry cows kept on deep-bedding (positive predictive value = 0.96 vs. 0.85, respectively). Typically, the devices are attached just above the fetlock and rely on the correct orientation of the limb to accurately record time spent standing or lying. These are examples of where the animals, housing, management, and underfoot conditions have affected the accuracy of the device and, therefore, researchers should consider these factors alongside other cow-, farm-, and management-related factors when determining an appropriate LB criterion to remove false LB and applying recommendations to their own work. Ideally, the experimental conditions under which a device was validated should be similar to the experimental conditions under which the device is to be used in prospective studies to ensure the LB criterion is appropriate (Kok et al., 2015).

SUMMARY

The use of accelerometers to record behavior in dairy cows is still evolving, and it is important that authors of both validation studies and experimental publications carefully consider and report methods accurately to allow inter-study comparisons. We have reviewed the studies published to date that have validated accelerometers for measuring lying behavior in dairy cows. Despite the abundance of recently published studies investigating aspects of lying behavior and aspects of behavior affected by cow-, farm-, and management-related factors using accelerometers, methodological issues involving the appropriate use of many accelerometers for the accurate recording of lying behavior still exist. Some validation studies have been able to systematically evaluate the accuracy of a range of devices to determine appropriate sampling frequencies, sampling intervals, and LB criteria under a range of conditions (e.g., Ledgerwood et al., 2010; Mattachini et al., 2013;

Kok et al., 2015), but many have failed to establish a robust methodology for measuring lying behavior in the device of interest (e.g., Müller and Schrader, 2003; Felton et al., 2012; Siegford et al., 2012; Charlton et al., 2017). This review highlights the importance of understanding the relationships between sampling frequency, sampling interval, and LB criteria and the data obtained from accelerometers to make better informed decisions. It reflects how these factors, as well as cow-, farm-, and management-related factors can affect the outcome of validation studies and, therefore, influence recommendations.

Researchers must understand the limitations of validation studies when applying recommendations to their own data and should ensure appropriate application of the accelerometers and carefully consider and report methods accurately to allow inter-study comparisons. At a minimum, the following should be reported: (1) the system under which the animals were managed (e.g., housing, underfoot conditions, feeding); (2) the accelerometers used and how the device was attached to the animal; (3) a description of the method, software, and statistical packages used to classify the raw accelerometer data into specific lying behaviors and to interpret the data; (4) the sampling frequency and sampling interval selected; and (5) the LB criterion applied to remove false positives with reference to validation work. Thorough reporting of methodology will support universal gold standards for the management and analysis of accelerometer data.

Protocols for the validation of accelerometer-derived lying behaviors should be developed to build on those proposed in the current review. Protocols should determine appropriate gold standards and device performance limits similar to those outlined by Kamphuis et al. (2013) for evaluating in-line mastitis detection systems. Similarly, if device performance limits (e.g., minimum Se and Sp values) were set for accelerometers marketed for measuring lying time and LB, this would give the user a high level of confidence in the data recorded. In the future, we recommend that manufacturers provide information regarding the performance limits of the accelerometers marketed for recording lying behavior.

Following the development of robust protocols for validation, more studies are needed to improve the accuracy of data derived from accelerometers, and researchers are encouraged to consider future research to (1) validate accelerometers under conditions outside of the range of previously published validation studies (e.g., different housing, physiological states, underfoot conditions, ages); (2) investigate a range of sampling intervals and LB criteria to determine the optimal sampling intervals and LB criteria for the prediction of

lying behavior; (3) ensure sufficiently sized and varied data sets are collected to achieve sufficient statistical power and report findings using appropriate agreement statistics; and (4) validate accelerometers in grazing dairy cows using criteria comparable to those undertaken in housed cows.

ACKNOWLEDGMENTS

This work was funded from partnership (DRCX1302) between the New Zealand Ministry of Business, Innovation and Employment and New Zealand dairy farmers through DairyNZ Inc. (Hamilton, New Zealand). S. Hendriks is also supported by the Colin Holmes Dairy Scholarship and the Ministry for Primary Industries Postgraduate Scholarship. The authors have not stated any conflicts of interest.








