The transect method: a novel approach to on-farm welfare assessment of commercial turkeys

Joanna Marchewka,* Inma Estevez,*,† Giuseppe Vezzoli,‡ Valentina Ferrante,§ and Maja M. Makagon‡[1]

*Neiker-Tecnalia, Animal Production, Vitoria-Gasteiz, Spain; †IKERBASQUE, Basque Foundation for Science, Bilbao, Spain; ‡Purdue University, Animal Sciences, 125 S. Russell St., W. Lafayette, Indiana, USA; and §Università degli Studi di Milano, Milan, Italy

ABSTRACT Currently, no animal-based protocol for on-farm welfare assessment of commercial turkeys is available. The birds' size and flighty nature make obtaining a representative sample using traditional methods difficult. The transect walks (TW) approach provides a potential alternative for on-farm assessments of turkey welfare. We compared the TW approach with a traditional method, and data collected as the birds were moved out of the house during the load out process (L). Ten commercial 19- to 20-week-old Hybrid turkey flocks were evaluated (1 flock/house/farm). Half of the flocks were housed on farms deemed as "faring well" by the company, the other half were on "suboptimal" farms. Each house was subdivided longitudinally into 4 transects. Two observers walked the transects in random order, recording the total number of birds per transect that were immobile; lame; aggressive towards a mate; interacting with humans; with visible head, vent, or back wounds; engaging in mounting behaviors; small; featherless; dirty; sick; terminal; or dead. Flocks were re-evaluated on the same day using the individual sampling method (S), where randomly selected birds were scored as they took 10 steps. Flocks were re-assessed within 48 h of the transect evaluation, as birds were funneled out of the house during load out. Using ANOVAs we determined the effects of observers, method, management, and their interactions on proportions of turkeys per house within each category. Outcome parameters were not affected by management ($P > 0.05$ for all) or observer ($P > 0.05$ for most), but an assessment method effect was detected ($P < 0.05$). S differed from the 2 other methods ($P < 0.05$) for most parameters except aggression towards a mate, back wounds, dirty, sick, and vent wounds. Differences were not detected between data collected using TW and during L, except for dead ($P = 0.0007$) and immobile ($P = 0.007$). Results suggest that the TW method is a promising tool for on-farm turkey welfare assessment as it produced results similar to those obtained at L when all birds could be scored individually.

Key words: commercial turkey, transect walks, welfare indicators, load out, meat poultry

2015 Poultry Science 94:7–16
http://dx.doi.org/10.3382/ps/peu026

INTRODUCTION

The lack of effective and efficient protocols for the evaluation of commercial turkeys impedes the ability of turkey producers to evaluate the effects of management practices on bird productivity and welfare, or to provide stakeholders with science-based assurances as to the welfare status of flocks. The ability for producers to monitor bird welfare can have important impacts on their economic revenue. Welfare-related issues, including leg and mobility problems, and aggression toward other turkeys, have been cited as major causes of economic loss for this industry (Krautwald-Junghanns et al., 2011). The development and validation of universal, reliable, quantitative, and easy-to-apply methodologies for on-farm turkey welfare assessment is a critical step toward monitoring the incidences of, understanding the causes of, and formulating remedies for these types of concerns.

Available science-based welfare assessment protocols for other meat poultry (e.g. Welfare Quality Assessment Protocol for Poultry; Welfare Quality, 2009) often require the corralling and handling of birds. These types of methods are not practical for use on turkey farms as the large body sizes, heavy weights, and active and flighty nature of turkeys make their handling difficult and potentially dangerous for the birds and handlers. Methods that do not require the handling of birds

Received August 3, 2014.

Accepted September 12, 2014.

[1]Corresponding author: Purdue University, Animal Sciences, 125 S. Russell St. W. Lafayette, IN 47907-2042, USA. Tel: +001 765.496.8816; FAX: +001 765.494.9347; Email: mmakagon@purdue.edu

Table 1. Description of the birds' behavior and appearance in each of the welfare indicator categories. Birds meeting any of the descriptors within a category were counted as belonging to that category. Individual turkeys could be classified as belonging to more than one category.

Indicator	Description
Immobile	Bird not moving when approached, or after being gently touched. Birds are only able to move by propping themselves up on their wings.
Lame	Bird walks with obvious difficulty. One or both legs are not placed firmly on the ground. Bird is moving away from the observer but stopping after 2 to 3 paces to rest. Bird has shaky leg syndrome.
Head wounds	Bird has visible marks on the head, snood, beak, or neck related to fresh or older wounds.
Back wounds	Bird has visible fresh or older, including bleeding, wounds on the back and/or wings.
Vent wounds	Bird has visible wounds around tail, or on its sides, including fresh, older, or bleeding wounds.
Aggression toward mate	Bird chases or pecks, hits, flies into, or leaps onto another bird.
Human interaction	Bird perceptibly hits human with the wings, or runs into, jumps onto, or pecks the human.
Mounting	Bird mounts another bird.
Dirty	Very clear and dark staining of the back, wing, and or tail feathers of the bird, not including light discoloration of feathers from dust, covering at least 50% of the body area.
Featherless	Missing feather on the majority of the back area, including the wings.
Small	Easily distinguishable females or individuals that are approximately $\frac{1}{2}$ the size of an average bird in the flock.
Sick	Bird showing clear signs of impaired health, including a small and pale comb, red-watery eyes, and disarranged feathers. These birds are usually found in a resting position. Also birds with pendulous crops. Birds with a pendulous crop hanging in front of the breast, with missing or deformed body parts (excluding birds with leg deformations accounted for as lamed), or with pale/yellowish body color.
Terminal	Bird with large wounds or lying on the ground with head rested on the ground or back, usually with half-closed eyes. Bird must be breathing visibly.
Dead	Dead

provide a more feasible option. For example, Dawkins et al. (2004) proposed a method for the assessment of broiler gait that relies on the visual inspection of a subsample of birds as they take 10 steps. However, a limitation of this and similar approaches is that they are time consuming and allow only for the assessment of a relatively small proportion of the flock, particularly given the relatively large sizes of turkey flocks. This could lead to skewed flock level estimates of incidences of welfare issues.

