Repeatability and reproducibility of transrectal pelvic measurements with the Rice-pelvimeter in South African beef cattle measured by experienced and unexperienced observers

L. L. Algie - 3514773
Begeleiders: J.C.M. Vernooij (MSc) - Faculty of Veterinary Medicine, Utrecht University
Prof D.E. Holm – Faculty of veterinary Science, University of Pretoria

Content

Abstract

The repeatability and reproducibility of pelvis measurements with the Rice-pelvimeter of four South African beef cattle done by an experienced and unexperienced was subject of this research. Three observers, one experienced veterinarian and two master veterinary science students, measured the pelvis height and width of 106 animals twice. Comparing these measurements within and between observer established the repeatability and reproducibility for the pelvis measurements. The experienced observer had a difference between the first and second measurement -0,17 cm for the horizontal pelvic measurements (95% limits of agreement -1.37 – 1.03) and -0.06 cm for the vertical pelvic measurements (95% limits of agreement -1.16 – 1.04). The unexperienced observer had a bias of -0.21 cm (95% limits of agreement -1.95 – 1.53) and -0.01 cm (95% limits of agreement -1.59 – 1.56) for the horizontal pelvic measurements and a bias of -0.12 cm (95% limits of agreement -1.77 – 1.53) and -0.07 cm (95% limits of agreement-2.27 – 1.61) for the vertical pelvic measurements. The repeatability and reproducibility were influenced by the observers and the different breeds. We concluded that the Rice-pelvimeter is reliable measurements tool when used by an experienced observer and less reliable when used by an unexperienced observer.

Introduction

Dystocia is a well-known topic due to its welfare and economic impact for both cattle and farmers (Funnell & Hilton, 2016; Price & Wiltbank, 1978a). A device like the Rice-pelvimeter could be used for future on-farm culling of cattle, in order to prevent as many cases of dystocia as possible(Holm, Webb, & Thompson, 2014). Dystocia with cattle heifers of 2 years old is mainly the result of disproportion between the pelvic area and the calf size and weight (Hickson, Morris, Kenyon, & Lopez-Villalobos, 2006; Kolkman et al., 2009; Price & Wiltbank, 1978a). The pelvis area of the dam is considered to be the most important maternal factor for dystocia.(Murray, Cartwright, Downham, Murray, & De Kruif, 2002) For a maximized production of a beef cow, for economic profit it is preferable that she calves for the first time by the time she reaches the age of two years(Price & Wiltbank, 1978b). On the other hand at the age of two years a dam isn't fully grown yet which increases the risk of dystocia with 19.2% in comparison to three year old dams(Price & Wiltbank, 1978b).

In previous studies done by Johnson et al (1988), Basarab et al. (1993) and Van Donkersgoed et al. (1990, 1993) there has been different results when it comes to predicting dystocia by using the measurements of the pelvic width and height. Johnson et al. (1988) predicted dystocia correctly in 66.7% by measuring the pelvic width and height of heifers with a Krautmann pelvic meter, all measurements were made by one experienced person. (Johnson, Deutscher, & Parkhurst, 1988). Basarab et al. (1993) used three different methods to calculate a predicted dystocia percentage in heifers by measurering pelvic width and heights with a Krautmann pelvic meter, the measurements were done by local veterinarians(Basarab, Rutter, & Day, 1993). The pelvic width and height were multiplied to calculate the pelvic area (PA), which was then used as a part of the different ratios that were used for predicting dystocia. Deutscher et al. (1989) used the ratio PA / calf birth weight (BWT), but the BWT is only known after birth. Basarab et al. (1993) tested three different ratios to predict dystocia with heifers using de PA so he could divide the group of heifers into a group which were predicted to have dystocia and a group which were predicted not to have any birth problems. Method 1 predicted dystocia right in 40% of the casus, method 2 in 78.5% of the cases, method 3 in 79.4% of the cases(Basarab et al., 1993). They concluded that method 1 was not a usefull on-farm tool for predicting dystocia, in method 2 and 3 86.0% and 85.7% respectively of the predicted dystocia were eventually easy calvers therefore they concluded that both these methods were also not usefull as an on-farm culling tool(Basarab et al., 1993). Van Donkersgoed et al. (1990) used the Rice-pelvimeter to measure the pelvic height and width at prebreeding and at pregnancy check, the measurements were done by one person. The heifers were predicted to have dystocia when their PA belonged to the 25% smallest pelvis areas (Van Donkersgoed, Ribble, Townsend, & Janzen, 1990). The positive predicted value (PPV) and sensitivity of the measurements taken prebreeding were respectively 27.5% and 22.06%, measurements taken at pregnancy check gave a PPV of 37.5% and a sensitivity of 37.75%, they concluded that the pelvic area could not be used as an on-farm tool for predicting dystocia with heifers(Van Donkersgoed et al., 1990).

As noted by van Donkersgoed et al. (1990) the prediction of dystocia by using the PA could be influenced by the reliability of the horizontal and vertical measurements taken by a pelvimeter. Reliability of the measurements with a pelvimeter is dependent on the repeatability, being the reliability of multiple measurements within one operator, and reproducibility also called the method agreement, being the reliability of measurements between different operators.

Short et al. (1979) reported a correlation coefficient for the repeatability within observers with a range between 0.87 and 0.92 when a small number of heifers (N=10) were measured twice by an observer. The correlation coefficient for the reproducibility between three observers ranged between 0.73 and 0.86(Short et al., 1979). When a larger group of heifers (N=45) was measured the correlation coefficient for the repeatability was 0.61 and for the reproducibility 0.87 which is lower than with a lower number of heifers measured (Short et al., 1979).

Van Donkersgoed et al. (1993) examined the reliability between and within experienced veterinarians of pelvic measurements (width and height) done with the Rice-pelvimeter and the Krautmann pelvimeter. For each heifer, the PA was calculated. Three different PA cut-off points were used for dividing the heifers in a predicted dystocia and a non-dystocia predicted group. The different PA cut-off points were the standard cut-off point (140 cm^2 at prebreeding, 180 cm^2 at pregnancy check, 220 cm^2 at precalving), the 10% heifers with the smallest PA (lower 10%) and the 25% heifers with the smallest PA (lower 25%)(Donkersgoed et al., 1993). The measured heifers were divided into two groups, a positive group being the heifers with a PA below cut-off point and a negative group being the heifers with a PA above cut-off point(Donkersgoed et al., 1993). By dividing the heifers into two groups the researchers used the Cohen's kappa coefficient as a measure of agreement to analyse the repeatability and reproducibility (Donkersgoed et al., 1993; Petrie & Watson, 2013). The Rice-pelvimeter had a Cohen's kappa coefficient for the repeatability of veterinarian 1 of 0.53 (standard and lower 25% cut-off point) and 0.63 (lower 10% cut-off point)(Donkersgoed et al., 1993). The repeatability of veterinarian 2 had a Cohen's kappa coefficient of 0.74 (lower 25% cut-off point) and 1.00 (lower 10% cut-off point) the repeatability for the standard cut-off point could not be calculated because there were no PA's below the cut-off point. The Cohen's kappa coefficient for the reproducibility was -0.09 (standard cut-off point), 0.39 (lower 25% cut-off point) and 0.43 (lower 10% cut-off point)(Donkersgoed et al., 1993). Their conclusion therefor was that the Rice-pelvimeter was a poor on-farm test because the repeatability and reproducibility was low to moderate (Donkersgoed et al., 1993).

Kolkman et al. (2009) measured pelvic height and width with the Rice-pelvimeter two hours before slaughter and compared those measurements with measurement made with a ruler 2h after slaughter when the carcass was split in half. The measurements on the live animal were taken three times, an epidural anaesthesia was used during the measurements and the average was used for comparison with the measurement obtained from the carcasses(Kolkman et al., 2009). All the measurements were taken by the same observer, the experience of the observer in pelvimetry wasn't noted(Kolkman et al., 2009). The difference between the mean pelvic width obtained before slaughter and the pelvic width from the carcass was -0.2 cm (95% limits of agreement are -2.5 cm to 2.1 cm)(Kolkman et al., 2009).The bias for the pelvis height was 1.2 cm (95% limits of agreement are -1.8 cm to 4.1 cm)(Kolkman et al., 2009). The Pearson's correlation coefficient between the mean measurement before slaughter and the measurements of the carcass were r = 0.56 (P<0,001) and r = 0.46 (P<0,001) for respectively pelvic width and pelvic height(Kolkman et al., 2009).