REFERENCES

- Anderson, D. M., R. E. Estell, and A. F. Cibils. 2013. Spatiotemporal cattle data—A plea for protocol standardization. *Positioning* 4:115–136. <https://doi.org/10.4236/pos.2013.41012>.
- Bewley, J. M., M. R. Borchers, K. A. Dolecheck, A. R. Lee, A. E. Stone, and C. M. Truman. 2017. Precision dairy monitoring technology implementation opportunities and challenges. Pages 1251–1264 in *Large Dairy Herd Management*. 3rd ed. D. K. Beede, ed. Am. Dairy Sci. Assoc., Champaign, IL.
- Borchers, M. R., Y. M. Chang, I. C. Tsai, B. A. Wadsworth, and J. M. Bewley. 2016. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. *J. Dairy Sci.* 99:7458–7466. <https://doi.org/10.3168/jds.2015-10843>.
- Brown, D. D., R. Kays, M. Wikelski, R. Wilson, and A. P. Klimley. 2013. Observing the unwatchable through acceleration logging of animal behaviour. *Anim. Biotelem.* 1:20–36. <https://doi.org/10.1186/2050-3385-1-20>.
- Charlton, G. L., C. Gauld, E. Bleach, and S. M. Rutter. 2017. The accuracy of accelerometer-based, leg-mounted sensors for measuring dairy cow locomotion and lying behaviour at pasture. Page 129 in *Proc. Int. Soc. Appl. Ethol. Conf.*, Aarhus, Denmark. Wageningen Academic Publishers, Wageningen, the Netherlands.
- Chen, K. Y., and D. R. Bassett Jr.. 2005. The technology of accelerometer-based activity monitors: Current and future. *Med. Sci. Sports Exerc.* 37(Suppl.):S490–S500. <https://doi.org/10.1249/01.mss.0000185571.49104.82>.
- Darr, M., and W. Epperson. 2009. Embedded sensor technology for real time determination of animal lying time. *Comput. Electron. Agric.* 66:106–111. <https://doi.org/10.1016/j.compag.2009.01.004>.
- Dolecheck, K. A., W. J. Silvia, G. Heersche Jr., Y. M. Chang, D. L. Ray, A. E. Stone, B. A. Wadsworth, and J. M. Bewley. 2015. Behavioral and physiological changes around estrus events identified using multiple automated monitoring technologies. *J. Dairy Sci.* 98:8723–8731. <https://doi.org/10.3168/jds.2015-9645>.
- Dow, C., K. E. Michel, M. Love, and D. C. Brown. 2009. Evaluation of optimal sampling interval for activity monitoring in companion dogs. *Am. J. Vet. Res.* 70:444–448. <https://doi.org/10.2460/ajvr.70.4.444>.
- Elischer, M. F., M. E. Arceo, E. L. Karcher, and J. M. Siegford. 2013. Validating the accuracy of activity and rumination monitor data from dairy cows housed in a pasture-based automatic milking system. *J. Dairy Sci.* 96:6412–6422. <https://doi.org/10.3168/jds.2013-6790>.
- Felton, C. A., M. G. Colazo, P. Ponce-Barajas, C. J. Bench, and D. J. Ambrose. 2012. Dairy cows continuously-housed in tie-stalls failed

