Horizon 2020 Programme

INFRAIA-02-2017
Integrating Activities for Starting Communities

SmartCow: an integrated infrastructure for increased research capability and innovation in the European cattle sector

an integrated infrastructure for increased research capability and innovation in the European cattle sector

Project ID: 730924

Deliverable number: D7.1

Deliverable title : Guidelines for validation of sensor output

EC version : V2

Due date of milestone	**31/01/2019 (M12)**
Actual submission date	**06/06/2019 (M17)**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°730924

DOCUMENT INFO

1. Author(s)

Organisation name lead contractor	**Aarhus University**

Author	Organisation	e-mail
Lene Munksgaard	**Aarhus University**	**Lene.munksgaard@anis.au.dk**

2. Revision history

Version	Date	Modified by	Comments
1	04.04.02019		
2	06.05.2019		

3. Dissemination level

PU	Public	X
CO	Confidential , only for members of the consortium (including the Commission Services)	☐

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

EXECUTIVE SUMMARY

Background	*At SmartCow RIs a large amount of data are recorded using new smart technologies on a routine basis or for experimental purposes. However, there is a lack of a common approach for validation of such devise.*
Objectives	*The objective of Task 7.1 is to develop and test uniform guidelines for validation of outputs from sensors for the recording of animal behaviour to be used within the SmartCow infrastructure.*
Methods	Meetings between by AU, WU, IRTA, INRA was held to build consensus on how to validate data from sensors and draft guidelines. Then the draft of the guidelines was sent to all other SmartCow partners who use sensors for recording of behaviour for commenting. This will include outputs from sensors available in the different RIs measuring eating, grazing, lying, standing, walking, steps, activity, rumination, and position of the animals. Furthermore, the draft of the report was discussed at the first annual meeting at a workshop.
Results & implications	A report has been produced including a checklist to consult before a study to validate a sensor for measuring cattle behaviour. The chapters give an overview of the main steps in a validations study that needs consideration and description in the report of the validation. Many new sensors for automatic measuring of cattle behaviour are developed and new devices are introduced in these years using different technologies. Therefore, we chose to make a checklist rather than a detailed description on how to validate specific sensor output. The guideline will be published at the SmartCow webpage and may be included in the book of guidelines to be produced in WP3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

Table of contents

SmartCow

an integrated infrastructure for increased research capability and innovation in the European cattle sector

A checklist to validate sensor output for the recording of cattle behaviour

Bouchon[1)], M., Bach[2)], A., Meunier[1)], B., Ternman[3)], E., Van Reenen[4)], K., Veissier[5)], I., Munksgaard[3)], L.

[1)] Experimental Unit Herbipole*, INRA, France* [2)]*, Department of Ruminant Production, IRTA, Spain* [3)]*, Department of Animal Science, Aarhus University, Denmark* [4)]*, Wageningen University* [5)]*, The Nederland's, Joint Research Group of Herbivores, INRA, France*

1 Introduction

A process of validation assesses the appropriateness and usefulness of a tool for its intended purpose within a specific context. Ideally, the validation of a tool should describe the range of purposes and contexts in which it is appropriate. This generally cannot be done completely. Therefore, when such a wide validation cannot be done, the validation process needs to refer clearly to the purpose of the use of the tool and to which animals it is going to be applied.

The technological development has led to an increase in the number of sensors and devices that are available for measuring cattle behaviour. The technical solutions behind the sensors can differ substantially, sensors use accelerometer technology, sound recording or image analysis to mention some, and even within each solution, there is a large amount of factors that can influence the quality of the sensor's information output or battery capacity. The raw data from the sensor is usually translated through algorithms to estimate the behaviours in question and such algorithms may vary in the way they work. There might be data points for individual animals or whole groups missing for some periods. The attachment of the sensor itself sometimes may affect the animals' behaviour. Furthermore, some sensors are developed for use on commercial farms and others for research use only, with differences in precision and accuracy. A careful validation of the outputs from sensors for measuring cattle behaviour is required before use, and the validation needs to be related to the purpose i.e. whether it is the behaviour of the cows or the status of the cows that is in focus (figure 1). To be able to compare the output from different devices across facilities a standardized validation protocol would be helpful. Therefore, this report discusses important factors that can affect the outcome of a validation.

For instance: Small ruminants vs cattle: be aware that a system that has been validated for small ruminants may not be valid for cattle, and vice-versa. Indeed the anatomy of the animal is already a difference that should be considered when testing a behaviour sensor.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