The research of Kolkman et al. (2009) was done with Belgian Blue cattle, the research of Van Donkersgoed et al. (1993) was done by different beef cattle breeds in the United States and Short et al. (1979) in crossbreeds Angus X Hereford (AH), AH X Charolai and AH X Simmental.

As noted by Kolkman et al. (2009) the breed is of influence on the pelvic conformation and therefore could be of influence on the repeatability of pelvic measurements.
Another explanation of differences of degrees in estimates for reliability with pelvic measurement in different breeds could be that the temperament of different breeds ensures more of less movement of the animal during the measurements. Using an epidural anaesthesia like Kolkman et al. (2009) could influence the obtained results and an epidural anaesthesia would not be part of an easy on-farm culling tool because of time and veterinary expenses. The temperament of different breeds has been researched as a subjective crush movement score by different researches(Gibbons, Lawrence, & Haskell, 2011; Grandin, 1993). Grandin et al (1993) recorded the temperament on a scale of 1-5 of each cow right after the crush was closed. They found that from the four measurements, 17% of the cows were agitated (score 4 or 5) in the crush for at least three out of four measurements, 51% of the cows stayed calm (score 1 or 2) during all four sessions and the rest of the cows (32%) had mixed temperament (score 3 or higher in 1 or 2 out for four sessions)(Grandin, 1993). Gibbons et al. (2011) observed the crush movement score of dairy cattle 30 seconds after being held in a crush, with the use of the same scoring system as Grandin et al. (1993). The repeatability of the crush movement score has been evaluated over 4 different measurements on Angus cattle(Halloway & Johnston, 2003). There were two measurements days which were 73 days apart and on each day they measured the crush movement score twice(Halloway & Johnston, 2003). The repeatability score between two measurements on the same day were 0,60 (day 1) and 0.58 (day 2) and the repeatability score between the first measurement on day 1 and the first measurement on day 2 was 0.44(Halloway & Johnston, 2003).
Another variable which could be of influence on the repeatability and reproducibility of pelvic measurements is the experience of the observers. Van Donkersgoed et al. (1993) noted that the observer was a veterinarian with experience on pelvimetry, Kolkman et al. (2009) and Short et al. (1979) didn't report the experience of the observers.
The aim of this study is to find out if transrectal pelvic measurements, horizontal and vertical pelvic measurements, with the Rice-pelvimeter in South African beef cattle is repeatable and reproducible. The influence of the different breeds (Brahmans, Bonsmara's, Nguni's and Herefords), crush movement score and the experience of the observer on the repeatability and the reproducibility of the pelvic measurements will be subject of this research as well.

Materials and methods

Material

For this research we used the Rice-pelvimeter (Lane Manufactering, 2075 So. Balentia St., Unit C,Denver, Colorado, USA), figure 1. All measurements were made by one and the same instrument. The Rice-pelvimeter consist of two arms with a scale in between. The scale goes from 0 to 20 cm with increments of 0.5 cm, the measurements can be read on the inside from the movable arm next to the arrow,

***Figure 1**: Rice-pelvimeter*

figure 2. There are two measurements which can be measured via the rectum, the width or the horizontal measurement (PH) and the height or the vertical measurement (PV), figure 3. The horizontal measurements are made by bringing the Rice-pelvimeter closed in the rectum as far cranial so that the two ends are just behind the pelvic. Then you should carefully open the two jaws by squeezing the measuring end. After opening the jaws in the rectum, you slowly move the pelvimeter caudal, while keeping a little tension on the measuring ends so that the jaws will pass the pelvis on the widest point, the poas tubercles. The vertical measurements are made by bringing the jaws closed in the rectum and carefully opening the jaws placing one jaw on the symphysis pubis and the other jaw on the sacral vertebrae.

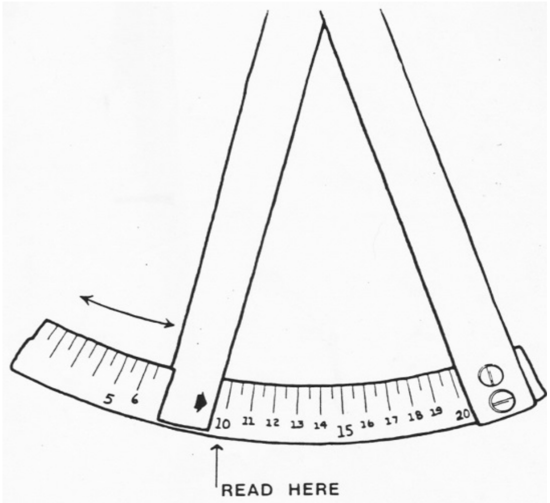

Figure 3*: Measurement scale, where the measurements are red at the point of the arrow on the movable arm*

Animals

A total of 111 heifers and bulls from five different farms were used for this research. The heifers used for this research came from farms in the area around Pretoria, South Africa. The farms we visited had already contributed to previous measurements done by two other Dutch veterinarian students and professor D.E. Holm in June/July 2015. The heifers and bulls we measured were the same as measured in June/July 2015, unfortunately less heifers were available because some cows were already culled, sold or the farmer couldn’t find his cattle. The five farms that joined this research were selected by professor D.E. Holm because of their pedigree breed beef cattle and their distance from the faculty of veterinary studies in Onderstepoort, South Africa. Each farm had a different beef cattle breed except for two farms which both had Bonsmara cattle, we will refer to them as the UP herd and the PGM herd, these letter combinations where the start of the animal identification numbers. The different breeds which participated on this research were Nguni (n=10), Hereford (n=36), Brahman (n=24) and Bonsmara (n=41) cattle (UP: n=16 and PGM: n=25). The 99 heifers and 12 bulls selected for the measurements should be originated from the farm herd and not bought, should be healthy and from the typical breed of the farm. To determine the repeatability and reproducibility of the Rice-pelvimeter it should not matter what age or in which pregnancy stage the animals are.

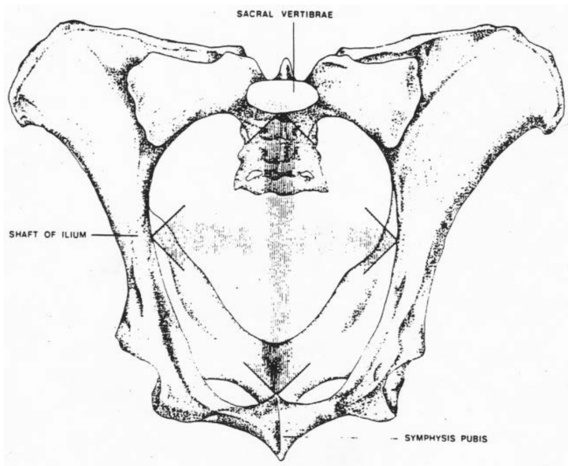

Figure 2*: Horizontal and vertical measurements measured between the arrows.*

Measurements

The measurements took place from November till December 2015. The measurements were done by professor and veterinarian Dietmar Holm (DH) and two Dutch veterinarian master students Marloes Wansink (MW) and Lisa Algie (LA). DH is experienced in measuring horizontal and vertical pelvis measurements with the Rice-pelvimeter. MW and LA had no experience measuring pelvis areas with the Rice-pelvimeter. The two students did practice the use of the Rice-pelvimeter for approximately two hours a day for five days before starting with the measurements for the research, half of the time DH was there to confirm or refute the measurements of MW and LA. The practice was done on the herd of the Faculty of Veterinary Science of the University of Pretoria. Each observer measured the height (PV) and width (PH). Each observer measured the PV and PH twice at each cow, between these two measurements the conditions stayed the same. The measurements were taken blind and randomly twice and the observers couldn't see the values of the measurements of the other observers. From each cow and bull, the animal identification number was noted, so that at the end the first and second measurements could be paired in the data. Other data which was measured with the first measurement of each cow was breed and the crush movement score. There were four South African beef cattle breeds represented in the research, Brahmans, Bonsmara, Nguni and Hereford. They belonged to five different herds, there were 2 Bonsmara herds. Each herd was measured on a different day., except the PGM herd the heifers (n=20) and the bulls (n=5) were measured on separate days. The crush movement score was determined in consultation by the three observers approximately 30 seconds after the cow was caught in the crush for the first measurement. The scale used was the same scale as used by Gibbons (2011) and consisted of 5 scores, table 1.