- to manifest activity changes during estrus. *Can. J. Anim. Sci.* 92:189–196. <https://doi.org/10.4141/cjas2011-134>.
- Gibbons, J., C. Medrano-Galarza, A. Marie de Passillé, and J. Rushen. 2012. Lying laterality and the effect of IceTag data loggers on lying behaviour of dairy cows. *Appl. Anim. Behav. Sci.* 136:104–107. <https://doi.org/10.1016/j.applanim.2011.12.005>.
- González, L. A., B. J. Tolkamp, M. P. Coffey, A. Ferret, and I. Kyriazakis. 2008. Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. *J. Dairy Sci.* 91:1017–1028. <https://doi.org/10.3168/jds.2007-0530>.
- Halsey, L. G., J. A. Green, R. P. Wilson, and P. B. Frappell. 2009. Accelerometry to estimate energy expenditure during activity: Best practice with data loggers. *Physiol. Biochem. Zool.* 82:396–404. <https://doi.org/10.1086/589815>.
- Henriksen, J. C., and L. Munksgaard. 2019. Validation of AfITagII, a device for automatic measuring of lying behaviour in Holstein and Jersey cows on two different bedding materials. *Animal* 13:617–621. <https://doi.org/10.1017/S1751731118001623>.
- Hibbert, D. B. 2007. Systematic errors in analytical measurement results. *J. Chromatogr. A* 1158:25–32. <https://doi.org/10.1016/j.chroma.2007.03.021>.
- Huzzey, J. M., M. A. G. von Keyserlingk, and D. M. Weary. 2005. Changes in feeding, drinking, and standing behavior of dairy cows during the transition period. *J. Dairy Sci.* 88:2454–2461. [https://doi.org/10.3168/jds.S0022-0302\(05\)72923-4](https://doi.org/10.3168/jds.S0022-0302(05)72923-4).
- IceRobotics Ltd. 2017. Specialist data collection solutions for researchers and educators. Accessed Jan. 20, 2018. http://www.icerobotics.com/wp-content/uploads/2017/05/IceRobotics_Product_Brochure_2017.pdf.
- Itle, A. J., J. M. Huzzey, D. M. Weary, and M. A. G. von Keyserlingk. 2015. Clinical ketosis and standing behavior in transition cows. *J. Dairy Sci.* 98:128–134. <https://doi.org/10.3168/jds.2014-7932>.
- Kamphuis, C., B. Dela Rue, G. Mein, and J. Jago. 2013. Development of protocols to evaluate in-line mastitis-detection systems. *J. Dairy Sci.* 96:4047–4058. <https://doi.org/10.3168/jds.2012-6190>.
- Kok, A., A. T. M. van Kneegsel, C. E. van Middelaar, H. Hogeveen, B. Kemp, and I. J. M. de Boer. 2015. Technical note: Validation of sensor-recorded lying bouts in lactating dairy cows using a 2-sensor approach. *J. Dairy Sci.* 98:7911–7916. <https://doi.org/10.3168/jds.2015-9554>.
- Kokin, E., J. Praks, I. Verrmäe, V. Poikalainen, and M. Vallas. 2014. IceTag3D™ accelerometric device in cattle lameness detection. *Agron. Res. (Tartu)* 12:223–230.
- Ledgerwood, D. N., C. Winckler, and C. B. Tucker. 2010. Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle. *J. Dairy Sci.* 93:5129–5139. <https://doi.org/10.3168/jds.2009-2945>.
- Martiskainen, P., M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen. 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Appl. Anim. Behav. Sci.* 119:32–38. <https://doi.org/10.1016/j.applanim.2009.03.005>.
- Mattachini, G., E. Riva, C. Bisaglia, J. C. A. M. Pompe, and G. Provololo. 2013. Methodology for quantifying the behavioral activity of dairy cows in freestall barns. *J. Anim. Sci.* 91:4899–4907. <https://doi.org/10.2527/jas.2012-5554>.
- McGowan, J. E., C. R. Burke, and J. G. Jago. 2007. Validation of a technology for objectively measuring behaviour in dairy cows and its application for oestrus detection. *Proc. N.Z. Soc. Anim. Prod.* 67:136–142.
- Mitlöchner, F. M., J. L. Morrow-Tesch, S. C. Wilson, J. W. Dailey, and J. J. McGlone. 2001. Behavioral sampling techniques for feedlot cattle. *J. Anim. Sci.* 79:1189–1193. <https://doi.org/10.2527/2001.7951189x>.
- Müller, R., and L. Schrader. 2003. A new method to measure behavioural activity levels in dairy cows. *Appl. Anim. Behav. Sci.* 83:247–258. [https://doi.org/10.1016/S0168-1591\(03\)00141-2](https://doi.org/10.1016/S0168-1591(03)00141-2).
- Munksgaard, L., M. B. Jensen, L. J. Pedersen, S. W. Hansen, and L. Matthews. 2005. Quantifying behavioural priorities—effects of time constraints on behaviour of dairy cows, *Bos taurus*. *Appl. Anim. Behav. Sci.* 92:3–14. <https://doi.org/10.1016/j.applanim.2004.11.005>.
- Munksgaard, L., C. G. Reenen, and R. Boyce. 2006. Automatic monitoring of lying, standing and walking behavior in dairy cattle. *J. Anim. Sci.* 84(Suppl. 1):304. (Abstr.)
- Neave, H. W., J. Lomb, D. M. Weary, S. J. LeBlanc, J. M. Huzzey, and M. A. G. von Keyserlingk. 2018. Behavioral changes before metritis diagnosis in dairy cows. *J. Dairy Sci.* 101:4388–4399. <https://doi.org/10.3168/jds.2017-13078>.
- Nielsen, P. P., I. Fontana, K. H. Sloth, M. Guarino, and H. Blokhuis. 2018. Technical note: Validation and comparison of 2 commercially available activity loggers. *J. Dairy Sci.* 101:5449–5453. <https://doi.org/10.3168/jds.2017-13784>.
- O'Driscoll, K., L. Boyle, and A. Hanlon. 2008. Short communication: A brief note on the validation of a system for recording lying behaviour in dairy cows. *Appl. Anim. Behav. Sci.* 111:195–200. <https://doi.org/10.1016/j.applanim.2007.05.014>.
- Onset. 2019. HOB0 Pendant G Data Logger. Accessed Oct. 14, 2019. <https://www.onsetcomp.com/products/data-loggers/ua-004-64>.
- Robert, B., B. J. White, D. G. Renter, and R. L. Larson. 2009. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Comput. Electron. Agric.* 67:80–84. <https://doi.org/10.1016/j.compag.2009.03.002>.
- Rodríguez-Jimenez, S., K. J. Haerr, E. Trevisi, J. J. Loo, F. C. Cardoso, and J. S. Osorio. 2018. Prepartal standing behavior as a parameter for early detection of postpartal subclinical ketosis associated with inflammation and liver function biomarkers in periparturient dairy cows. *J. Dairy Sci.* 101:8224–8235. <https://doi.org/10.3168/jds.2017-14254>.
- Rutten, C. J., A. G. J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: Sensors to support health management on dairy farms. *J. Dairy Sci.* 96:1928–1952. <https://doi.org/10.3168/jds.2012-6107>.
- Rutter, S. M., N. E. Atkins, and S. A. Birch. 2014. The accuracy of the automatic recording of lying behaviour is affected by whether dairy cows are inside or outside. Page 73 in *Proc. 48th Congr. Int. Soc. Appl. Ethol., Vitoria-Gasteiz, Spain*. Wageningen Academic Publishers, Wageningen, the Netherlands.
- Scheibe, K. M., and C. Gromann. 2006. Application testing of a new three-dimensional acceleration measuring system with wireless data transfer (WAS) for behavior analysis. *Behav. Res. Methods* 38:427–433. <https://doi.org/10.3758/BF03192796>.
- Sepúlveda-Varas, P., D. M. Weary, and M. A. G. von Keyserlingk. 2014. Lying behavior and postpartum health status in grazing dairy cows. *J. Dairy Sci.* 97:6334–6343. <https://doi.org/10.3168/jds.2014-8357>.
- Shepard, E. L. C., R. P. Wilson, L. G. Halsey, F. Quintana, A. Gómez Laich, A. C. Gleiss, N. Liebsch, A. E. Myers, and B. Norman. 2008. Derivation of body motion via appropriate smoothing of acceleration data. *Aquat. Biol.* 4:235–241. <https://doi.org/10.3354/ab00104>.
- Shepley, E., M. Berthelot, and E. Vasseur. 2017. Validation of the ability of a 3D pedometer to accurately determine the number of steps taken by dairy cows when housed in tie-stalls. *Agriculture* 7:53–56. <https://doi.org/10.3390/agriculture7070053>.
- Siefford, J. M., M. F. Elischer, and E. L. Karcher. 2012. The accuracy of activity monitor data from dairy cows housed in a pasture-based automatic milking system. Pages 113–117 in *Proc. Dairy Cattle Welf. Symp., Ontario, Canada*. University of Guelph, Ontario, Canada.
- Tamura, T., Y. Okubo, Y. Deguchi, S. Koshikawa, M. Takahashi, Y. Chida, and K. Okada. 2019. Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers. *Anim. Sci. J.* 90:589–596. <https://doi.org/10.1111/asj.13184>.
- Telezhenko, E., M. A. von Keyserlingk, A. Talebi, and D. M. Weary. 2012. Effect of pen size, group size, and stocking density on activity in freestall-housed dairy cows. *J. Dairy Sci.* 95:3064–3069. <https://doi.org/10.3168/jds.2011-4953>.