Currently available information about the state of commercial turkey flocks has been obtained mainly from animal-centered protocols focused on the condition of the birds at the slaughter plant (dead on arrivals, condemnations, bruising, etc.), both pre- and post-mortem (e.g., St-Hilaire et al., 2003). This information does not necessarily represent the on-farm welfare condition of the birds as it is confounded by the effects of loading, transportation, and lairage. Additionally, it does not provide information about many of the welfare categories that can be assessed on-farm.

The TW approach, an assessment method that has recently been deemed practical for evaluating the welfare of broilers (Marchewka et al., 2013a), may resolve many of the challenges associated with the assessment of large turkey flocks. The TW method is based on line transect methodology, a technique routinely used in ecological studies to estimate animal biodiversity and abundance (Burnham et al., 1980; Butler et al., 2006). In short, an assessor walks the house along predetermined paths counting the incidences of birds' representative of predefined welfare indicator categories (Table 1). The method requires no animal handling and allows for the visual assessment of the entire flock. An additional strength of the approach is that it bears similarity to the daily poultry flock checks conducted by farmers, and is therefore easy to adopt.

The overall goal of this study was to examine the suitability of the TW method for the on-farm assessment of turkey welfare. We compared the results and interrater reliabilities of turkey welfare assessments made using the TW method to those made using 2 other approaches: an individual scoring approach modified from Dawkins et al. (2004), and the individual scoring of turkeys during L. The assessments focused on large toms at the end of the production cycle, as we hypothesized that welfare-related issues would be most evident within this production group, and because assessments made at the end of the production cycle would be most likely to be consistent with the evaluation made during L. We additionally compared the results of the 3 assessments with the company's perception of how the flocks on the farms would be faring based on past production data.

MATERIALS AND METHODS

The study was conducted between September 23 and November 18, 2013 on 10 commercial turkey tom farms located in the midwestern United States. One flock per farm was included in the study. The farms were selected based on the focal flock's age at the time of the study, and the company's opinion as to whether the farm fared well (5 farms) or needed improvement (5 farms) with regard to bird performance. The research team was blind as to each farm's classification until data collection was completed.

Facilities and Animals

All farms belonged to a single turkey company and were managed using standardized protocols. Hybrid turkey flocks were raised in grow-out houses from 6 wks of age until the end of the production cycle, which took place at approximately 19 to 20 wks of age when the birds reached an average BW of 20 kg. The stocking density at the beginning of the production cycle was between 3.5 and 3.6 birds/m^2. The turkeys were originated from 5 breeder flocks. They were raised on wood shavings and/or rice hulls. All houses had mesh windows on the sides of the buildings, were equipped with automatic drinkers and feeders, and with either manually or automatically controlled ventilation systems. Natural light, which entered the house through the windows, was supplemented with artificial lighting for a total of 23 h of light per day. Flock management information is summarized in Table 2.

With the exception of 1 house, which measured 15.3 × 152.4 m, all of the study houses had identical dimensions (12.5 × 152.4 m). House measurements were confirmed using laser range finder (BOSCH GLM 825, Stuttgart, Germany). For the purposes of data collection, each house was divided longitudinally into 4 transects. Transects were approximately 3 m wide, and their widths were limited by the feeder and drinker lines, or the wall and adjacent drinking line (Figure 1). The presence of these physical barriers hindered the birds' movement between adjacent transects as they walked away from approaching humans.

Data Collection

Each of the 10 flocks was evaluated at 19/20 wks of age by 2 observers using the TW and S methods within the same day. A final flock evaluation, L, took place during the load-out process, when birds were moved out of the house and loaded onto transportation trucks, which occurred within 48 h of the initial evaluation period. All 3 evaluation methods included the same welfare indicator categories, which are presented in Table 1. The duration of each evaluation, and cumulative mortality data were obtained from flock records at 19/20 wks of age are shown in Table 2.

Two observers assessed each flock simultaneously, but independently, to allow for a subsequent evaluation of interobserver reliability. Both observers had previous experience conducting poultry welfare assessments, but had limited experience working with turkey flocks. Whereas 1 of the observers had previous experience

Table 2. Total number of birds placed, management details, cumulative mortality calculated up to 19/20 wk of age, and duration of each of the data collection procedure (L, TW and S) is listed for each focal house.

Farm	Total nr of birds placed/house	Management	Drinker	Litter	Antibiotic use allowed	Light	Cumulative mortality-19/20 wk (%)	Load time (h)	Transect time (min)	Individual time (min)
1	6545	optimal	bell	rice hull	no	incandescent	24.0	3.5	28.3	30
2	6660	suboptimal	bell	rice hull + wood shavings	yes	incandescent	11.4	5	35.1	27
3	6485	optimal	nipple	rice hull + wood shavings	yes	incandescent	13.3	4.5	35.8	32
4	6460	optimal	bell	rice hull + wood shavings	no	incandescent	20.2	3.5	36.6	29.5
5	6660	suboptimal	bell	rice hull	yes	incandescent	13.0	5	46.6	40
6	6560	suboptimal	nipple	rice hull + wood shavings	no	compact fluorescent	18.8	3.5	26.9	32
7	6560	optimal	bell	rice hull + wood shavings	no	compact fluorescent	11.4	3	25.8	34
8	8462	optimal	nipple	rice hull + wood shavings	yes	incandescent	11.4	7.5	43.0	35
9	6502	suboptimal	bell	rice hull + wood shavings	yes	compact fluorescent	15.9	4	29.4	31
10	6660	suboptimal	bell	rice hull + wood shavings	yes	incandescent	10.9	4.5	32.6	45
Mean							15.03	4.40	34.01	33.55
SE							1.44	0.41	2.17	1.69

Figure 1. Observer during a TW data collection. The transects are limited by the drinkers (left) and feeder line (right).

evaluating broilers using TW, the other was new to this methodology. Prior to the onset of data collection the observers were provided with on-farm training during which they walked through a house with a producer and discussed examples of turkeys that they deemed to be representative of each animal welfare indicator category.