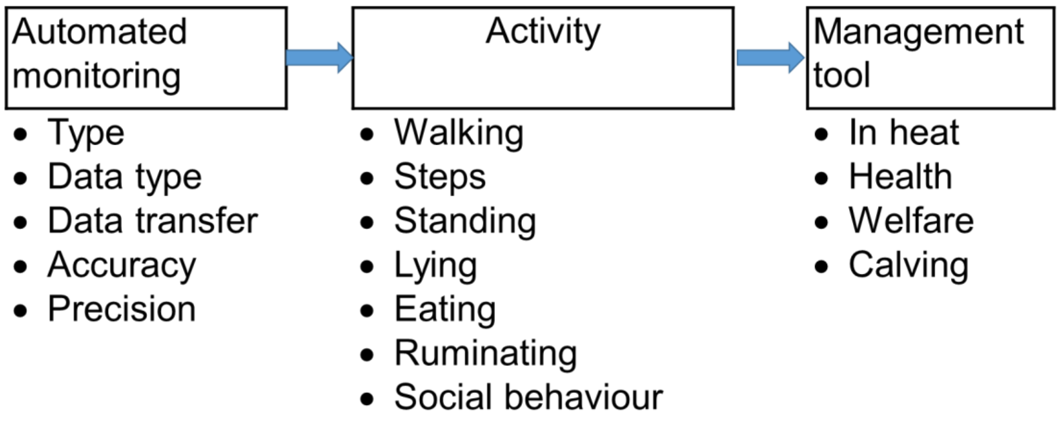

Figure 1. Illustration of the pathways automatic recording of behaviour and the use as a management tool (draft).

This guideline is developed in the SmartCow project as collaboration between the partners in WP 7 with input from other partners in the project working with devices that automatically measure behaviour. Many new sensors for automatic measuring of cattle behaviour are developed and new devices regularly introduced in these years using different technologies. Therefore, this paper is a checklist to consult before the onset of a study to validate a sensor for measuring cattle behaviour rather than a detailed description on how to validate specific sensor output. The following chapters provide an overview of the main steps in a validations study that needs consideration and description in the report of the validation. Some of the issues are relevant not only for validation studies but also for any type of behavioural studies.

2 Description of the equipment used

In order to compare the outcome of validation studies it is important that a proper description of the equipment in use be provided. A description should include minimum requirements, described in the two following paragraphs (sensor technology and data from the sensor), and non-exhaustive additional information, which can be added if relevant to interpret the dataset provided by the device. Criteria retained to describe devices have been chosen as a result of literature review of validation studies. We will illustrate most of the retained items by examples, taken from the literature or built by our own when nothing relevant was found.

2.1 Sensor technology

Some basic information is needed to describe the sensor. This information is mainly provided by the manufacturer and easy to catch. It can help understanding how the device works and what implication might have on data acquisition and/or animal behaviour.

(i) Type of sensor technology

This information is the most important to describe the sensor under study. Technological solutions vary between sensors, for example accelerometer, Real Time Location System (RTLS), Global Positioning System (GPS), weighing scales, thus it is important to be precise in the description of the technology.

(ii) Commercial Name/Company/Version of the Sensor

The aim is to give information about the exact model/version of the sensor used, to easily identify it and make it findable for other applications. This is information given by the manufacturer.

Example: "The acceleration sensor used was the ADXL335 (Analog Devices, One Technology Way, P.O. Box 9106, Norwood, MA 02062–9106, USA), a complete tri-axial accelerometer" (Giovanetti et al. 2017) *(iii) Sensor weight and size, mounting system and recoverability*

The interaction between the sensor and the natural behaviour of the animal may affect the behaviour and the sensor output. Indeed, a small lightweight accelerometer mounted on a collar does not have the same impact on cow behaviour as a large and heavy GPS device fixed on a halter. From an ethical and economic point of view, it also seems relevant to describe the recoverability of the sensor, e.g. an intra-ruminal bolus that cannot be taken back from the animal may need specific authorization, as it might be considered an invasive measurement.

Example 1: Moreau et al. (2009) used GPS tracking collar on goats. "The components were placed in a sturdy casing and mounted on a polyester harness adjustable to the animal's neck. Gross mass of the collar was 610g." Meunier et al. (2017) equipped dairy cows with "a counter-weighed collar (565g) on which active tag (6x5x4cm – 146g) was mounted". On these two examples, we can easily assess that 600g on a goat (which represent 1% of its BW), has not the same impact than 600g on a dairy cow (which represent approximately 0.1% of its BW).

Example 2: "Rumen pH was monitored continuously throughout the experiment using a commercial sensor […] immersed in the reticulum through the oesophagus with the use of dedicated balling gun." (Villot et al. 2017)

(iv) Memory capabilities and battery life

This point aims at ensuring the device will work well during the validation trial. It has to be expressed in terms of duration. Special attention will be paid to the adequacy between the duration of the study and these two criteria: memory capacity and battery life. Additionally, technology used to store data and battery technology may be described.

Example 1: "Onset Pendant G data loggers […] were set to record […] at a 6-s interval until the memory storage in the data loggers was full (i.e., approximately 1.5 d)" (Ledgerwood *et al.* 2010)

Example 2: "The duration of recording varies between 1.5 and 681.3 days (e.g. 45 days at a sample-interval of 1min)" Müller and Schrader (2009)

Example 3: "Such batteries would need to operate properly and autonomously for long periods of time (e.g., five years) without being recharged or replaced" (Benaissa *et al.* 2017)

(v) Sampling rate

If known, sampling rate could be useful in understanding how the measurement is done and/or processed (see below). It has to be expressed in terms of frequency or period (e.g. in Hz or points per minutes). For instance, to estimate the speed of a moving leg by accelerometer data will demand a much higher sampling rate than to estimate whether an animal is lying or standing.