- Tolkamp, B. J., M. J. Haskell, F. M. Langford, D. J. Roberts, and C. A. Morgan. 2010. Are cows more likely to lie down the longer they stand? *Appl. Anim. Behav. Sci.* 124:1–10. <https://doi.org/10.1016/j.applanim.2010.02.004>.
- Trénel, P., M. B. Jensen, E. L. Decker, and F. Skjøth. 2009. Technical note: Quantifying and characterizing behavior in dairy calves using the IceTag automatic recording device. *J. Dairy Sci.* 92:3397–3401. <https://doi.org/10.3168/jds.2009-2040>.
- UBC (University of British Columbia). 2013. UBC Animal Welfare Program: SOP-HOBO Data Loggers. Univ. British Columbia, Vancouver, Canada. Accessed Mar. 18, 2018. <http://lfs-awp.sites.olt.ubc.ca/files/2013/11/SOP-HOBO-Datalogger-november-2013.pdf>.
- Ungar, E. D., Y. Nevo, H. Baram, and A. Arieli. 2018. Evaluation of the IceTag leg sensor and its derivative models to predict behaviour, using beef cattle on rangeland. *J. Neurosci. Methods* 300:127–137. <https://doi.org/10.1016/j.jneumeth.2017.06.001>.
- von Keyserlingk, M. A. G., A. Barrientos, K. Ito, E. Galo, and D. M. Weary. 2012. Benchmarking cow comfort on North American freestall dairies: Lameness, leg injuries, lying time, facility design, and management for high-producing Holstein dairy cows. *J. Dairy Sci.* 95:7399–7408. <https://doi.org/10.3168/jds.2012-5807>.
- Watson, P. F., and A. Petrie. 2010. Method agreement analysis: A review of correct methodology. *Theriogenology* 73:1167–1179. <https://doi.org/10.1016/j.theriogenology.2010.01.003>.
- Weary, D. M., J. M. Huzzey, and M. A. G. von Keyserlingk. 2009. Board Invited Review: Using behavior to predict and identify ill health in animals. *J. Anim. Sci.* 87:770–777. <https://doi.org/10.2527/jas.2008-1297>.

ORCIDS

- S. J. Hendriks  <https://orcid.org/0000-0002-6725-9290>
 C. V. C. Phyn  <https://orcid.org/0000-0002-4912-4069>
 J. M. Huzzey  <https://orcid.org/0000-0003-1644-6608>
 K. R. Mueller  <https://orcid.org/0000-0001-7023-5356>
 S-A. Turner  <https://orcid.org/0000-0003-2654-0673>
 D. J. Donaghy  <https://orcid.org/0000-0002-3002-5773>
 J. R. Roche  <https://orcid.org/0000-0002-4165-9253>

APPENDIX

Table A1. Description of validation studies undertaken to determine the accuracy of various commercially available activity monitors typically used in research for estimating lying time and lying bouts in dairy cows