Transect Walks (TW) The TW approach for on farm welfare evaluation of turkey flocks was based on methodology previously described for broilers (Marchewka et al., 2013a). The 2 observers walked the length of each transect recording all observed incidences of birds falling into any of the predefined welfare indicator categories. The order in which transects were walked were selected randomly, except that sequential observations of contiguous transects were avoided to account for double counting of birds. Observers moved slowly in order to minimize disruption to the flock during scoring (Figure 1).

Individual Scoring (S) The individual scoring method was adapted from Dawkins et al. (2004). One hundred and four randomly-chosen turkeys, 26 birds per transect from 2 random locations along it, were evaluated. Each bird was followed visually as it took up to 10 steps, and then scored using the predefined welfare indicator criteria (Table 1). In order to ensure that both observers were evaluating the same bird, a laser pointer was used to identify the focal birds. With his or her eyes closed, 1 observer fixed the pointer on a spot. If the pointer indicated a bird, that turkey was assessed by both observers. The procedure was repeated if no bird was present in the indicated spot. The turkeys seldom appeared to notice the spot indicated by the pointer, most likely because it appeared for a very brief period of time.

Load Out Evaluation (L) During L, consecutive batches of 40 to 50 birds were herded into a corridor made out of wooden panels that funneled the turkeys towards the loading belt. The corridor was divided by a middle panel separating the birds into 2 groups of approximately 20 to 25. Turkeys were individually evaluated as they walked toward the loading belt and past the observers who stood at the sides of the corridors. The data collected in this fashion was considered to be the "gold standard," as it provided observers with the opportunity to assess each bird within a flock from a close distance. The same indicator categories were considered as for the other 2 methods (Table 1), with the exception of interactions with humans as the observers were separated from the birds by a wooden barrier. Once the majority of birds were moved out of the house, the observers walked the house, recording mortalities and evaluating any birds that may have been left behind.

Slaughter Plant Data Data collected routinely at the slaughter plant was acquired for each focal flock. The following indicators were used in the current study: livability; condemned: DOA, whole, parts; age at slaughter; weight gain per day; and average weight gain.

Statistical Analysis

Incidence of welfare indicators were calculated for each flock, therefore the analyses were conducted with house as the experimental unit (10 houses total). Prior to statistical analysis, the recorded frequency data on the numbers of individuals counted within each welfare indicator category were transformed into proportions per transect based on the known flock population in each house at the time of assessment. It was assumed that turkeys distributed randomly throughout the house.

In order to meet normality and homogeneity of residual variance assumptions, all variables were arcsine square root transformed. An independent mixed-model repeated-measures ANOVA was performed for each of the 14 welfare indicator categories using the PROC MIXED procedure in SAS 9.3 (SAS Institute Inc., 2011). The model included method of assessment, observer, and producer assigned management level (faring well or suboptimal) as fixed factors, and the interactions between observer by method and management, as well as management by method. Farm nested within the management category was included as a random statement, as well as its interactions with method, observer, and the 3-way interaction with method and observer. Least squares mean differences were adjusted for multiple comparisons using the post-hoc Tukey test. As turkeys were not able to interact with assessors during loading, the model for this indicator variable was run only for the other 2 methods.

Spearman correlations calculated using the PROC CORR script in SAS 9.3 (SAS Institute Inc., 2011) software were used to test the relationships between all variables collected during TW, S and L, data collected at the slaughter plant, and cumulative mortality levels reported to 19/20 wks of turkey age.

RESULTS

Effects of fixed factors and interactions included in the analysis of variance on all evaluated indicators are summarized in Table 3.

Welfare assessment remained consistent across observers for all indicators (Table 4). Interaction between observer and method did not affect the incidences of the welfare indicators with the exception of immobility as evaluated by TW. Nonetheless, the significant differences across observers, which ranged between 0.74% ± 0.2% vs. 0. 75% ± 0.2% for the incidence of immobile birds, were very small in numerical terms.

The effects of assessment method on welfare indicator outcomes are presented in Table 5. As compared to the TW and L methods, the S method yielded higher estimates of the incidence of lame turkeys and turkeys with head wounds, and lower estimates of the incidence of featherless, terminal and dead birds, and birds engaging in mounting behaviors. The percentages of sick birds evaluated using TW or S differed from the percentage calculated during L assessments. All 3 methods differed in the detected percentages of immobile and dead birds.

The results of the Spearman correlations analyses between welfare indicators as collected by TW and L (excluding interaction with humans, which was not observed during L), cumulative mortality levels reported at 19/20 wks and the parameters collected at the slaughter plant are presented in Table 6. We did not find

Table 4. Mean values (±SEM) of incidence of birds within each welfare indicator category expressed as percentages for each observer.