Example 1: "The collars were programmed to collect GPS data at 4Hz (i.e. 345000 data points/day) and accelerometer data at 10 Hz (i.e. 862500 data points/day)" (Gonzalez *et al.* 2015)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

(vi) *Data storage/transmission*

The description of the technology used to transmit data (e.g. wireless transmission such as Bluetooth, Wi-Fi or Radiofrequency, or manual data transmission) is relevant because it can have an influence on cow behaviour: data that are automatically transmitted do not need intervention on the animal, whereas data manually downloaded (USB, SD-card) may disturb the animal. On the other hand, wireless data often entail a loss of data. If data loss occurs, it would be good to quantify and specify when and how much data are missing.

Example 1: On MSR Rumiwatch pressure noseband, used to register chewing or ruminating behaviour, data are stored on a SD-card (Rombach *et al.* 2018, Werner *et al.* 2017). If there is a need to collect the data during the trial, it could create a disturbance in cow behaviour, as the halter has to be unmounted to download the SD-card through a computer.

Example 2: "The collar-mounted accelerometers […] measurements were gathered once a second using wireless data acquisition network […]. Data packets were then sent to a PC […]." (Martiskainen *et al.* 2009)

Example 3: "Data were downloaded every 15 days using eCow handset (smartphone + antenna) with eCow android application." (Villot *et al.* 2017)

2.2 Data output and handling

A second step to describe the sensor consists in describing the data it provides, as every system has its own way to process, deliver, present, aggregate etc… the data.

(i) *Nature and type of data collected*

The nature of the data obtained using sensor varies: acceleration, voltage, weight, inclination, position, duration… Moreover, some sensors can provide already processed data, like activity levels or direct behaviour (e.g. ingestion or rumination), or simply filtered or averaged values. Thus, a proper description of the nature (including units) and the type (raw vs processed) of data is required. Data processing will be described later.

Example 1: "The data recorded by the GPS included the spatial position of the animal on the earth-centered earth-fixed coordinate system (X-, Y- and Z-axis coordinates in cm) the time as in the GPS system, and the 3-dimensional travel speed (cm/s)" (Gonzalez *et al.* 2015)

Example 2: "Five variables were recorded by the accelerometers over each epoch: average acceleration in each of the three axes (XavgG, YavgG, ZavgG), Vector Magnitude Average (VMavgG), and Vector Magnitude Max (VMmaxG). Average axis acceleration (XavgG, YavgG, ZavgG) and Vector Magnitude Average (VMavgG) were calculated by summing the inputs (100samples/s) and dividing by the specified reporting interval (3, 5, or 10s). Vector Magnitude (formula (1)) is calculated by

$$\text{Vector Magnitude} = \sqrt{XavgG^2 + YavgG^2 + ZavgG^2} \qquad (1)$$

The VMmaxG is simply the highest combined axis instance per reporting interval" (Robert *et al.* 2009)

(ii) *Data rate and timestamp/dating*

Data rate can be rather different to sampling rate, mainly in cases where data are processed. For instance data may be expressed as steps per 15 minutes even though the accelerometer data is sampled for instance with 12 HZ. It can be expressed in terms of frequency, epochs etc. A precise description of time- and date-related information is also necessary, as it has repercussions on data analysis, and especially in data comparison with gold standard.

Indeed, data can be summarized in UTC (Coordinated Universal Time), GMT±h (Greenwich Mean Time) or collected with a non-synchronized clock (e.g. chronometer and sensor synchronized at the beginning of the trial).

Example 1: “The clocks of the observer, the video recording system, and the sensors were synchronized at the start and at the end of the observation period so that observation data could be aligned accurately with the tri-axial accelerometer data retrieved from the sensors.” (Benaissa *et al.* 2017)

Example 2: “The device is connected to Ethernet to ensure a perfect clock synchronization using the Network Time Protocol (NTP), dating in the epoch time system” (Meunier *et al.* 2017)

(iii) Data processing

Some pieces of information are helpful to interpret the dataset. The point here is not to describe precisely the algorithm used, but more to describe its impact on the dataset. Data can be acquired by a single-shot measurement, but it can also be averaged, filtered, or processed by an algorithm, and thus it can introduce a delay, for example data presentation that has to be characterized.

Example 1: Axel ® Medria normalized dataset present data obtained by the identification of the main behaviour registered during the previous 5 minutes. For example, if the dataset indicates that the cow was grazing at 0815h, it means that between 0810h and 0815h, the algorithm has calculated that the cow was mainly grazing during this period.

Example 2: Various filter (e.g. Kalman filter, median filter) can be applied to clean the data provided by RTLS, as Pastell *et al.* 2017 does, leading to a reduced data output or data capturing frequency.

(iv) Data calibration

Some systems need to be calibrated regularly, in order to improve the matching between the physical measurement and the generated data. This calibration must be stated in frequency (e.g. monthly).