Table 1
Score table for the crush movement score of the animals while standing in the crush.

Score	Description
0	Standing stationary, calm, no movement, no resistance
1	Slight calm movement of feet/tail, but mainly stationary
2	Uneasy, head, body, tail and feet movement
3	Struggling, continuous movement, straining at head restraint
4	Active escape behaviour, animal may kneel/fall

Statistics

The total data set of 111 cows and bulls was cleaned from incomplete data because either one of the observers forgot to write down one of the measurements (n=2) or because some cows escaped between the first and the second measurement (n=3). The data set of 106 cows was used for the repeatability and reproducibility analyses. Analyses were done with SPSS 24. The repeatability of the measurements (PV and PH) was determined by the Pearson's correlation coefficient, Cohen's kappa coefficient and the limits of agreement between the first and the second measurement of an observer. The limits of agreements are determined as the mean of the differences $\pm$ 2 times the standard deviation of the differences this gives the limits within which 95% of the differences hold. We used a weighted Cohen's kappa coefficient where a difference of half a centimetre also counted as an exact agreement between the first and the second measurement. The reproducibility of the measurements (PV and PH) was determined by the Pearson's correlation coefficient, Cohen's kappa coefficient and the limits of agreement between the means of the within observer measurements. To determine the influence of the variables, breed and crush movement score on the repeatability and the reproducibility we used a generalized linear mixed model and a logistic regression model. A generalized linear mixed model for the repeatability was used because each animal had three dependent variables, the difference between the first and the second measurements for each observer. For the reproducibility, a generalized mixed model was because each animal had two dependent variables, the difference between the mean measurements of DH and the mean measurements of MW or LA. A generalized linear mixed model was fitted by choosing the model with the lowest AIC with at least the observer as a fixed factor for the horizontal and the vertical measurements in the model. This was done for the measurement taken between observers and within observers. The data variable 'difference' between the first and second measurements and between the mean measurement of DH and the mean measurements of MW and LA was then transformed to a binary data set with difference > 0.5 being a 1 and difference $\leq$ 0.5 being a 0. A logistic regression was done on the binary data set to determine the odds on a difference with variables observer, breed and crush movement scores as explanatory variables. A model was created where the variables stayed in the model that had significant odds on a difference.

Results

Descriptive statistics

The age of the 111 animals during the research was only recorded for 59 of the animals, they were aged between 12 and 42 months. The twelve bulls were aged between the 12 and 38 months. From the cows seven had already calved for the first time, 33 were pregnant during the measurements and 59 heifers were not pregnant or in such early stage it wasn't palpable. The mean crush movement score was 1.7 $\pm$ 1.3 for all the animals, the mean crush movement score and standard deviation per breed is shown in table 2. The Hereford cattle had the lowest crush movement scores, while the Brahman cattle had the highest crush movement scores.

Table 2

The number of animals, gender, age, the mean $\pm$SD of the horizontal pelvic measurement, the mean $\pm$SD of the vertical pelvic measurement and the mean $\pm$SD crush movement score per breed.

Breed	n	Gender	Age	PH	PV	Crush movement score
Brahman	24	7 ♂ 17 ♀	12-33 months	10.9 $\pm$ 1.4	16.3 $\pm$ 1.0	3.1 $\pm$ 0.8
Bonsmara	41	5 ♂ 36 ♀	23-42 months (16 animals age unknown)	12.1 $\pm$ 1.5	15.6 $\pm$ 1.4	1.9 $\pm$ 1.1
Nguni	10	10 ♀	24-34 months	10.0 $\pm$ 0.6	13.9 $\pm$ 0.8	2.3 $\pm$ 1.0
Hereford	36	36 ♀	Age unknown	14.9 $\pm$ 0.8	17.4 $\pm$ 1.2	0.6 $\pm$ 0.7

The pelvic measurements measured by DH had a horizontal pelvic mean and SD and vertical pelvic mean and SD of 12.7 $\pm$ 2.2 cm and 16.3 $\pm$ 1.5 cm respectively. The horizontal and vertical pelvic mean and SD measured by MW were 12.8 $\pm$ 2.1 cm 16.2 $\pm$ 1.5 cm respectively. LA had horizontal and vertical pelvic means and SDs of 13.4 $\pm$ 2.1 cm and 16.5 $\pm$ 1.7 cm respectively.
The first and the second measurement of each observer was exactly the same or differed half a centimeter in more than 66% of the repeated measurements, table 3. The absolute difference of 0 to 0.5 cm between the first and second horizontal measurements of DH occurred in 75.5% (95% CI: 67.3% - 83.7%) of the animals. For the vertical measurements, the difference between the first and the second measurements of DH was zero to half a centimeter in 82.1% (95% CI: 74.8% – 89.4%). The less experienced observers MW and LA had a difference of zero or half a centimeter for the horizontal measurements in 66% (95% CI: 57.0% – 75.1%) and 76.4% (95% CI: 68.3% – 84.5%) respectively. Between the vertical measurements the difference was zero or half a centimeter in 73.6 % (95% CI: 65.2% – 82.0%) of the measurements of MW and 67.9% (95% CI: 59.0% – 76.8%) of the measurements of LA. The more experienced observer DH had the difference spread from 0 to 2 and from 0 to 1.5 for the horizontal and vertical measurements respectively. The unexperienced observers had differences between the horizontal and vertical measurements spread from 0 to 3.5 (MW, for both directions), from 0 to 3 (LA, horizontal measurements) and from 0-2.5 (LA, vertical measurements) (table 3).

Table 3
Frequency of the differences between the first and second horizontal or vertical measurements of all the animals (n=106) for each observer, expressed in percentages.

		Difference between the first and the second measurement							
	Observer	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
Horizontal pelvis measurements	**DH**	38.7	36.8	20.8	2.8	0.9	0	0	0
	LA	34.0	42.5	12.3	7.6	0.9	1.9	0.9	0
	MW	28.3	37.7	22.6	6.6	0.9	1.9	0.9	0.9
Vertical pelvis measurements	**DH**	46.2	35.9	14.2	3.8	0	0	0	0
	LA	29.3	38.7	21.7	4.7	1.9	3.8	0	0
	MW	42.5	31.1	16.0	6.6	0.9	0.9	0	1.9

The means, standard deviation and minimum and maximum measurement of the first, second and mean horizontal and vertical pelvic measurement of each observer is shown in table 4. The means of DH are compared with the means of the unexperienced observers. The means for the horizontal and vertical measurements of LA are on average 0.7 cm and 0.27 cm higher than the means for those measurements of DH, respectively. Also, the means of MW are on average 0.13 cm for the horizontal measurements and 0.07 cm for the vertical measurements higher than the means of DH. Yet the standard deviations of the measurements are comparable between observers.

Table 4
The mean, standard deviation (SD) and minimum and maximum measurement in cm of the horizontal and vertical pelvic measurements for each measurement (sequence 1 and 2) and the mean of the first and second measurement (Mean) measured with the Rice-pelvimeter by three observers (DH, MW and LA). (n=106)

Measurement	Observer	Sequence	Mean	SD	Min	Max
Horizontal pelvic measurement	**DH**	1	12.6	2.1	8.50	16.50
		2	12.8	2.2	8.50	16.50
		Mean	12.7	2.2	8.50	16.50
	MW	1	12.7	2.1	8.50	17.00
		2	13.0	2.2	8.50	16.50
		Mean	12.8	2.1	8.50	16.50
	LA	1	13.4	2.2	9.50	19.00
		2	13.4	2.2	9.50	17.00
		Mean	13.4	2.1	9.50	18.00
Vertical pelvic measurement	**DH**	1	16.2	1.6	13.00	20.00
		2	16.3	1.6	13.00	19.50
		Mean	16.3	1.5	13.00	19.75
	MW	1	16.1	1.6	12.00	20.00
		2	16.3	1.5	13.00	19.50
		Mean	16.2	1.5	13.00	19.50
	LA	1	16.5	1.8	13.00	20.00
		2	16.6	1.8	13.00	20.00
		Mean	16.5	1.7	13.00	20.00

Repeatability

The first and de second measurement of the horizontal and the vertical measurement of each observer were plotted in scatter plots, figure 4. The scatter plots of DH's measurements are narrower than the scatter plots of MW and LA. The dots for the horizontal measurements lie closer to the Y=X-line than the dots for the vertical measurements, independent of the observers. Observers MW and LA seem to have more outliners than DH. The Pearson's correlation coefficient, the Cohen's kappa coefficient as well as the weighted Cohen's kappa coefficient between the first and the second measurements for each observer are shown in table 5. The measurements of DH are comparable correlated with the measurements of the unexperienced MW and LA. Overall the horizontal measurements tend to be better correlated than the vertical measurements. The Cohen's kappa coefficient shows there is only a fair agreement between the first and the second measurement. The measurements by DH show a slightly better within animal agreement than the measurements of MW and LA. With the weighted Cohen's kappa coefficients, the agreement between the repeated measurements are classified as moderate to substantial (κ between 0.50 and 0.71). Observers DH an MW had a better agreement for the repeated horizontal measurements then for the vertical measurements. In contrary, the repeated vertical measurements had better agreement than the repeated horizontal measurement with observer LA.