Indicator	Observer 1 (%) Mean (SEM)	Observer 2 (%) Mean (SEM)
Immobile	0.741 (*0.161*)	0.752 (*0.162*)
Lame	6.428 (*1.314*)	6.371 (*1.244*)
Aggression toward mate	0 (*0*)	0.001 (*0.001*)
Mounting	0.009 (*0.003*)	0.011 (*0.004*)
Human interaction	0.253 (*0.072*)	0.154 (*0.053*)
Head wounds	3.416 (*1.001*)	3.491 (*1.027*)
Back wounds	0.350 (*0.077*)	0.291 (*0.056*)
Vent wounds	0.139 (*0.052*)	0.140 (*0.052*)
Small	0.921 (*0.016*)	0.899 (*0.158*)
Featherless	0.024 (*0.009*)	0.022 (*0.007*)
Dirty	0.069 (*0.032*)	0.1 (*0.044*)
Sick	0.445 (*0.0105*)	0.397 (*0.097*)
Terminal	0.032 (*0.009*)	0.032 (*0.008*)
Dead	0.168 (*0.041*)	0.168 (*0.041*)

Table 3. Effect of observer, method, management, and their interactions for all scored welfare indicators.

Indicator		Analysis of variance factors					
		Observer	Method	Management	Observer*Method	Management*Observer	Management*Method
Immobile	F	4.33	30.64	1.3	4.33	1.33	0.65
	P-value	*0.071*	*<0.0001*	*0.288*	*0.0292*	*0.2823*	*0.5361*
Lame	F	0.06	32.06	0.53	0.07	0.02	0.15
	P-value	*0.8095*	*<0.0001*	*0.488*	*0.9369*	*0.8822*	*0.8653*
Aggression toward mate	F	2.5	2.5	2.5	2.5	2.5	2.5
	P-value	*0.1526*	*0.1136*	*0.1526*	*0.1102*	*0.1526*	*0.1136*
Mounting	F	1	5.35	0.54	1	1	1.73
	P-value	*0.3466*	*0.0167*	*0.4843*	*0.3874*	*0.3466*	*0.2086*
Human interaction (excluding L)	F	2.87	34.39	0.04	1.26	0	0.15
	P-value	*0.1289*	*0.0004*	*0.8436*	*0.2905*	*0.9998*	*0.7094*
Head wounds	F	1.72	21.17	0.32	0.67	0.15	0.79
	P-value	*0.2267*	*<0.0001*	*0.5871*	*0.5223*	*0.7119*	*0.4715*
Back wounds	F	1.4	1.18	0	1.97	0.38	1.51
	P-value	*0.2703*	*0.3335*	*0.9613*	*0.1686*	*0.5552*	*0.2507*
Vent wounds	F	0.34	0.63	0.01	0.35	0.7	1.13
	P-value	*0.5759*	*0.544*	*0.9432*	*0.7083*	*0.4271*	*0.3479*
Small	F	0.36	0.24	0.68	0.74	0.02	0.09
	P-value	*0.5657*	*0.7895*	*0.432*	*0.4906*	*0.8852*	*0.9144*
Featherless	F	0.04	6.65	0.07	0.04	0.78	0.85
	P-value	*0.8406*	*0.0079*	*0.7949*	*0.9568*	*0.4039*	*0.4443*
Dirty	F	0.7	0.42	0.5	1.16	1.16	1.72
	P-value	*0.4268*	*0.6646*	*0.5006*	*0.335*	*0.3122*	*0.2097*
Sick	F	1.32	13.07	4.23	1.49	1.31	3.33
	P-value	*0.2832*	*0.0004*	*0.0738*	*0.2509*	*0.2857*	*0.0619*
Terminal	F	0.09	11.11	1.45	0.09	1.25	0.65
	P-value	*0.7714*	*0.0009*	*0.2635*	*0.9141*	*0.2957*	*0.537*
Dead	F	1	65.36	1.26	1	1	1.93
	P-value	*0.3466*	*<0.0001*	*0.2942*	*0.3874*	*0.3466*	*0.1772*

Table 5. Mean values (±SEM) of incidence of birds within each welfare indicator category expressed as percentages for each method.

Indicator	*Method*					
	Transect (%)		Individual sampling (%)		Load out (%)	
	Mean	SEM	Mean	SEM	Mean	SEM
Immobile	0.60[a*]	*0.11*	0.19[b]	*0.09*	1.45[c]	*0.24*
Lamed	2.36[b]	*0.34*	12.74[a]	*1.92*	4.10[b]	*0.67*
Aggression toward mate	0.002	*0*	0	*0*	0	*0*
Mating	0.02[a]	*0.01*	0[b]	*0*	0.01[a]	*0*
Human interaction	0.31[a]	*0.05*	0.10[b]	*0.07*	N/A	N/A
Head wounds	1.16[b]	*0.15*	7.50[a]	*1.81*	1.70[b]	*0.31*
Back wounds	0.22	*0.02*	0.38	*0.13*	0.35	*0.05*
Vent wounds	0.05	*0.01*	0.29	*0.10*	0.08	*0.01*
Small	0.59	*0.06*	1.35	*0.31*	0.79	*0.06*
Featherless	0.04[a]	*0.01*	0.00[b]	*0.00*	0.03[a]	*0.01*
Dirty	0.07	*0.02*	0.14	*0.08*	0.04	*0.01*
Sick	0.05[b]	*0.01*	0.50[b]	*0.16*	0.71[a]	*0.09*
Terminal	0.03[a]	*0.00*	0[b]	*0*	0.06[a]	*0.01*
Dead	0.14[b]	*0.02*	0[c]	*0*	0.37[a]	*0.06*

*small letter subscripts indicate P-value smaller then 0.05.

any differences between management classifications in the percentages of birds within particular welfare indicator categories. Effects of management by method, or management by observer interactions were also not significant.

DISCUSSION

Considering the large size and flighty nature of commercial turkeys, and particularly that of turkey toms, evaluation of their welfare status using strategies developed for other poultry species, which typically require the corralling or handling of birds, can be challenging. In addition to being inefficient with respect to time requirements, these methods are likely to disrupt the flock, and are potentially dangerous for both the turkeys and the evaluators.