Example: weighing scales, like those used in BioControl weighing trough, which basically measure voltages and convert it into weights, needs to be calibrated at least monthly to ensure that there is no deviation into the converting process.

For some sensors, a “training” period can also be necessary. Devices might indeed require an auto-calibration or recalibration period, e.g. an accelerometer need to “find” earth acceleration to determine the real acceleration direction it records.

Example 1: CowView RTLS used fixed tags to auto-evaluate its performance (Meunier *et al.* 2017, Pastell *et al.* 2018, Veissier *et al*. 2017)

Example 2: “It records both dynamic accelerations, related to changes in the movements of the sheep and static accelerations (–9.8 m s^{-2}) caused by earth's gravity” (Giovanetti *et al*. 2017)

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement N°730924

Manufacturers often provide this information, and a validation study needs to respect at least these minimum requirements.

(v) Theoretical accuracy, resolution and range of measurement

Manufacturers can provide information on the precision of the measurement made by the sensor. When it is available, it is relevant to present to characterize what can be expected from the device in use. It can also provide useful information in the discussion of the results obtained from the validation trial.

Example 1: CowView RTLS system is given to have an accuracy of 50cm in detecting the position of an individual in the barn (Tullo *et al.* 2016), based on the capabilities of its electronic. Meunier *et al.* (2017) shown that in real condition, it could be better, with a 16cm precision to detect a precise activity (eating salt).

Example 2: "The accelerometer measured a range of ±3.2 g with an accuracy of 0.075 g" (Chapinal *et al.* 2011)

3 Test environment

This part will provide a description of the important factors to be aware of both from an animal behaviour point of view and from a technology point of view.

3.1 Environment for the animal

The environment of the animals can have important impact on the results. Ideally, the range of environments in which the tool will provide valid results should be described. Because this can hardly be done, the tool should at least be validated for later use in the same type of environments. A large number of environmental factors can affect the behaviour of cattle, and since many of the sensors in use are measuring the movement of different parts of the body of the animal, it is important to keep the animals in an environment that allows for the movements that later on are in focus. Furthermore, since direct observations or video recordings very often are used as gold standard the environment should allow for proper observations of the behaviour of the animals. The animals have to be habituated to the environment before any data collection. Thus a clear description of housing (including size, floor type, bedding, cubicle type, feeding area etc.), milking, group size, climate conditions etc. should be included in the materials and methods section of a validation study.

Example 1: "Cows were housed in free stalls with mattress and sawdust bedding and concrete slatted floors. They were milked twice daily and supplied with fixed amounts of concentrates and ad libitum roughage." (Kok *et al.* 2015).

Example 2: "Lactating cows were kept in a loose-house system with slatted floor and cubicles with mattresses (Comfi Cushion, with permeable rubber, Egtved, Denmark). The facility was equipped with a free cow traffic automatic milking system (DeLaval, Tumba, Sweden) with an automatic milking unit for each section. Dry cows were kept in pens with deep bedding in the lying area and with slatted floors in front of the feeders." (Henriksen & Munksgaard, 2018).

3.2 Environment for the sensor

There is often limitation to which environment the sensor works properly. From a technological point of view, it is important to consider that some environmental condition can influence the quality of the dataset. Some information can be relevant to assess that pitfalls have to avoided or taken into account during the validation study.

Respect of environmental condition of use of the sensor and its electronics is important. As every electronic device, sensors are made to operate under given environmental conditions. The sensor can be sensitive to physical disturbance (e.g. altitude can influence the measurement made by an accelerometer as it modifies its reference acceleration due to a different gravity). The on-board electronic is also designed to run under appropriate climate conditions (e.g. electronic components are given to run in a precise range of temperature, clock and sensor's precision are often deviating with temperature). This information is often provided by the manufacturer too. It may also imply that a record of the environmental conditions during the study should be recorded and reported.

Example: Medria recommend to use Axel ® accelerometers between -20°C and +55°C, at an altitude lower than 2000m asl.

Some information can be relevant to assess that pitfalls have been avoided or taken into account during the validation trial. Here are some examples of pitfalls that have been avoided before the experiment, or discovered *a posteriori*, based on users' feedback.

Example 1: when a cow is lying on one side, is the CowView sensors masked by the cow posture or by the cubicle structure? When cow is in the milking parlour, which is not covered by the antennas, what activity is associated with?

Example2: accelerometer interprets feeding behaviour using the inclination of cow head. In case cows are fed with an automatic concentrate feeder, the angle can be insufficient to detect a feeding behaviour when the animal eats concentrate. Thus, using accelerometer with such feeding systems can lead to poor results in the validation process. This can also be the case with animals grazing in high grass or animals fed on a weighing trough.

Example 3: gyroscope based sensors are very sensitive to metallic environment and motions recorded with such devices can be false, whereas UWB signals are absorbed by wood, which can lead to a distortion of the signal.

Example 4: MSR Rumiwatch pressure noseband is very sensitive to low temperature and altitude. As they work with propylene glycol, there is a need to ensure that its density is stable and that it is in a liquid state.