Source; data collection period	n (cow) ¹	Subject ²	System ³	Device, direction; attachment ⁴	Behavior validated ⁵	Gold standard ⁶
Müller and Schrader (2003); 5 d	12	L HF	FS; TMR	Actiwatch OD; HL	LB and LT	Video
Munksgaard et al. (2006); not specified	12	Not specified	Not specified	IceTag 3D; HL	LT	Direct
McGowan et al. (2007); 3 d (~9.3 h total)	15	D Primi Multi H	P PA; G	IceTag 3D; HL	LB and LT	Direct
Ledgerwood et al. (2010; Exp. 1); 1.5 d	12	HF J	FS LS Y; TMR	HOB0 3D; HL	LB and LT	Video
Ledgerwood et al. (2010; Exp. 2); 1.5 d	12	L HF J	DB I Y; TMR	HOB0 3D; HL	LB and LT	Video
Felton et al. (2012); 6 d	19	L Multi H	I P S TS; TMR	AfiAct 3D; HL	LB and LT	Video
Siefford et al. (2012); Elischer et al. (2013); 15 d	15	L Primi Multi H	FS P; TMR	ACT; NC IceQube 3D; HL	LT	Direct
Mattachini et al. (2013); 1 d	5	L Multi HF	LS; TMR	HOB0 3D; HL	LB and LT	Video
Mattachini et al. (2013); 1 d	12	L Multi HF	LS; TMR	IceTag 2D; HL	LB and LT	Video
Rutter et al. (2014); 8 h	16	L HF	LS P	IceTag 3D	LB and LT	Direct
Kok et al. (2015); 6 d	28	L Multi	C FS; TMR	IceQube 3D; HL	LB	Logger
Borchers et al. (2016); 8 d (4 h/d)	48	L Primi Multi H	C FS Y; TMR	AfiAct 3D; HOB0 3D, IceQube 3D; HL TAC 3D; FL	LT	Direct
Charlton et al. (2017); 2 h	48	L HF	P	IceQube 3D; FL HL	LB and LT	Direct
Henriksen and Munksgaard (2019); 2 h	40	D Primi Multi H J	DB FS LS SL; PMR	AfiTagII 3D; HL	LB and LT	Direct
Nielsen et al. (2018); 19.8 h	6	L Primi Multi HF	FS	IceTag 3D; HL	LT	Video
Nielsen et al. (2018); 29.5 h	10	L Primi Multi HF	FS	CowScout 3D; FL	LT	Video

¹n (cow) describes the number of subjects that were enrolled in the study.

²Subject describes the physiological state, parity, and breed of the subjects, respectively. If either of these 3 phenotypes is not specified, then it was not described in the paper. Physiological state; D = dry; L = lactating; Parity; Primi = primiparous; Multi = multiparous; Breed; H = Holstein, HF = Holstein-Friesian, J = Jersey.

³System describes the management system used. Housing and underfoot conditions; C = concrete underfoot, DB = deep bedded underfoot, FS = freestall barn, I = individually housed, LS = loose-housed, P = pasture underfoot, PA = pasture access, S = sand underfoot, SL = slatted floor underfoot, TS = tiestall, Y = yard access. Feeding; G = grazing, PMR = partial mixed ration. If any of these systems is not specified, then it was not described in the paper.

⁴Device, direction; attachment describes the name of the activity monitor tested, the directions in which signals are received, and where the device was attached to the animal. Device: Actiwatch = Actiwatch Activity Monitoring System (Cambridge Neurotechnology, Fenstanton, Cambridgeshire, UK), AfiAct = AfiMilk Pedometer Plus and AfiTagII = AfiTag II Pedometer (AfiMilk, Kibbutz, Afikim, Israel), ACT = Automatic Milking System activity monitor (Lely, Maassluis, the Netherlands), HOB0 = HOB0 Pendant G Acceleration Data Logger (Onset Computer Corp., Pocasset, MA), IceTag 3D = IceTag 3D Activity Monitor, IceTag 2D = IceTag 2.004 Activity Monitor (ceased manufacture in 2008) and IceQube = IceQube Activity Monitor (IceRobotics Ltd., Edinburgh, UK), CowScout = CowScout Leg Sensor (GEA Farm Technologies, Bönen, Germany), TAC = The Track a Cow (ENGs, Rosh Pina, Israel). Direction: 2D = two-dimensional, 3D = three-dimensional, OD = omnidirectional. Attachment; FL = front leg, HL = hind leg, NC = neck-worn collar. If the attachment is not specified, then it was not described in the paper.

⁵Behavior validated describes the behaviors validated in the study. LB = lying bouts, LT = lying time.

⁶Gold standard describes the type of observation or device that the device being tested was compared with. Direct = behavior recorded by an observer, Logger = behavior recorded by a logger on the opposite leg of the test device, Video = behavior recorded using video monitoring and transcribed retrospectively.