The current study compared the TW approach with 2 other welfare assessment methods, which do not require that birds be handled: individual sampling of a random sample of birds and flock evaluation as birds were loaded out of the house for transportation. None of these methods had previously been evaluated for use on commercial turkey farms. The S and TW approaches were selected for this study as they have successfully been used for the evaluation of broilers (Dawkins et al., 2004; Marchewka et al., 2013a). To the best of our knowledge, this is the first time the results of a welfare evaluation conducted during the load-out process are presented for poultry. Although this approach is time consuming, and therefore not practical for everyday use, it provided us with the opportunity to sample all of the birds in the flock. In addition to comparing the welfare indicator data collected using each of the 3 methods, we examined practical aspects of the methodologies including the interobserver reliability of each method, and its time requirements.

Because a validated set of welfare indicators is not available for turkeys, we evaluated the birds based on welfare indicators developed for other species of meat poultry (Welfare Quality, 2009) and based on a review of available literature for turkeys (Marchewka et al., 2013b). Several indicators were added to account for differences we expected to see between turkeys and broilers based on their temperament and size, older age at slaughter, ability to reach sexual maturity, and relatively higher activity levels (Huff et al., 2003). These factors were suspected to result in higher severity of wounds, especially on the head and back, higher frequency of interactions with flock mates, and higher incidence of sexual behaviors, which has previously been noticed in turkey flocks (Marchewka et al., 2013b). Human-oriented behaviors, such as running into or jumping onto the observer, hitting the observer with the wings, or pecking at the individual, were broadly categorized human directed interactions, as the motivation (e.g., aggression, defense, curiosity) was not clear. The welfare indicator categories were developed based on preliminary observation of turkey toms, as only male flocks were included in this study. It is possible that additional or different indicators may be appropriate to use in female flocks (Martrenchar et al., 2001; Huff et al., 2007).

We found no differences between observers, as well as for observer and method interaction, except for the proportion of immobile birds as evaluated using the TW method. Differences obtained for immobile birds across observers were similar to the results obtained earlier for broilers by using TW (Marchewka et al., 2013a). In both studies although differences across observers for immobility were significant, the actual difference in values were minimal. In this study, the numerical differences between observers were only 0.01%, which is too small to be considered as a major constraint of the TW approach. Therefore the results suggest that the

Table 6. Significant ($P < 0.05$) Spearman correlations between indicators collected by TW, during L, the duration of each method, and mortality levels at 19/20 wks.