In every case encountered, it is essential to observe and annotate abnormal events to allow fully discussing the results.

Example 1: D'Andrea et al. (2017) encountered issues of misclassified behaviours due to prolonged activity of body scratching by one cow with the leg equipped of a pedometer.

Example 2: feeding visit at the weighing trough could be allocated to the wrong animal due to a smart "animal thief", able to eat over the gate.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

4 Ethic and need for permission

Remember to get the permissions needed for the handling and care of the animals included in the validation, and considered whether the impact on the animals' welfare can be justified. These considerations should also include facts about the effects of long term uses (see chapter 9).

5 Animals, feed and water

The type of animal (breed, physiological status, age, size,…) as well as feeding and access to water can have important impact on the results. Ideally, the range of animal types for which the tool will provide valid results should be described. Because this can hardly be done, the tool should at least be validated for later use with the same type of animals. Therefore, the description of the animals provided in the validation needs to include a minimum of information about the animals and their feeding.

Animals and feeding

Several animal characteristics to consider during the validation process will help for the interpretation of results:

(i) Breed: defining if sensors are tested in dairy and beef ruminants is important since they have different routines that may entail differences in the validation process.

(ii) Physiological status: animals will behave different depending of their physiological status (i.e., lactating or not, high vs low producing milk cows, unhealthy animals). Therefore, including animals in different physiological stages or defining them will be required for the validation process.

(iii) Animal age: young and adult ruminants have different sizes and routines that may alter the results of the validation or the adaptability of the sensor to the animal. Thus, young and adult animals will require different validations processes.

> Example 1: Large, heavy animals may have different patterns of movements than smaller animals.
>
> Example 2: Calves may run and jump more than older animals; thus, a device for measuring movement may not show the same results on calves and cows.

(iv) Diet: the type of feeds animals have access to (i.e., TMR, hay, pasture, composition, quality, temperature) and feeding management (i.e., ad libitum, restricted, frequency, stocking density) should be described to help the interpretation of the validation process.

> Example 1: A sensor that measures eating at pasture may not measure eating in a barn due to differences in the movements of cows when eating.

For number of animals and number of observations to include in the study see chapter 8 (Sample size).

6 Position of the sensor on the animal and need for a habituation period

The position of the sensor on the animal is very important, because in many cases, it is highly correlated with the measurement, and thus with the studied behaviour. A complete description of the position of the sensor should include at least:

- Anatomic indication (e.g. retro-mandibular microphone, leg- or neck-mounted accelerometer, Reticulum pH boluses, ear-tag…)
- Symmetry (e.g. right/left, front/rear) if relevant

Example 1: "accelerometers were attached to the lateral face of the rear leg, just proximal to fetlock" (Robert *et al.* 2009)

In some case, it could be meaningful to describe the orientation of the sensor. For example, with an accelerometer, indicating where does x-, y- and z-axis point can be essential to understand the data.

Example 1: "During the experiment, the bold end of the logger always pointed towards the animal's head and the truck ridge pointed to the top. Thus the instrument's X-axis corresponded to the vertical dimension, the Y-axis corresponded to one horizontal dimension, measuring the acceleration sidewise to the left and right, and the Z-axis corresponded to the second horizontal dimension measuring the acceleration forwards and backwards." (Moreau *et al.* 2009)

In general, several questions have to be addressed while choosing how to equip the animal with a particular sensor, and could help in describing and justifying how the sensor is placed on the animal:

- What sensor for what position and for what behaviour?

Example 1: Benaissa *et al.* (2017) showed that leg-mounted accelerometers are more accurate than neck-mounted accelerometer to detect lying behaviour, whereas it is the opposite in identifying feeding behaviour.

Example 2: For RTLS tag, "the dorsal position was chosen to minimize possible shielding of the signal by the body of the cow." (Pastell *et al.* 2017)

- Is there a risk of negative interaction if several sensors are placed on the same mounting system? Does the sensor measure the animal or the device (i.e., collar) behaviour? How to ensure a stable position of the sensor over time?

Example 1: CowView weighted-collars are designed to receive a CowView tag, and ensure its positioning on the top of the cow neck. What would happen if we had a retro-mandibular microphone on this collar? Would the RTLS tag stay at the right place?

Example 2: Some author showed that for certain sensors, tightly or loosely mounted collars have no significant impact on the quality of the dataset provided by an accelerometer (Oudshoorn *et al.* 2013)

- Is there a known risk of injury by the sensor if placed in a certain position/on certain mounting system? How can it influence the natural behaviour of the animal?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

Example1: RumiWatch halter is less adapted to young bulls. They may over-express scratching behaviour because the halter embarrasses them.

Lastly, the user needs to describe whether an habituation period is required by the animal (do not confuse with the training period of the sensor, as previously). This can be particularly relevant in cases that mounting the sensor affects animal behaviour for a few days.

Example 1: for pedometers, D'Andrea *et al*. (2017) applied a 2-day habituation period, allowing cows to return to a normal activity. Rombach *et al*. (2018) proceed similarly with Rumiwatch halters, which needed 4 days of habituation for the animal.