Table A2. Summary of reported outcomes for a range of accelerometers validated since 2003 for lying time in dairy cows

Source	Device ¹	Sampling interval ²	LB criterion ³	Reliability ⁴	Se ⁵	Sp ⁶
Müller and Schrader (2003)	Actiwatch	10 min	<1 min	$r_s = 0.65^{***}$	—	—
Munksgaard et al. (2006)	IceTag 3D	1 min	—	$r_p = 0.99^{NR}$	—	—
McGowan et al. (2007)	IceTag 3D	1 min	—	—	—	—
Ledgerwood et al. (2010; Exp. 1)	HOB0	6 s	—	$R^2 > 0.99$	99.3	99.8
	HOB0	30 s	—	$R^2 > 0.99$	99.3	99.7
	HOB0	1 min	—	$R^2 > 0.99$	99.3	99.8
	HOB0	5 min	—	$R^2 > 0.98$	99.3	99.8
	HOB0	6 s	≤ 6 s	$R^2 > 0.99$	99.8	99.9
	HOB0	30 s	≤ 30 s	$R^2 > 0.99$	99.8	99.9
	HOB0	1 min	≤ 1 min	$R^2 > 0.99$	99.7	99.9
	HOB0	5 min	≤ 5 min	$R^2 > 0.98$	98.8	99.5
	HOB0	6 s	≤ 12 s	$R^2 > 0.99$	99.9	100
	HOB0	30 s	≤ 1 min	$R^2 > 0.99$	99.7	99.9
	HOB0	1 min	≤ 2 min	$R^2 > 0.99$	99.6	99.9
	HOB0	5 min	≤ 10 min	$R^2 > 0.98^*$	97.1	98.9
	HOB0	6 s	≤ 36 s	$R^2 > 0.99$	99.8	99.9
	HOB0	30 s	≤ 3 min	$R^2 > 0.99$	99.5	99.8
	HOB0	1 min	≤ 6 min	$R^2 > 0.99$	98.9	99.5
	HOB0	5 min	≤ 30 min	$R^2 \leq 0.92$	91.4	95.5
	HOB0	6 s	—	$R^2 > 0.99^*$	97.6	99.4
	HOB0	30 s	—	$R^2 > 0.99^*$	97.6	99.5
	HOB0	1 min	—	$R^2 > 0.99^*$	97.4	99.4
	HOB0	5 min	—	$R^2 > 0.98^*$	96.2	99.0
	HOB0	6 s	≤ 6 s	$R^2 > 0.99$	99.2	99.8
	HOB0	30 s	≤ 30 s	$R^2 > 0.99$	99.4	99.9
	HOB0	1 min	≤ 1 min	$R^2 > 0.99$	99.3	99.8
	HOB0	5 min	≤ 5 min	$R^2 > 0.98$	97.1	99.3
	HOB0	6 s	≤ 12 s	$R^2 > 0.99$	99.4	99.9
	HOB0	30 s	≤ 1 min	$R^2 > 0.99$	99.4	99.9
	HOB0	1 min	≤ 2 min	$R^2 > 0.99$	99.0	99.8
	HOB0	5 min	≤ 10 min	$R^2 > 0.98^*$	95.3	98.9
	HOB0	6 s	≤ 36 s	$R^2 > 0.99$	99.5	99.9
	HOB0	30 s	≤ 3 min	$R^2 > 0.99$	98.8	99.7
	HOB0	1 min	≤ 6 min	$R^2 > 0.99$	97.9	99.6
	HOB0	5 min	≤ 30 min	$R^2 > 0.85$	82.3	95.5
Felton et al. (2012)	AfiAct	1 min	—	$\rho_c = 0.89$; $r_p = 0.91^*$	—	—
Siegford et al. (2012); Elischer et al. (2013)	IceQube	15 min	—	$r_p = 0.97^{***}$; $R^2 = 0.85$	—	—
Mattachini et al. (2013)	ACT	2 h	—	$r_p = -0.57^{**}$	—	—
	HOB0	1 min	≤ 2 min	—	99.0	>99.0
Mattachini et al. (2013)	IceTag 2D	1 s	≤ 25 s	—	99.7	100
	IceTag 2D	1 min	≤ 2 min	—	99.7	100
	IceTag 2D	1 min	≤ 2 min	$R^2 = 1.00$	—	—
	IceTag 2D	2 min	≤ 2 min	$R^2 = 1.00$	—	—
	IceTag 2D	3 min	—	$R^2 > 0.99$	—	—
	IceTag 2D	4 min	—	$R^2 > 0.99$	—	—
	IceTag 2D	5 min	—	$R^2 > 0.99$	—	—
	IceTag 2D	10 min	—	$R^2 > 0.99$	—	—
	IceTag 2D	15 min	—	$R^2 = 0.98$	—	—
	IceTag 2D	30 min	—	$R^2 = 0.96^*$	—	—
	IceTag 2D	60 min	—	$R^2 = 0.84^*$	—	—
Rutter et al. (2014)	IceTag 3D	5 min	—	—	—	—
Borchers et al. (2016)	IceQube	15 min	—	$\rho_c > 0.99$; $r_p > 0.99^{**}$	—	—
	HOB0	15 min	—	$\rho_c = 0.81-0.87$; $r_p = 0.83-0.87^{**}$	—	—
	HOB0	60 min	—	$\rho_c = 0.92$; $r_p = 0.93^{**}$	—	—
	TAC	1 min	—	$\rho_c > 0.99$; $r_p = 0.99^{**}$	—	—
	AfiAct	1 min	—	$\rho_c > 0.99$; $r_p = 0.99^{**}$	—	—
	AfiTagII	15 min	<3 min	$r_p = 0.98$	—	—