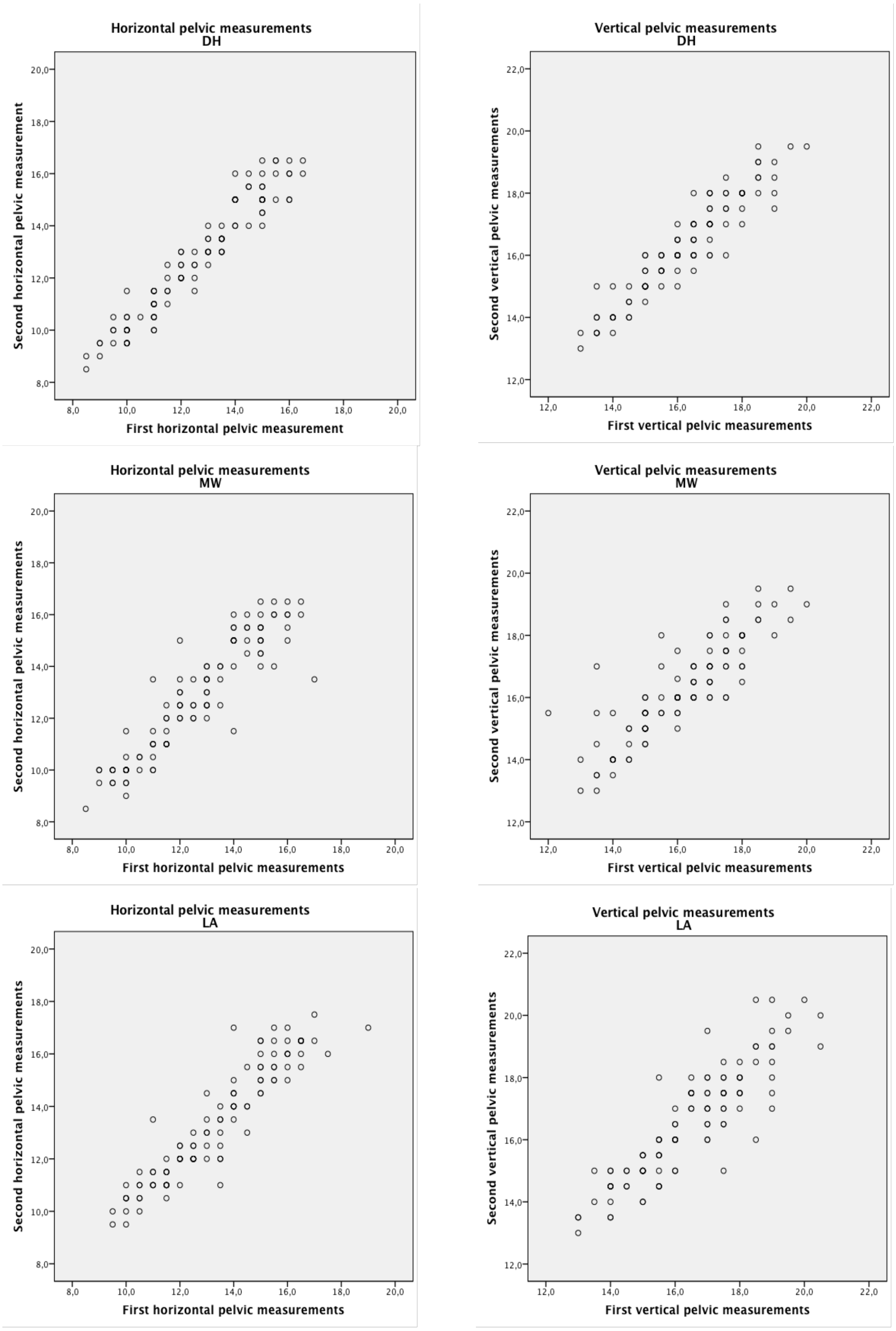

Figure 4*.Scatter plots of the first and the second horizontal and vertical measurements of each observer. The first measurements are on the X-axis and the second measurements are on the Y-axis. Bolder dots represent more than one animal with the same value on the X- and Y-axis.*

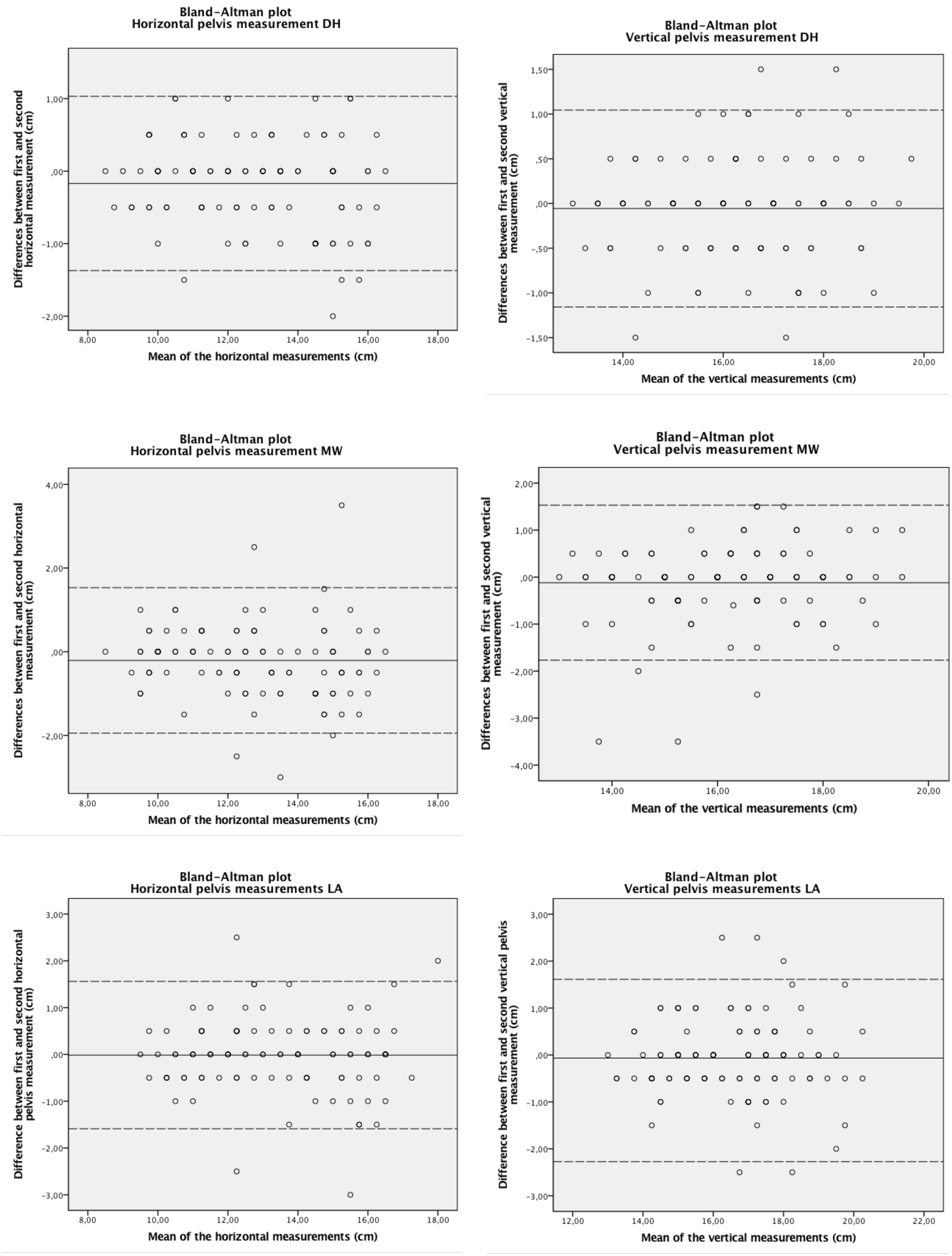

***Figure 5**: Bland-altman plots of the repeated horizontal and vertical measurement of each observer. On the X-axis the means of the first and the second measurement. On the Y-axis the difference between the first and the second measurement. The unbroken line represents the mean difference (bias) and the dashed lines represent the upper and lower limit of agreement.*