Example 2: While giving access to feed into a weighing trough, cows may need a few days to learn the gate access process.

7 How to measure/record the "gold standard"

The gold standard is the best estimate of the true behaviour performed or true status of the animal that is used to validate the accuracy, precision etc. of the output of the sensor. The quality of the gold standard is therefore very important for the outcome and sets the limit for the quality of the validation. However, it may be difficult to get gold standards that are correct /true for some type of behaviours. It is relatively easy to define whether a cow is standing or lying, but determining from either direct observation or video recordings whether a cow is walking, standing, or even taking a step can be much more complicated. Feeding behaviour is another example, feeding can be defined as when a cow is chewing or it can also be defined as when a cow has the head above the feeder. The choice of definition of the gold standard may have a dramatic impact on the outcome of validation. Furthermore, sometimes other types of equipment may provide better gold standard than observations of the behaviour. For instance may brush rotation provide a better estimate of grooming towards a brush than direct observation of the behaviour? Therefore, providing a careful description on how the gold standard is obtained is important.
Furthermore, there are a number of other factors to consider:

(i) Description of the behaviour. How do the behaviours in question look like, include drawings or pictures of the behaviour /position of the animal will often improve the clarity.

(ii) The choice of data sampling method, duration etc. should match the question that is addressed. If the question is whether a specific sensor measures rumination per day observations of ruminations per day may be enough, and here minor deviations in when rumination start and stop may not be important. However, if the question is: can the output from this sensor tell me exactly when a cow starts/stops to ruminate, then exact recordings of rumination including start and stop time stamps are needed.

(iii) The duration of the sampling the gold standard. In which time windows should data be collected and for how long on each animal.

(iv) Synchronization of time between recording "gold standard data" and sensor output.

(v) Intra- and inter-observer repeatability should be addressed. Well-trained observers are needed; and it is important to check intra and inter observer repeatability and report that along with the results.

8 Sample size and analysis

A validation of the output of a sensor or device to measure automatically a behaviour of cows should include both information about how accurate is the measure (how close to the true value) and how precise is the measure (will repeated measures show the same?). Furthermore, there might be a variation in the accuracy and precision depending on the level of the response (e.g. accuracy in the measures of short lying periods may be larger than accuracy in longer lying bouts). Most validations include an estimate of the average accuracy and precision in terms of an expression of the agreement between the gold standard and the automatic measure (see for instance Borchers et al. (2016). However, it will be more useful if information about the variation at different response levels is reported (see figure 2).
Furthermore, since we are dealing with live animals, the accuracy and precision may also fluctuate due to variation between animals e.g. the same sensor may work really well on some cows but not on other cows due to for instance slight differences in the position of the sensor. Therefore it is not a simple task to choose how many animals and how many repeated measures per animal should be included in the validation (se also chapter 5 and 7).

In order to choose the sample size, consider the following:

(i) How are the data distributed and what is the highest and lowest values that you wish to estimate using the equipment

(ii) Do the manufacture suggest a lowest and highest value the device can measure and what is the resolution?

(iii) Decide based on (i) and (ii) which interval the validation should cover and the resolution of the gold standard data.

(iv) Decide whether the validation should give an estimate of the average agreement or better also report the variation in the agreement over the interval the validation shall include

(v) Based on (i) or (ii) and best estimate of the variation in the response both within and between animals at a given response level perform appropriate power analysis.

It is beyond this guideline to give exact information about power analysis, since this may vary depending on the type of data and the questions that the validation shall address and the methods of analysis. We therefore suggest that you either contact a statistician or study the literature about power analysis.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924