Continued

Table A2 (Continued). Summary of reported outcomes for a range of accelerometers validated since 2003 for lying time in dairy cows

Source	Device ¹	Sampling interval ²	LB criterion ³	Reliability ⁴	Se ⁵	Sp ⁶
Nielsen et al. (2018)	IceTag 3D	15 min	—	$\rho_c > 0.99$	—	—
	CowScout	15 min	—	$\rho_c > 0.99$	—	—

¹Device: Actiwatch = The Actiwatch Activity Monitoring System (Cambridge Neurotechnology, Cambridgeshire, UK), AfiAct = AfiAct Pedometer Plus and AfiTagII = AfiTag II Pedometer (Afimilk, Kibbutz, Afikim, Israel), ACT = Automatic Milking System activity monitor (Lely, Maassluis, the Netherlands), HOBO = HOBO Pendant G Acceleration Data Logger (Onset Computer Corporation, Pocasset, MA), IceTag 3D = IceTag 3D Activity Monitor, IceTag 2D = IceTag 2.004 Activity Monitor (ceased manufacture 2008) and IceQube = IceQube Activity Monitor (IceRobotics Ltd., Edinburgh, Scotland), CowScout = CowScout Leg Sensor (GEA Farm Technologies, Bönen, Germany), TAC = The Track a Cow (ENGs, Rosh Pina, Israel).

²Sampling interval describes time between samples of the device and within each sampling interval is a summary of all registered counts of data recorded at a predetermined sampling frequency.

³Lying bout (LB) criterion describes the minimum LB duration that is considered a true bout of lying and $LB \leq$ the LB criterion are removed from the data to discard false LB before interpretation. If no LB criterion is presented, the study did not apply a LB criterion to the data.

⁴Reliability is described by the following agreement statistics: ρ_c = concordance correlation coefficient, R^2 = coefficient of determination. ^{NR}Statistical significance was not reported; *slope and intercept were significantly different ($P < 0.05$) from 1 and 0, respectively; **slope and intercept were significantly different ($P < 0.01$) from 1 and 0, respectively. If no superscript is presented, the slope and intercept were not different from 1 and 0. r_p = Pearson correlation coefficient (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$), r_s = Spearman rank correlation (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). If no reliability measures are presented, the study did not present reliability measures.

⁵Se = sensitivity.

⁶Sp = specificity.

Table A3. Summary of reported outcomes for a range of accelerometers validated since 2003 for lying bouts in dairy cows