		Load out														Transect walks														Sampling									Slaughter information							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Load out	Mortality (1)																																													
	Immobile (2)	-0,35																																												
	Lame (3)	-0,37	0,96***																																											
	Human interaction (4)	0,29	-0,29	-0,41																																										
	Mounting (5)	0,41	-0,03	-0,08	0,43																																									
	Head wounds (6)	0,03	0,09	0,10	-0,52	0,03																																								
	Back wounds (7)	-0,16	0,48	0,49	-0,41	0,15	0,08																																							
	Vent wounds(8)	0,35	-0,28	-0,21	-0,17	0,07	-0,37	0,05																																						
	Small (9)	-0,62	0,37	0,50	-0,41	0,06	0,03	0,59	0,02																																					
	Featherless (10)	0,19	-0,39	-0,55	0,31	0,50	-0,14	-0,10	0,30	-0,24																																				
	Dirty (11)	-0,07	0,34	0,50	-0,41	-0,05	0,20	0,30	-0,23	0,46	-0,54																																			
	Sick (12)	-0,38	0,71**	0,71**	-0,41	-0,03	0,65*	0,38	-0,69**	0,39	-0,47	0,51																																		
	Terminal (13)	-0,21	0,55	0,63	-0,18	0,12	-0,13	0,61	0,08	0,77**	-0,35	0,52	0,36																																	
	Dead (14)	-0,33	0,50	0,58	-0,06	0,20	-0,02	0,58	-0,20	0,81**	-0,36	0,51	0,50	0,93***																																
Transect walks	Immobile (15)	-0,22	0,89**	0,85**	-0,17	-0,03	0,07	0,42	-0,18	0,24	-0,46	0,04	0,56	0,44	0,39																															
	Lame (16)	-0,31	0,89***	0,85**	-0,29	-0,02	0,10	0,53	-0,09	0,36	-0,36	0,02	0,56	0,48	0,43	0,98***																														
	Agression towards mate (17)	0,05	0,04	-0,03	0,58	0,04	-0,49	0,16	-0,35	-0,03	-0,05	0,16	-0,03	0,32	0,39	-0,07	-0,15																													
	Human interaction (18)	0,47	-0,56	-0,66*	0,17	-0,01	0,10	-0,18	-0,12	-0,69**	0,38	-0,19	-0,31	-0,72**	-0,67*	-0,59	-0,62	0,16																												
	Mounting (19)	0,42	-0,44	-0,58	0,46	0,74**	-0,05	0,00	0,19	-0,26	0,9***	-0,42	-0,40	-0,28	-0,23	-0,46	-0,40	0,09	0,46																											
	Head wounds (20)	0,24	0,20	0,24	-0,41	0,23	0,89***	0,18	-0,22	0,03	-0,27	0,18	0,60	0,01	0,08	0,31	0,31	-0,51	-0,05	-0,07																										
	Back wounds (21)	-0,12	0,62	0,53	0,41	0,36	-0,26	0,16	-0,20	0,05	-0,10	-0,25	0,24	0,23	0,27	0,78**	0,72	0,13	-0,45	-0,03	0,03																									
	Vent wounds (22)	0,21	-0,22	-0,18	0,06	0,27	-0,32	0,02	0,85**	0,16	0,33	-0,44	-0,58	0,20	0,02	0,00	0,09	-0,34	-0,40	0,26	-0,11	0,11																								
	Small (23)	-0,38	0,13	0,26	-0,17	0,23	-0,01	0,48	0,20	0,83	-0,19	0,09	0,14	0,62	0,67*	0,25	0,35	-0,16	-0,68*	-0,11	0,16	0,21	0,49																							
	Featherless (24)	0,21	-0,25	-0,46	0,53	0,50	-0,17	-0,09	0,09	-0,44	0,83**	-0,79**	-0,39	-0,48	-0,42	-0,13	-0,09	0,02	0,40	0,82**	-0,15	0,31	0,25	-0,20																						
	Dirty (25)	-0,27	0,43	0,48	0,29	0,12	-0,21	-0,01	-0,57	0,33	-0,43	0,62	0,44	0,45	0,58	0,22	0,12	0,53	-0,36	-0,32	-0,18	0,33	-0,48	0,07	-0,42																					
	Sick (26)	-0,47	0,56	0,60	-0,52	-0,24	0,68*	0,21	-0,56	0,30	-0,56	0,25	0,85**	0,08	0,21	0,59	0,59	-0,40	-0,32	-0,55	0,67*	0,24	-0,43	0,21	-0,36	0,16																				
	Terminal (27)	0,03	0,18	0,07	0,17	-0,52	-0,30	-0,18	-0,35	-0,52	-0,19	0,06	0,01	-0,23	-0,27	0,05	-0,05	0,54	0,41	-0,28	-0,43	-0,01	-0,60	-0,77**	-0,07	0,23	-0,12																			
	Dead (28)	-0,38	0,42	0,47	-0,52	0,01	0,61	0,53	-0,35	0,68*	-0,25	0,51	0,77**	0,62	0,71*	0,24	0,33	-0,02	-0,39	-0,20	0,49	-0,14	-0,22	0,44	-0,44	0,21	0,55	-0,31																		
Sampling	Immobile (29)	0,09	0,44	0,44	-0,17	0,37	0,09	0,52	0,00	0,09	-0,09	0,09	0,26	0,00	0,00	0,52	0,52	-0,25	0,00	0,05	0,35	0,52	-0,04	0,17	0,18	0,00	0,35	-0,17	-0,09																	
	Lame (30)	-0,30	0,83**	0,81**	-0,29	-0,34	0,10	0,24	-0,25	0,08	-0,59	0,06	0,55	0,21	0,16	0,93***	0,88**	-0,15	-0,45	-0,66*	0,26	0,62	-0,18	0,03	-0,25	0,18	0,67*	0,27	0,13	0,44																
	Human interaction (31)	0,52	-0,17	-0,17	-0,17	0,00	0,44	-0,17	-0,17	-0,61	-0,09	0,18	0,09	-0,62	-0,61	-0,17	-0,26	-0,25	0,7*	0,05	0,44	-0,26	-0,48	-0,61	0,00	-0,17	0,17	0,26	-0,26	0,38	0,00															
	Head wounds(32)	0,24	-0,11	-0,07	-0,23	0,37	0,76*	0,10	-0,09	0,01	-0,05	-0,07	0,32	-0,19	-0,07	0,10	0,12	-0,59	0,05	0,16	0,90	0,02	0,09	0,30	0,09	-0,34	0,51	-0,65*	0,27	0,39	0,02	0,39														
	Back wounds (33)	0,03	0,42	0,49	-0,26	-0,35	-0,15	-0,04	0,38	-0,04	-0,51	-0,09	-0,03	0,15	-0,06	0,64*	0,60	-0,39	-0,51	-0,62	0,12	0,39	0,38	0,12	-0,31	-0,13	0,25	0,03	-0,23	0,25	0,72*	-0,05	0,00													
	Vent wounds (34)	-0,04	0,42	0,34	0,51	0,20	-0,65*	-0,19	0,11	-0,11	0,04	-0,27	-0,19	0,12	0,04	0,49	0,42	0,16	-0,42	-0,04	-0,42	0,8**	0,27	-0,04	0,27	0,34	-0,19	0,19	-0,49	0,22	0,42	-0,33	-0,42	0,47												
	Small (35)	-0,35	-0,21	-0,06	0,18	0,02	-0,34	-0,14	0,12	0,37	-0,26	-0,09	-0,24	0,04	0,13	-0,01	-0,03	-0,17	-0,41	-0,23	-0,18	0,22	0,32	0,63	-0,13	0,18	0,05	-0,46	-0,24	0,13	-0,03	-0,31	0,11	0,21	0,24											
	Dirty (36)	0,39	-0,06	0,11	-0,17	-0,13	-0,22	-0,16	0,42	-0,16	-0,46	0,36	-0,28	-0,06	-0,23	0,02	-0,08	-0,25	-0,02	-0,39	0,01	-0,08	0,13	-0,08	-0,46	0,07	-0,10	0,04	-0,45	0,31	0,13	0,43	0,00	0,56	0,16	0,31										
	Sick (37)	0,07	0,15	0,14	0,19	0,82**	0,35	0,19	-0,33	0,20	0,19	0,15	0,39	0,03	0,23	0,12	0,12	-0,13	-0,05	0,44	0,48	0,39	-0,12	0,28	0,29	0,30	0,28	-0,52	0,22	0,56	-0,09	0,19	0,59	-0,36	0,04	0,16	-0,13									
Slaughter information	Liveabilty (38)	-0,95***	0,49	0,49	-0,29	-0,46	0,08	0,16	-0,48	0,58	-0,30	0,13	0,55	0,31	0,43	0,36	0,42	0,04	-0,52	-0,52	-0,13	0,16	-0,31	0,30	-0,30	0,35	0,56	0,09	0,52	-0,17	0,43	-0,52	-0,25	0,01	0,04	0,16	-0,47	-0,12								
	Condemned (39)	-0,58	0,82**	0,78**	-0,29	-0,20	0,13	0,26	-0,36	0,26	-0,38	0,02	0,61	0,14	0,18	0,84**	0,84**	-0,24	-0,48	-0,51	0,20	0,66*	-0,23	0,15	-0,09	0,24	0,75*	0,10	0,19	0,52	0,89***	-0,09	0,05	0,49	0,42	0,12	-0,06	0,16	0,62							
	DOA (40)	-0,08	0,58	0,41	0,41	0,21	0,00	-0,05	-0,43	-0,22	-0,01	-0,38	0,36	0,01	0,09	0,71**	0,63*	0,13	-0,26	0,00	0,15	0,86**	-0,11	-0,12	0,40	0,23	0,33	0,21	-0,06	0,26	0,63	-0,13	0,04	0,26	0,65*	-0,13	-0,34	0,27	0,22	0,63						
	Whole (41)	-0,25	0,89***	0,85**	-0,17	0,07	0,15	0,37	-0,25	0,24	-0,41	0,07	0,62	0,35	0,35	0,98***	0,95***	-0,16	-0,56	-0,40	0,38	0,81**	-0,06	0,24	-0,08	0,26	0,66*	-0,02	0,24	0,61	0,9***	-0,09	0,20	0,58	0,49	0,03	0,02	0,30	0,36	0,89***	0,73**					
	Parts (42)	-0,72*	0,82**	0,85**	-0,52	-0,37	0,21	0,27	-0,38	0,50	-0,52	0,37	0,73	0,35	0,37	0,68*	0,71**	-0,21	-0,59	-0,69*	0,16	0,35	-0,34	0,20	-0,46	0,40	0,77**	0,10	0,47	0,26	0,77**	-0,17	-0,08	0,38	0,19	0,06	-0,04	0,03	0,79**	0,88***	0,33	0,72*				
	Av GW (43)	0,66*	-0,49	-0,59	0,41	0,27	-0,02	-0,56	0,04	-0,76*	0,51	-0,12	-0,41	-0,57	-0,60	-0,60	-0,67*	0,13	0,70	0,55	-0,13	-0,33	-0,16	-0,78**	0,37	-0,11	-0,55	0,32	-0,47	-0,26	-0,54	0,52	-0,11	-0,39	-0,04	-0,48	0,08	0,02	-0,65*	-0,60	-0,10	-0,55	-0,64*			
	Age (44)	0,69*	-0,38	-0,50	0,23	0,05	0,33	-0,43	-0,09	-0,91***	0,22	-0,38	-0,18	-0,71*	-0,7*	-0,23	-0,32	-0,08	0,73*	0,33	0,32	-0,17	-0,19	-0,71*	0,40	-0,44	-0,11	0,31	-0,39	-0,04	-0,12	0,7*	0,33	-0,07	-0,19	-0,49	0,03	-0,04	-0,61	-0,31	0,16	-0,22	-0,53	0,7*		
	Gain/day (45)	064*	-0,37	-0,43	0,41	0,25	-0,16	-0,36	0,01	-0,58	0,34	0,18	-0,33	-0,28	-0,32	-0,56	-0,65*	0,42	0,62	0,44	-0,25	-0,35	-0,26	-0,71*	0,15	0,15	-0,62	0,43	-0,35	-0,26	-0,54	0,44	-0,30	-0,44	-0,04	-0,49	0,16	-0,02	-0,61	-0,65*	-0,20	-0,55	-0,59	0,92***	0,50	