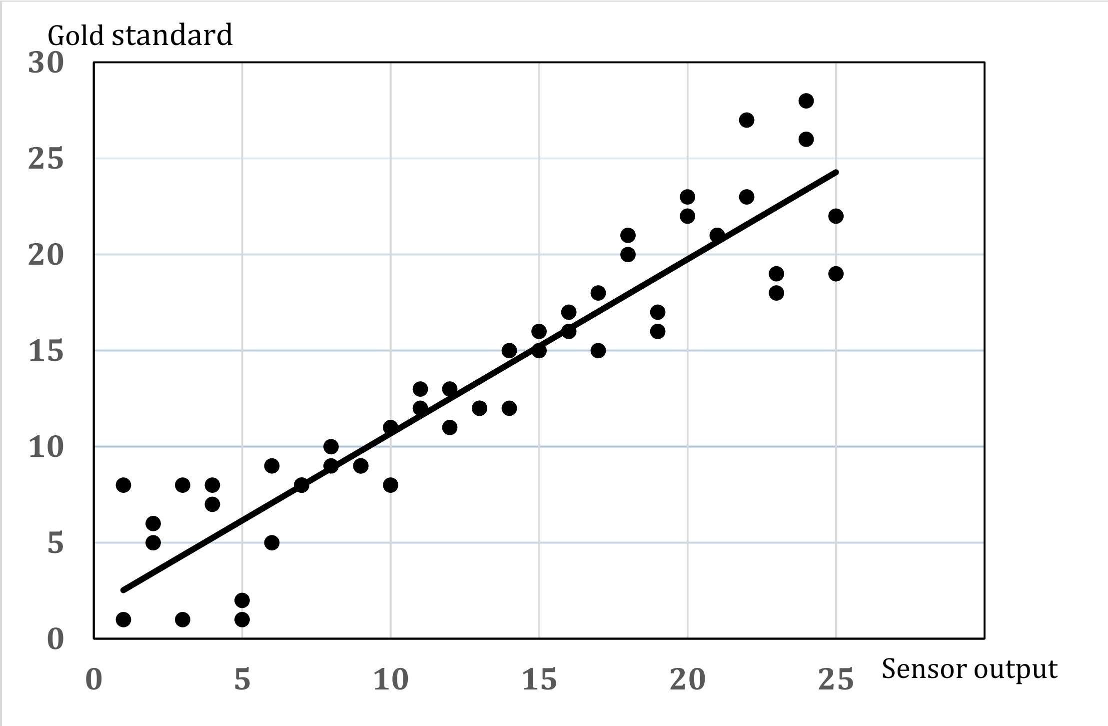

Figure 2. Illustration of the relationship between sensor data (X-axis) and gold standard (Y- axis), and the variation in the accuracy depending on the response level.

Analysis of data

Analysis of the data should give information about the agreement between the gold standard and the behaviour recorded with the sensor (e.g. the sensor data should predict the behaviour of the animal). Simple correlation coefficients can be misleading, and correlations do not provide information about the deviation from the gold standard (Bland and Altman, 1986). Linear regression analysis should provide a confidence interval. An intercept of zero and a slope close to one will indicate a good agreement, and a proper confidence interval can show the variation at different response levels. The Bland Altman method (Bland & Altman, 1995, 2007) has been used widely in medical research, for more information see also Hartnack (2014). However, the most appropriate choice of analysis depends on the type of data (e.g. distribution, continuous or categorical). However, whether this method can be used depend on whether the assumptions are fulfilled.

(i) Plot and check the quality of the data , provide a clear description of data that eventually are omitted from the dataset and any other pre-handling of both gold standard data and automatically collected data

(ii) Check the distribution of the data

(iii) Check the assumptions for choice of statistical model are fulfilled

9 Robustness of the system in use

Besides the actual validation of the output from a specific device, it will be helpful for other users to have as much information as possible on the following:

(i) Effects on the animals. Any injuries on the animals derived from the devices depending on how and where on the animal the sensor is attached, and for how long time the sensors is attached to the animal determines for how long a specific sensor can be used.

(ii) Drift in output over time. Some sensors may show a drift in for instance the timestamp or the actual output from the sensor. This type of information should be provided by the manufacture but, in some instances it may be lacking, thus any information about this or approach to overcome the drift will help other users.

(iii) Need for calibration. This type of information should be provided by the manufacture but, in some occasions, it may be lacking, thus any information about this will help other user.

(iv) Maintenance needs of the sensor and the system.

10 Conclusion

First of all the validation process needs to refer clearly to the purpose of the use of the tool and to which animals it is going to be applied. Further, careful considerations about the equipment used, the test environment from both a technical and animal point of view and how the sensor is attached to the animal should be included. The choice of gold standard is very important and shall reflect the purpose of the purpose of the use of the tool. The sample size should be large enough to report the agreement between the sensor output and gold standard over the interval that the validation covers rather than an average agreement.

11 Acknowledgement

The authors thank Sadjad Danesh Mesgaran; Bjoern Kuhla from FBN Leibniz – Institute for Farm Animal Biology, Germany and René Baumont from INRA, France for valuable comments on earlier versions of the report.
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement n°730924.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement N°730924

12 References

Altman, D.G. and J.M. Bland. 1983. Measurement in Medicine: The Analysis of Method of comparison studies. Journal of the Royal Statistical Society. Series D (The Statistician) 32:307-17.

Benaissa, S., Tuyttens, F.A.M., Plets, D., de Pessemier, T., Trogh, J., Tanghe, E., Martens, L., Vandaele, L., Van Nuffel, A., Joseph, W., Sonck, B., 2017. On the use of on-cow accelerometers for the classification of behaviours in dairy barns. Research in Veterinary Science. https://doi.org/10.1016/j.rvsc.2017.10.005

Bland, J.M. and Altman, D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurements. Lancet (London, England) 1:307.

Bland JM, Altman DG (1995) Comparing methods ofmeasurement: why plotting difference against standardmethod is misleading. Lancet 346, 1085–1087.

Bland JM, Altman DG (2007) Agreement between methodsof measurement with multiple observations perindividual. J Biopharm Stat 17, 571–582.

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.