The means of the repeated measurements (x-axis) and the difference between the measurements (Y-axis) were plotted against each other to create two Bland-Altman plots (horizontal measurements and vertical measurements) for each observer, figure 5.
The bias for the repeated horizontal and the repeated vertical measurements for all observers are close to zero ranging from -0.21 to -0.01, as shown in table 6. The standard deviation of the differences (SD_{Diff}) of MW for the horizontal (0.89 cm) and vertical measurements (0.84 cm) and the SD_{Diff} of LA for the horizontal (0.80 cm) and vertical measurements (0.86 cm) are both higher than the SD_{Diff} of 0.61 cm and 0.56 cm for the horizontal and vertical measurements, respectively. As a consequence, the limits of agreements of the unexperienced observers are wider compared to this range for DH (Table 6).

Table 5:
Pearson's correlation coefficient (r), Cohen's kappa coefficient (κ) and weighted Cohen's kappa coefficient (κ) between the first and the second horizontal measurement (PH) and vertical measurement (PV) for each observer.

Observer	Measurements	Pearson's correlation coefficient (r)	Cohen's Kappa coefficient (κ)	Weighted Cohen's kappa coefficient (κ)
DH	PH	r = 0.96*	κ = 0.34	κ = 0.60
	PV	r = 0.94*	κ = 0.41	κ = 0.71
MW	PH	r = 0.92*	κ = 0.23	κ = 0.50
	PV	r = 0.86*	κ = 0.37	κ = 0.60
LA	PH	r = 0.93*	κ = 0.29	κ = 0.64
	PV	r = 0.88*	κ = 0.23	κ = 0.54

**P* ≤0.01*

Table 6:
The mean difference (bias), the standard deviation of the differences(SD_{Diff}) and the upper and lower limit of agreement (95% of the differences are between these limits) between the first and the second measurements of each observer in centimeters.

	Horizontal measurements			Vertical measurements		
	DH	**LA**	**MW**	**DH**	**LA**	**MW**
$Mean_{Diff}$	-0.17	-0.01	-0.21	-0.06	-0.07	-0.12
SD_{Diff}	0.61	0.80	0.89	0.56	0.86	0.84
Upper limit of agreement	1.03	1.56	1.53	1.04	1.61	1.53
Lower limit of agreement	-1.37	-1.59	-1.95	-1.16	-2.27	-1.77

The mean differences between the first and the second horizontal and vertical measurements per observer, breed and crush movement score are shown in table 7 and table 8 respectively. The mean difference between the first and the second horizontal measurement for the variables are for the observers between -0.21 and -0.04, for the breeds between -0.32 and 0 and for the crush movement scores between -0.29 and 0.04. Comparable, but less spread are the mean differences between the first and second vertical measurement for the variables of -0.11 to -0.06 for the observers, -0.17 and -0.01 for the breeds and -0.14 to 0.02 for the crush movement scores. The SD_{Diff} between the first and the second horizontal measurement of MW and LA are with 0.89 and 0.81 respectively higher than the SD_{Diff} between the first and the second horizontal measurement of 0.62 of DH, figure 6. Similarly, the SD_{Diff} between the first and second vertical measurement of MW (0.86) and LA (0.85) are higher than the SD_{Diff} between the first and second vertical measurement of DH (0.56), figure 7. The SD_{Diff} for the horizontal measurements for the Nguni's is with 0.45 smaller than the SD_{Diff} of the Brahmans (0.76), the Bonsmara's (0.77) and the Herefords (0.84). Similarly, the SD_{Diff} for the vertical measurements for the Nguni's is with 0.46 smaller than the SD_{Diff} of the Brahmans (0.85), the Bonsmara's (0.79) and the Herefords (0.74).

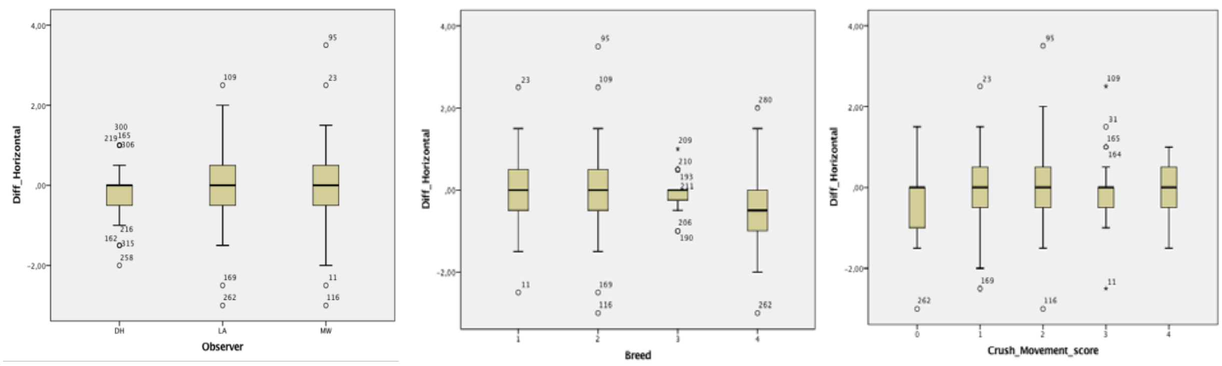

Figure 6: *Boxplots of the differences between the first and the second horizontal measurements in cm per variable observer, breed (Brahmans (1), Bonsmara (2), Nguni (3) and Hereford (4)) and crush movement score.*

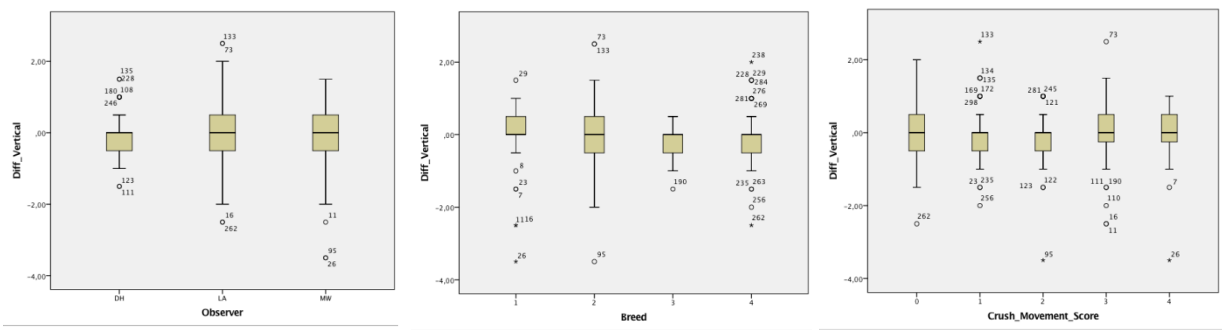

Figure 7: *Boxplots of the differences between the first and second vertical measurements in cm per variable observer, breed (Brahmans (1), Bonsmara (2), Nguni (3) and Hereford (4)) and crush movement score.*