detection of important and well-defined welfare indicators in turkeys can be reliably achieved by multiple observers with minimal training.

On the other hand, differences in the data collected using the 3 assessment methods were found for all but 5 of the indicator categories. Overall, the identified differences separated the S method from the TW and L assessments. The S method yielded different results for 8 of the 13 welfare indicators evaluated during L, including the proportion of turkeys that were lame, featherless, terminal, showing head wound, exhibiting mounting behaviors, sick, immobile, or dead. In contrast, data obtained using the TW differed from L data in only 3 categories: sick, immobile, and dead. The numerical differences between TW and L were, relatively small, below 1% for all 3 differing indicators. The lack of agreement between TW and L versus S method may be explained by potential constraints of the number of birds sampled using the latter method. A sample size of approximately 100 individuals is typically suggested in available welfare assessment protocols for poultry (Welfare Quality, 2009; De Jong et al., 2012). Therefore we sampled 104 turkeys using the S method. However, sampling only 104 turkeys from a flock of approximately 4,500 may have led to high estimates for some of the indicators, as identification of a welfare indicator in just 1 bird is equivalent to a 0.96% increase in the incidence of that indicator at the flock level, and to low estimates of other indicators, as only 2.3% of the entire flock is evaluated. The sampling scheme may have further accounted for the underestimation of the incidence of immobile and dead individuals. Individuals in these categories were typically lying on the ground, obstructed from view by standing turkeys, and were therefore unlikely to be randomly selected for sampling using the laser pointer method employed in this study. These limitations in the sample obtained using the S versus TW method, in combination with the relatively large sizes of turkeys versus broilers, may explain why the differences found between TW and S for turkeys were smaller than ones reported for broilers (Marchewka et al., 2013a).

The proportions of sick, immobile and dead turkeys were highest when evaluated during L than using either of the other methods. This likely reflects the increased visibility of these birds during L, as sick, lame, and immobile individuals were likely to be left behind after the load out process was completed, and could easily be counted by the observers. These birds would have likely been sitting on the ground, obstructed from view by surrounding turkeys during S and T assessments. Additional factors that may have contributed to the increased number of observed immobile and dead birds include bird fatigue and flock disruption that occurs during the load out process, as well as differences in the timing of the 2 evaluations as L was carried out up to 48 h later than TW (Buchwalder and Huber-Eicher, 2003; Marchewka et al., 2013b). The difference in the incidence of sick birds could have been attributed to the different views from which the birds were scored. During TW birds are typically evaluated from behind as they moved away from the observers, while during L the turkeys were scored as they moved towards or stood next to the observers. As the most frequent reason for assigning a turkey to the sick category was the pendulous crop (personal observation; earlier: Rigdon et al., 1960), it is possible that this condition was underestimated when the turkeys were observed from behind.