Chapinal, N., de Passillé, A.M., Pastell, M., Hänninen, L., Munksgaard, L., Rushen, J., 2011. Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle. Journal of Dairy Science 94, 2895–2901. https://doi.org/10.3168/jds.2010-3882

D'Andrea, L., Guccione, J., Alsaaod, M., Deiss, R., Di Loria, A., Steiner, A., Ciaramella, P., 2017. Validation of a pedometer algorithm as a tool for evaluation of locomotor behaviour in dairy Mediterranean buffalo. Journal of Dairy Research 84, 391–394. https://doi.org/10.1017/S0022029917000668

Giovanetti, V., Decandia, M., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., Cossu, R., Serra, M.G., Manca, C., Rassu, S.P.G., Dimauro, C., 2017. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer. Livestock Science 196, 42–48. https://doi.org/10.1016/j.livsci.2016.12.011

González, L.A., Bishop-Hurley, G.J., Handcock, R.N., Crossman, C., 2015. Behavioral classification of data from collars containing motion sensors in grazing cattle. Computers and Electronics in Agriculture 110, 91–102. https://doi.org/10.1016/j.compag.2014.10.018

Hartnack, S. 2014.Issues and pitfalls in method comparison studies. Veterinary Anaesthesia and Analgesia. 41:227-241.

Henriksen, J.C,and L. Munksgaard. 2018. Validation of AfiTagII, a device for automatic measuring of lying behaviour in Holstein and Jersey cows on two different bedding materials. Animal (in press). doi:10.1017/S1751731118001623

Kok, A, van Knegsel, A.T.M. van Middelaar, C.E. Hogeveen, H. Kemp, B. 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 http://dx.doi.org/10.3168/jds.2015-9554

Ledgerwood, D.N., Winckler, C., Tucker, C.B., 2010. Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle. Journal of Dairy Science 93, 5129–5139. https://doi.org/10.3168/jds.2009-2945

Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M., Mononen, J., 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Applied Animal Behaviour Science 119, 32–38. https://doi.org/10.1016/j.applanim.2009.03.005

Meunier, B., Pradel, P., Sloth, K.H., Cirié, C., Delval, E., Mialon, M.M., Veissier, I., 2018. Image analysis to refine measurements of dairy cow behaviour from a real-time location system. Biosystems Engineering 173, 32–44. https://doi.org/10.1016/j.biosystemseng.2017.08.019

Moreau, M., Siebert, S., Buerkert, A., Schlecht, E., 2009. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. Applied Animal Behaviour Science 119, 158–170. https://doi.org/10.1016/j.applanim.2009.04.008

Müller, R., Schrader, L., 2003. A new method to measure behavioural activity levels in dairy cows. Applied Animal Behaviour Science 83, 247–258. https://doi.org/10.1016/S0168-1591(03)00141-2

Oudshoorn, F.W., Cornou, C., Hellwing, A.L.F., Hansen, H.H., Munksgaard, L., Lund, P., Kristensen, T., 2013. Estimation of grass intake on pasture for dairy cows using tightly and loosely mounted di- and tri-axial accelerometers combined with bite count. Computers and Electronics in Agriculture 99, 227–235. https://doi.org/10.1016/j.compag.2013.09.013

Pastell, M., Frondelius, L., Järvinen, M., Backman, J., 2018. Filtering methods to improve the accuracy of indoor positioning data for dairy cows. Biosystems Engineering 169, 22–31. https://doi.org/10.1016/j.biosystemseng.2018.01.008

Robert, B., White, B.J., Renter, D.G., Larson, R.L., 2009. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. Computers and Electronics in Agriculture 67, 80–84. https://doi.org/10.1016/j.compag.2009.03.002

Rombach, M., Münger, A., Niederhauser, J., Südekum, K.-H., Schori, F., 2018. Evaluation and validation of an automatic jaw movement recorder (RumiWatch) for ingestive and rumination behaviors of dairy cows during grazing and supplementation. Journal of Dairy Science 101, 2463–2475. https://doi.org/10.3168/jds.2016-12305

Tullo, E., Fontana, I., Gottardo, D., Sloth, K.H., Guarino, M., 2016. Technical note: Validation of a commercial system for the continuous and automated monitoring of dairy cow activity. Journal of Dairy Science 99, 7489–7494. https://doi.org/10.3168/jds.2016-11014

Veissier, I., Mialon, M.-M., Sloth, K.H., 2017. Short communication: Early modification of the circadian organization of cow activity in relation to disease or estrus. Journal of Dairy Science 100, 3969–3974. https://doi.org/10.3168/jds.2016-11853

Villot, C., Meunier, B., Bodin, J., Martin, C., Silberberg, M., 2018. Relative reticulo-rumen pH indicators for subacute ruminal acidosis detection in dairy cows. animal 12, 481–490. https://doi.org/10.1017/S1751731117001677

Werner, J., Leso, L., Umstatter, C., Niederhauser, J., Kennedy, E., Geoghegan, A., Shalloo, L., Schick, M., O'Brien, B., 2018. Evaluation of the RumiWatchSystem for measuring grazing behaviour of cows. Journal of Neuroscience Methods 300, 138–146. https://doi.org/10.1016/j.jneumeth.2017.08.022.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement N°730924