The linear mixed effect model shows that the mean difference between the first and second horizontal measurement (Diff H_{1-2}) of MW differs -0.03 cm (P = 0.714, 95% CI: -0.24 – 0.17) from Diff H_{1-2} of DH, while the Diff H_{1-2} of LA differs 0.16 cm (P = 0.132, 95% CI: -0.05 – 0.36).The mean difference between the first and second vertical measurement (Diff V_{1-2}) of MW differs -0.06 cm (P=0.528, 95% CI: -0.25 – 0.13) from Diff V_{1-2} of DH and the Diff V_{1-2} of LA differs -0.01 cm (P=0.923, 95% CI: -0.20 – 0.18) from Diff V_{1-2} of DH. The logistic regression showed that LA has 2.168 times bigger odds on a difference of more than half a centimetre between the first and second vertical measurement than DH (P=0.018, 95% CI: 1.139 – 4.125). Similarly, MW has 1.646 times bigger odds on a difference of half a centimetre or more between the first and the second vertical measurement than DH (P=0.138, 95% CI: 0.851 – 3.181). The odds on a difference (>0.5) between the first and second horizontal measurement was influenced by the breed of cows, this was significant. The odds on a difference (>0.5 cm) between the first and the second horizontal measurement is 2.318 for the Herefords compared to the Brahmans (P=0.018, 95% CI: 1.153 – 4.547). The influence of the observers on the occurrence of a difference was not significant, but we kept this variable in the model to estimate the effect of observer. The influence of the observer on the difference between the first and second horizontal measurement was not significant. The odds ratio that the difference between the first and second horizontal measurement was bigger than 0,5 in comparison to DH was 0.946 (P=0.868, 95% CI: 0.492-1.181) for LA and 1.646 (P=0.117, 95% CI: 0.882 – 3.070) for MW.

Table 7:
The mean differences ($Mean_{Diff}$), the standard deviation of the differences(SD_{Diff}) and the minimum (Min_{Diff}) and maximum (Max_{Diff}) difference between the first and second horizontal measurement per variable observer, breed and crush movement score in 106 animals of 4 South African beef breeds.

	$\mathbf{Mean_{Diff}}$	$\mathbf{SD_{Diff}}$	$\mathbf{Min_{Diff}}$	$\mathbf{Max_{Diff}}$
Observer				
DH	-0.17	0.62	-2	1
LA	-0.04	0.81	-3	2.5
MW	-0.21	0.89	-3	3.5
Breed				
Brahmans (1)	0.00	0.76	-2.5	2.5
Bonsmara (2)	-0.05	0.77	-3	3.5
Nguni (3)*	-0.06	0.45	-1	1
Hereford (4)	-0.32	0.84	-3	2
Crush movement score				
0	-0.29	0.74	-3	1.5
1	-0.11	0.83	-2.5	2.5
2	-0.13	0.92	-3	3.5
3	-0.09	0.68	-2.5	2.5
4**	0.04	0.61	-1.5	1

* Only eight Nguni’s took part in the research
** Only twelve cows had a crush movement score of 4

Table 8:
The mean difference (Mean$_{Diff}$), the standard deviation of the differences(SD$_{Diff}$) and the minimal (Min$_{Diff}$) and maximum (Max$_{Diff}$) differences between the first and second vertical measurement per variable observer, breed and crush movement score in 106 animals of 4 South African beef breeds.

	Mean$_{Diff}$	**SD$_{Diff}$**	**Min$_{Diff}$**	**Max$_{Diff}$**
Observer				
DH	-0.06	0.56	-1.50	1.50
LA	-0.06	0.86	-2.50	2.50
MW	-0.11	0.85	-3.50	1.50
Breed				
Brahmans (1)	-0.01	0.85	-3.50	1.50
Bonsmara (2)	-0.07	0.79	-3.50	2.50
Nguni (3) *	-0.17	0.46	-1.50	0.50
Hereford (4)	-0.11	0.74	-2.50	2.00
Crush movement score				
0	-0.09	0.78	-2.50	2.00
1	-0.11	0.72	-2.00	2.50
2	-0.14	0.72	-3.50	1.00
3	0.02	0.84	-2.50	2.50
4 **	-0.04	0.82	-3.50	1.00

* Only eight Nguni's took part in the research
** Only twelve cows had a crush movement score of 4

Reproducibility

The mean of the measurements of LA and MW are compared with the mean of the measurements of DH to ascertain the level of agreement between the unexperienced and the experienced observer. The Pearson's correlation coefficient and the Cohen's kappa coefficient of the difference between the mean measurements of DH and the mean measurements of MW and LA for the horizontal and vertical measurements are shown in table 9. The mean of measurements of MW are a slightly better correlated than the mean of measurements of LA with the mean of measurements of DH. Further the kappa coefficient shows a far better agreement between the measurements of MW and DH (0.539 and 0.545 respectively the horizontal and vertical measurements) than the measurements of LA and DH (0.048 and 0.182, respectively horizontal and vertical measurements). The mean of the vertical measurements of MW and LA are less correlated with the mean of the vertical measurements of DH than the mean of the horizontal measurements. Whereas the kappa coefficients of the horizontal measurements of MW and LA show lesser agreement with DH than the kappa coefficients of the vertical measurements of MW and LA. The Fleiss kappa coefficient shows less agreement between observers for the second horizontal and vertical measurements than with the first horizontal and vertical measurements (table 10).

Table 9:
The Pearson's correlation coefficient and Cohen's kappa coefficient between the mean of measurements of DH and the mean of measurements of MW or LA for the horizontal and vertical pelvic measurements.

	Mean DH - observer	Pearson's correlation coefficient (r)	Cohen's kappa coefficient (κ)
Horizontal pelvic measurements	**Mean LA**	0.95*	0.048**
	Mean MW	0.97*	0.539*
Vertical pelvic measurements	**Mean LA**	0.93*	0.182*
	Mean MW	0.96*	0.545*

* $p < 0.001$
** $p = 0.004$

Table 10:
Fleiss kappa coefficient between the three observers for the first and second measurement and the mean of the measurements.

	Fleiss kappa coefficient (κ)
Mean horizontal pelvic measurements	0.253
First horizontal pelvic measurement	0.318
Second horizontal pelvic measurement	0.296
Mean vertical pelvic measurements	0.317
First vertical pelvic measurement	0.412
Second vertical pelvic measurement	0.346

To create a Bland-Altman plot for the reproducibility the mean of the mean measurements of DH and the mean measurements of MW or LA are plotted on the X-axis and difference between the mean measurements of DH and the mean measurements of MW or LA are plotted on the y-axis (figure 8). Table 11 shows the bias between the mean measurements of DH and the mean measurements of MW and LA. The bias between the mean measurements of DH and MW are 0.03 cm for the horizontal measurements and -0.05 cm for the vertical measurements. In contrast to the bias between the mean measurements of DH and LA of 0.69 cm and 0.27 cm for the horizontal and vertical measurements respectively. Additionally, the SD_{Diff} for the measurements of LA of 0.71 cm for the horizontal measurements and 0.63 cm for the vertical measurements are higher than the SD_{Diff} for the measurements of MW for the horizontal and vertical measurements, 0.56 cm and 0.44 cm respectively. The mean measurements of MW have a better agreement with the mean measurements of DH in comparison to the mean measurements of LA. The mean vertical measurements have a better agreement with the mean measurements of DH for both unexperienced observer than the mean horizontal measurements (table 11).
The mean differences ($mean_{Diff}$) and standard deviation (SD_{Diff}) of the differences between the measurements of DH and the measurements of MW and LA are shown in table 12 for the horizontal measurements and table 13 for the vertical measurements per variable observer, breed and crush movement score. The $mean_{Diff}$ of the variable Hereford is 0.38 cm for the vertical measurements which is bigger than the $mean_{Diff}$ for the Brahmans and Nguni's of both -0.05 cm and the Bonsmara of 0.01 cm. The SD_{Diff} for the Nguni's is 0.25 cm for the vertical measurements, while the SD_{Diff} for the Bonsmara's 0.44 cm is and the SD_{Diff} for the Brahmans and Herefords are 0.64 and 0.61 respectively. The crush movement score

shows a $mean_{Diff}$ which is getting smaller the higher the crush movement score is for the vertical measurements (table 13). Though the SD_{Diff} doesn't shown the same pattern with the different crush movement score for the vertical measurements. Furthermore, the $mean_{Diff}$ and the SD_{Diff} for the horizontal measurements per different crush movement score are comparable with each other (table 12). The Nguni's show the lowest $mean_{Diff}$ and SD_{Diff} for both the horizontal and vertical measurements compared to the other breeds.
The linear mixed effect model for the horizontal measurements shows that the difference between the mean measurements of LA differ 0.65 cm more from the mean measurements of DH than the mean measurements of MW (P<0.001, 95% CI: 0.51 – 0.80). There is no significant difference between the different breeds for the horizontal measurements. The linear mixed effect model for the vertical measurements show that there is a significant difference between the two observers and the four breeds. The difference between the mean measurements of LA differ 0.32 cm more from the mean measurements of DH than the mean measurements of MW (P<0.001, 95% CI: 0.20 – 0.45). The difference between the mean measurements of DH and the mean measurements of MW or LA are for the Brahmans 0.45 cm (P<0.001, 95% CI: -0.66 – -0.24), the Bonsmara's 0.37 cm (P<0.001, 95% CI: -0.54 – -0.20) and the Nguni's 0.47 cm (P=0.002, 95% CI: -0.79 – -0.17) smaller than for the Herefords. The logistic regression shows that LA has 9.436 times bigger odds that the difference between DH and LA is bigger than 0.5 cm for the horizontal measurements than MW (P<0.001, 95% CI 4.697 – 18.956). For the vertical measurements LA has 7.031 times bigger odds that the difference between DH and LA is bigger than 0.5 cm than MW (P<0.001, 95% CI: 2.964 – 16.678). The logistic regression for vertical measurement show no significant difference for the variable breeds.