Source	Device ¹	Sampling interval ²	LB criterion ³	Reliability ⁴	Se ⁵	Sp ⁶
Müller and Schrader (2003)	Actiwatch	10 min	<1 min	$r_p = 0.80^{**}$	—	—
McGowan et al. (2007)	IceTag 3D	1 min	—	—	—	—
Ledgerwood et al. (2010; Exp. 1)	HOBO	6 s	—	$R^2 < 0.25$	99.3	99.8
	HOBO	30 s	—	$R^2 < 0.60$	99.3	99.7
	HOBO	1 min	—	$R^2 < 0.70$	99.3	99.8
	HOBO	5 min	—	$R^2 \geq 0.92$	99.3	99.8
	HOBO	6 s	≤ 6 s	$R^2 < 0.80$	99.8	99.9
	HOBO	30 s	≤ 30 s	$R^2 \geq 0.93$	99.8	99.9
	HOBO	1 min	≤ 1 min	$R^2 \geq 0.93$	99.7	99.9
	HOBO	5 min	≤ 5 min	$R^2 > 0.90$	98.8	99.5
	HOBO	6 s	≤ 12 s	$R^2 < 0.80$	99.9	100
	HOBO	30 s	≤ 1 min	$R^2 \geq 0.93$	99.7	99.9
	HOBO	1 min	≤ 2 min	$R^2 \geq 0.93$	99.6	99.9
	HOBO	5 min	≤ 10 min	$R^2 > 0.90$	97.1	98.9
	HOBO	6 s	≤ 36 s	$R^2 \geq 0.97$	99.8	99.9
	HOBO	30 s	≤ 3 min	$R^2 > 0.90$	99.5	99.8
	HOBO	1 min	≤ 6 min	$R^2 > 0.90$	98.9	99.5
	HOBO	5 min	≤ 30 min	$R^2 > 0.90^*$	91.4	95.5
	HOBO	6 s	—	$R^2 < 0.30$	97.6	99.4
Ledgerwood et al. (2010; Exp. 2)	HOBO	30 s	—	$R^2 < 0.50$	97.6	99.5
	HOBO	1 min	—	$R^2 < 0.60$	97.4	99.4
	HOBO	5 min	—	$R^2 \geq 0.92^*$	96.2	99.0
	HOBO	6 s	≤ 6 s	$R^2 < 0.60$	99.2	99.8
	HOBO	30 s	≤ 30 s	$R^2 \geq 0.93$	99.4	99.9
	HOBO	1 min	≤ 1 min	$R^2 \geq 0.93^*$	99.3	99.8
	HOBO	5 min	≤ 5 min	$R^2 < 0.70^*$	97.1	99.3
	HOBO	6 s	≤ 12 s	$R^2 < 0.80$	99.4	99.9
	HOBO	30 s	≤ 1 min	$R^2 \geq 0.93$	99.4	99.9
	HOBO	1 min	≤ 2 min	$R^2 \geq 0.93^*$	99.0	99.8
	HOBO	5 min	≤ 10 min	$R^2 < 0.60^*$	95.3	98.9
	HOBO	6 s	≤ 36 s	$R^2 \geq 0.97$	99.5	99.9
	HOBO	30 s	≤ 3 min	$R^2 < 0.90^*$	98.8	99.7
	HOBO	1 min	≤ 6 min	$R^2 < 0.80^*$	97.9	99.6
	HOBO	5 min	≤ 30 min	$R^2 < 0.10^*$	82.3	95.5
Felton et al. (2012)	AfiAct	1 min	—	$\rho_c = 0.94$; $r_p = 0.94^*$	—	—
Mattachini et al. (2013)	IceTag 2D	1 min	≤ 2 min	$R^2 = 0.98$	—	—
	IceTag 2D	2 min	≤ 2 min	$R^2 = 0.92$	—	—
	IceTag 2D	3 min	—	$R^2 = 0.89$	—	—
	IceTag 2D	4 min	—	$R^2 = 0.83$	—	—
	IceTag 2D	5 min	—	$R^2 = 0.82$	—	—
Mattachini et al. (2013)	IceTag 2D	10 min	—	$R^2 = 0.72$	—	—
	IceTag 2D	15 min	—	$R^2 = 0.59$	—	—
	IceTag 2D	30 min	—	$R^2 = 0.36$	—	—
	IceTag 2D	60 min	—	$R^2 = 0.05^*$	—	—
Rutter et al. (2014)	IceTag 3D	5 min	—	—	—	—
Kok et al. (2015)	IceQube	1 s	≤ 1 s	Acc = 0.96	100	0.00
	IceQube	1 s	≤ 4 s	Acc = 0.96	100	14.1
	IceQube	1 s	≤ 33 s	Acc = 0.99	99.3	97.7
	IceQube	1 s	≤ 55 s	Acc = 0.99	98.9	100
	IceQube	1 s	≤ 90 s	Acc = 0.99	98.4	100
	IceQube	1 s	≤ 148 s	Acc = 0.98	97.5	100
	IceQube	1 s	≤ 245 s	Acc = 0.97	96.7	100
Charlton et al. (2017)	IceQube	1 s	—	$R^2 = 0.76^{***}$	—	—
Henriksen and Munksgaard (2019)	AfiTagII	1 min	<3 min	$R^2 = 0.97^{**}$	—	—

¹Device: Actiwatch = The Actiwatch Activity Monitoring System (Cambridge Neurotechnology, Fenstanton, Cambridgeshire, UK), AfiAct = AfiAct Pedometer Plus and AfiTagII = AfiTag II Pedometer (Afirmilk, Kibbutz, Afikim, Israel), ACT = Automatic Milking System activity monitor (Lely, Maassluis, the Netherlands), HOBO = HOBO Pendant G Acceleration Data Logger (Onset Computer Corp., Pocasset, MA), IceTag 3D = IceTag 3D Activity Monitor, IceTag 2D = IceTag 2.004 Activity Monitor (ceased manufacture 2008) and IceQube = IceQube Activity Monitor (IceRobotics Ltd., Edinburgh, UK), CowScout = CowScout Leg Sensor (GEA Farm Technologies, Bönen, Germany), TAC = The Track a Cow (ENGs, Rosh Pina, Israel).

²Sampling interval describes time between samples of the device and within each sampling interval is a summary of all registered counts of data recorded at a predetermined sampling frequency.

³Lying bout (LB) criterion describes the minimum LB duration that is considered a true bout of lying and LB \leq the LB criterion are removed from the data to discard false LB before interpretation. If no LB criterion is presented, the study did not apply a LB criterion to the data.

⁴Reliability is described by the following agreement statistics: ρ_c = concordance correlation coefficient, R^2 = coefficient of determination. Acc = accuracy (defined as the sum of correctly discarded false LB records and correctly retained true LB records divided by the total amount of LB records). *Slope and intercept were significantly different ($P < 0.05$) from 1 and 0, respectively; **slope and intercept were significantly different ($P < 0.01$) from 1 and 0, respectively. If no superscript is presented, the slope and intercept were not different from 1 and 0. r_p = Pearson correlation coefficient (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$), r_s = Spearman rank correlation (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). If no reliability measures are presented, the study did not present reliability measures.

⁵Se = sensitivity.

⁶Sp = specificity.