Table 11
The mean differences ($Mean_{Diff}$)and the standard deviation of the differences (SD_{Diff}) between the mean measurements of DH and the mean measurements of MW and LA. As well as the upper and lower limit of agreement between the mean measurements of DH and MW and LA.

	Horizontal pelvic measurements		**Vertical pelvis measurements**	
Mean DH -	**Mean MW**	**Mean LA**	**Mean MW**	**Mean LA**
$Mean_{Diff}$	0.03	0.69	-0.05	0.27
SD_{Diff}	0.56	0.71	0.44	0.63
Upper limit of agreement	1.13	2.07	0.82	1.50
Lower limit of agreement	-1.06	-0.70	-0.91	-0.96

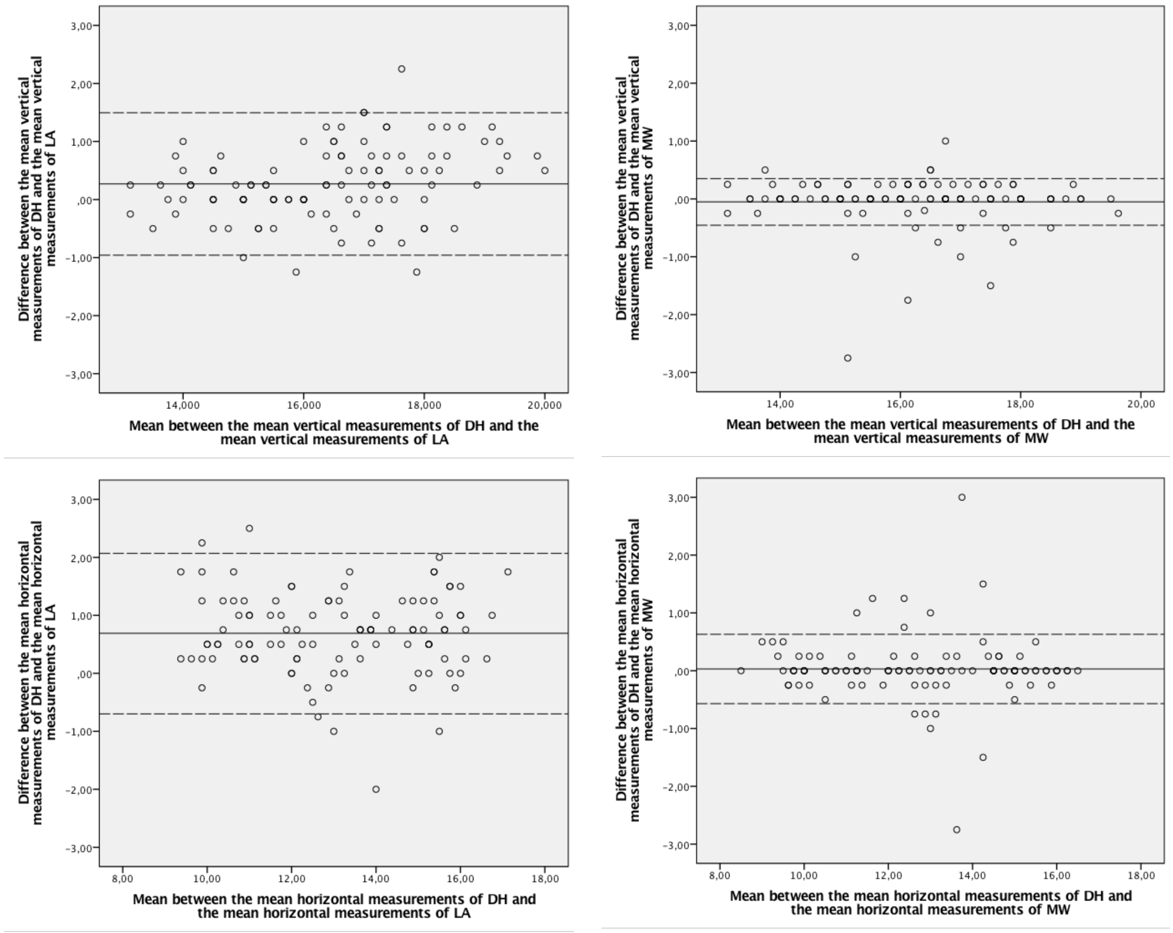

Figure 8: *Bland-altman plots of the mean between the mean horizontal and vertical measurement of DH with the mean measurements of MW or LA. On the X-axis the mean between the mean measurements of DH and the mean measurements of MW or LA. On the Y-axis the difference between the mean measurements of DH and the mean measurements of MW or LA. The unbroken line represents the mean difference (bias) and the dashed lines represent the upper and lower limit of agreement.*

Table 12:
The mean differences (mean$_{Diff}$), the standard deviation of the differences (SD$_{Diff}$) and the minimum and maximum differences between the measurements of DH and the measurements of MW and LA per variable observer, breed and crush movement score for the horizontal pelvic measurements.

	Mean$_{Diff}$	**SD$_{Diff}$**	**Min$_{Diff}$**	**Max$_{Diff}$**
Observer				
MW	0.03	0.56	-2.75	3.00
LA	0.69	0.71	-2.00	2.50
Breed				
Brahmans (1)	0.63	0.63	-0.25	2.25
Bonsmara (2)	0.24	0.85	-2.75	3.00
Nguni (3) *	0.11	0.24	-0.25	0.50
Hereford (4)	0.41	0.62	-1.00	2.00
Crush Movement				
0	0.30	0.65	-1.00	1.50
1	0.27	0.77	-2.75	2.50
2	0.41	0.74	-1.50	3.00
3	0.47	0.72	-0.75	1.75
4 **	0.44	0.69	-0.75	2.25

* Only eight Nguni's took part in the research
** Only twelve cows had a crush movement score of 4

Table 13:
The mean differences (mean$_{Diff}$), the standard deviation of the differences (SD$_{Diff}$) and the minimum and maximum differences between the measurements of DH and the measurements of MW and LA per variable observer, breed and crush movement score for the vertical pelvic measurements.

	Mean$_{Diff}$	**SD$_{Diff}$**	**Min$_{Diff}$**	**Max$_{Diff}$**
Observer				
MW	-0,05	0,44	-2,75	1,00
LA	0,28	0,62	-1,25	2,25
Breed				
Brahmans (1)	-0,05	0,64	-2,75	0,75
Bonsmara (2)	0,01	0,44	-1,75	1,00
Nguni (3) *	-0,05	0,25	-0,50	0,25
Hereford (4)	0,38	0,61	-1,00	2,25
Crush Movement				
0	0,27	0,69	-1,00	2,25
1	0,18	0,46	-1,00	1,25
2	0,10	0,53	-1,75	1,25
3	-0,02	0,44	-1,50	0,75
4 **	-0,03	0,71	-2,75	1,00

* Only eight Nguni's took part in the research
** Only twelve cows had a crush movement score of 4

Discussion

In the present study, we assessed the reliability of pelvic measurements done with the Rice-pelvimeter done by experienced and unexperienced observers in four South African beef cattle. This was done by comparing the six pelvis measurements of 106 animals measured by three observers. The within and between observer agreement was evaluated. The results show small difference in the agreement of the horizontal and vertical pelvic measurements for the experienced observer. However, the results showed a varying agreement of the horizontal and vertical pelvic measurements for the unexperienced observers. Further, the between observer agreement with the experienced observer differed for both unexperienced observers.

The agreement within and between observers is analyzed by a Pearson's correlation coefficient, a Cohen's kappa coefficient and the limits of agreement. The Pearson's correlation coefficient is an incorrect way of analyzing agreement between two measurements (Bland & Altman, 1986). A correlation coefficient doesn't measure the agreement between two measurements but the strength of the relation between two measurements(Bland & Altman, 1986). Short et al. (1979) reported a good repeatability by using the correlation coefficient as measure of agreement. Similarly, this study reported high correlation coefficients for the measurements within (r between 0.86 and 0.96) and between (r between 0.93 and 0.97) observers, but as Bland & Altman (1986) reported this is not the same as a good agreement.

The results of this study show a good repeatability for pelvimetry done by an experienced observer. The bias and limits of agreement for the experienced observer are for the horizontal measurements -0.17 cm (-1.37 – 1.03) and for the vertical measurements -0.06 cm (-1.16 – 1.04). The repeatability for the unexperienced observers is less good with broader limits of agreement for the horizontal (limits of agreement MW -1.95 – 1.53 and LA -1.59 – 1.26) and vertical pelvic measurements (limits of agreement MW -1.77 – 1.53 and LA -2.27 – 1.61). We have to keep in mind that the Rice-pelvimeter has a scale with increments of half a centimeter, a small difference is easily made. Even though there is a statistical difference between the measurements this doesn't mean the difference is of biological relevance. For instance, the mean PA of the Hereford from this study is 259 cm^2, following the limits of agreement of DH the PA could differ 40 cm^2 which is 15%. More research is needed to find out is this difference in measurements would have a biological relevance in preventing or predicting dystocia.

The difference for within agreement of the vertical measurements was only influenced by the observer, not by the breed or the crush movement score. The odds of a difference >0.5 of LA (2.168) were bigger than the odds of MW (0.946). The horizontal measurements were not significantly influence by the observer, but was influenced by breed. The Hereford had an odds ratio of 2.318 on a difference >0.5 compared with the Brahmans.

The agreement between the experienced and unexperienced observers varied. The horizontal measurement of MW (0.03 cm, -1.06 – 1.13) agreed better with measurements of DH than the measurements of LA (0.69 cm, -0.70 – 2.07). For the vertical measurements showed an equal pattern, MW (-0.05 cm, -0.92 – 0.82) agreed better with DH than LA (0.27 cm, -0.96 – 1.50). The reproducibility of the horizontal and vertical measurements was only significantly influenced by the observer. The odds ratio for a difference >0.5 cm of LA was 9.436 and 7.031 for the horizontal and vertical measurements respectively compared with MW.

The difference between observers might be explained by the fact that the height of the observers varied, which could have caused a different angle during measuring the pelvic dimensions. With a length of 1.59 m LA is considerably smaller than the 1.72 and 1.79 of MW and DH respectively. Or the unexperienced observer LA didn't take the vertical measurement between the symphysis pubis and sacral vertebrae. The difference of the breed, could also be a difference between two herds instead of the breed. Different pelvic conformation for both or one of the breed could be an explanation of the breed differences, like Kolkman et al. (2009) cited for the Belgian Blue breed. Another difference between the different breeds in this study was the number of animals that was measured on one day. The Nguni, showed a small bias and small range between the upper and lower limit of agreement, this might be caused by the fact that only a small number of animals was measured on that day. These findings would agree with the hypothesis of Short et al. (1979) that the number of animals influence the reliability of the pelvic measurements.
An influence of the crush movement score wasn't found, but this might be due to the fact that there was only on crush movement score per animal, while the animals were standing in the crush for a longer time because all observer had to take their measurements. The better way of analyzing the influence of the crush movement score would have been to score a crush movement score during each measurement.
Before we could use pelvimetry reliable further research should find out, when someone could call himself/herself an experienced observer. Based on our results we could say that pelvimetry is repeatable when done by an experienced observer.

Conclusion

These results show that measurements done with the Rice-pelvimeter by an experienced observer are repeatable. The repeatability of measurements done by an unexperienced observer is less good. Therefore, our conclusion is that the pelvimetry isn't a good on-farm tool for unexperienced farmers to measure pelvic dimensions to select their herd for larger pelvic areas.

References

Basarab, J. A., Rutter, L. M., & Day, P. A. (1993). The efficacy of predicting dystocia in yearling beef heifers: I. Using ratios of pelvic area to brith weight or pelvic arear to heifer weight. *Journal of Animal Science*, *71*(6), 1359–1371.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *327*(fig 1), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Donkersgoed, J. Van, Ribble, C. S., Booker, C. W., Mccartney, D., & Janzen, E. D. (1993). The Predictive Value of Pelvimetry in Beef Cattle. *Canadian Journal Veterinary Research*, (57), 170–175.

Funnell, B. J., & Hilton, W. M. (2016). Management and Prevention of Dystocia. *Veterinary Clinics of North America - Food Animal Practice*, *32*(2), 511–522. https://doi.org/10.1016/j.cvfa.2016.01.016

Gibbons, J. M., Lawrence, A. B., & Haskell, M. J. (2011). Consistency of flight speed and response to restraint in a crush in dairy cattle. *Applied Animal Behaviour Science*, *131*(1–2), 15–20. https://doi.org/10.1016/j.applanim.2011.01.009

Grandin, T. (1993). Behavioral agitation during handling of cattle is persistent over time. *Applied Animal Behaviour Science*, *36*(1), 1–9. https://doi.org/10.1016/0168-1591(93)90094-6

Halloway, D. R., & Johnston, D. J. (2003). Evaluation of flight time and crush score as measures of temperament in Angus cattle. *AAAABG*, *15*, 261–264.

Hickson, R. E., Morris, S. T., Kenyon, P. R., & Lopez-Villalobos, N. (2006). Dystocia in beef heifers : A review of genetic and nutritional influences. *New Zealand Veterinary Journal*, *54*(6), 256–264. https://doi.org/10.1080/00480169.2006.36708

Holm, D. E., Webb, E. C., & Thompson, P. N. (2014). A new application of pelvis area data as culling tool to aid in the management of dystocia in heifers. *Journal of Animal Science*, *92*(5), 2296–2303. https://doi.org/10.2527/jas2013-6967

Johnson, S. K., Deutscher, G. H., & Parkhurst, A. (1988). Relationships of pelvic structure, body measurements, pelvic area and calving difficulty. *Journal of Animal Science*, *66*, 1081–1088.

Kolkman, I., Hoflack, G., Aerts, S., Murray, R. D., Opsomer, G., & Lips, D. (2009). Evaluation of the Rice pelvimeter for measuring pelvic area in double muscled Belgian Blue cows. *Livestock Science*, *121*, 259–266. https://doi.org/10.1016/j.livsci.2008.06.022

Murray, R. D., Cartwright, T. A., Downham, D. Y., Murray, M. A., & De Kruif, A. (2002). Comparison of external and internal pelvic measurements of belgian blue cattle from sample herds in belgium and the united kingdom. *Reproduction in Domestic Animals*, *37*(1), 1–7. https://doi.org/10.1046/j.1439-0531.2002.00327.x

Petrie, A., & Watson, P. (2013). *Statistics for Veterinary and Animal Science* (3rd ed.). John Wiley & Sons, Incorporated.

Price, T. D., & Wiltbank, J. N. (1978a). Dystocia in cattle: A review and implications. *Theriogenology*, *9*(3), 195–219.

Price, T. D., & Wiltbank, J. N. (1978b). Predicting dystocia in heifers. *Theriogenology*, *9*(3), 221–249.

Short, R. E., Bellows, R. A., Staigmiller, R. B., & Carr, J. B. (1979). Multiple linear and nonlinear regression analyses of factors causing calving difficulty. *Theriogenology*, *12*(3), 121–130.

Van Donkersgoed, J., Ribble, C. S., Townsend, H. G. G., & Janzen, E. D. (1990). The usefulness of pelvic area measurements as an on-farm test for predicting calving difficulty in beef heifers. *Can Vet J*, *31*, 190–193. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480770/pdf/canvetj00076-0048.pdf