Additional support for the accuracy of the TW method can be obtained from the correlation analysis. Overall, a high number of strong correlations were detected with regards to the incidence of indicators scored by TW and L. The number of birds having problems in terms of immobility, lameness and head and vent wounds assessed using the 2 methods had particularly high ($r > 0.85$) degrees of agreement. Conversely, the incidence of lameness was the only category evaluated using S that correlated with evaluations made using TW and L. However, the lack of correlation in the numbers of immoble turkeys detected, supports that S may not be suitable for detecting this welfare issue.

Data collected using the on-farm assessments correlated with important production parameters collected at the slaughter facility. Bird condemnations (whole or part) correlated positively with mobility assessments made using all 3 on-farm assessment methods. Mobility assessments made using TW additionally correlated with birds found DOA. Back wounds evaluated using TW, but not S or L, were highly and positively correlated with the proportion of birds DOA, and proportion of birds condemned whole. This partially could be related with the angle from which the birds were evaluated. TW and S were conducted from behind the birds, while the birds were evaluated front and side during L. The differences between S and TW could be explained by differences in sample size, as described above. The angle of assessment is a less obvious explanation for why the proportion of condemned (total, whole, and parts) birds was correlated with the frequency of sick birds as evaluated using the TW method only. The majority of birds deemed “sick” had a pendulous crop, which one would assume would have been easier to observe from the front of the birds, as was the case during L.

Several expected correlations were noted. For example, flocks affected by lameness may be expected to be less active, which is likely the reason behind the negative correlation between the recorded amount of interactions with humans and proportion of birds affected by lameness. Similarly, as expected, high frequencies of mounting observed during TW were correlated with the proportion of featherless birds in the flock assessed using L ($r = 0.9$), but also TW ($r = 0.74$).

From the perspective of the practicality of on-farm application, not surprisingly, the L data collection method proved to be the most time consuming. The time requirements of this method were limited by the pace of the loading crews, the number of birds in the flock at the time of L, as well as their condition. The time necessary for conducting TW (34.0 ± 2.2 min) was similar to that needed for the S (33.5 ± 1.7 min), highlighting the efficiency of the TW method, which allowed for the evaluation of the entire flock versus 104 randomly-selected individuals.

No differences in welfare measures were detected between houses assigned by the company to the 2 management categories. This study focused on flocks raised in 1 geographical area, during one season within the same year, and on farms belonging to the same company and therefore using very similar husbandry protocols, and house lay outs. This overall similarity may be one reason why we were not able to detect differences in the management levels assigned by the company using any of the turkey welfare assessment methods. This result may also be due to differences in the criteria used for identifying "well-managed" farms (Botreau et al., 2007; Duncan et al., 2012). Such issues as mobility problems, dirtiness or aggression toward humans cannot be directly detected after finishing the production cycle and obtained from the performance reports. In the case of this study, where turkey carcasses were used for parts as a final product, it can be especially difficult to deduce from slaughter plant data the welfare of the birds on farm (St-Hilaire et al., 2003). The company's impression of the farms was additionally based on long-term observations of their performance. We have obtained data from 1 flock only, getting a snapshot of the current on-farm situation. It is likely that data collected from multiple flocks from each farm would have better aligned with the producer's perception, underlining the importance of developing methodology that allows for this type of systematic data to be collected in a simple manner.

Overall, evidence collected during this study indicates that the transect-based on-farm welfare assessment method for meat poultry could be a feasible method for assessing the welfare of turkeys, including large toms. The method can be a time-efficient and practical tool for turkey companies and farmers to evaluate the welfare status of the flocks. The results of such evaluations can be linked to historical and management data as well as economically important outputs. Information collected over several flocks could be used to control and predict arising welfare issues giving producers the ability to develop and implement preventative strategies. In light of the increasingly high demand among consumers for guaranteed standards in animal welfare (Barbut, 2010), the TW could also serve as an animal-based assessment during internal or external welfare evaluations or audits.

CONCLUSION

The data supports that the TW method is a reliable tool for on-farm assessment of turkey welfare. This method is practical, efficient, and easy to implement under field conditions as it resembles the techniques typically used by farmers to check on their flocks, and requires minimal training to produce reliable data when used by different observers. Importantly, for 10 of the 13 welfare indicator categories sampled during L, when all birds in the flock were individually assessed, the TW method yielded similar results. The differences that were identified, the percentage of immobile, dead and sick turkeys, may equally well have been due to the increased visibility of affected birds during the L procedure, or the impact of the L procedure on these welfare indicators.

The nonhandling S method, which was used to evaluate a random sample of 104 birds per flock, either over- or under-estimated the prevalence of many of the welfare indicators (8 out of 13) as competed to the results obtained during L. The utility of the S method seems to be constrained by the sample size and strategy. Although it is possible that the utility of this method could be improved by increasing the size of the sample, given that the time required evaluating the 104 turkeys using the S methods is similar to that required for the entire house to be assessed using the TW method, increasing the sample size may not be practical.

ACKNOWLEDGMENT

The authors gratefully acknowledge the EU financial support provided under the VII Framework Program for Research, Technological Development and Demonstration Activities of the project Animal Welfare Indicators (FP7-KBBE-2010-4). We are also grateful to our industry collaborator for help with data collection at the farms and constructive input during manuscript preparation. We thank Ms. Longjie Cheng and Dr. Thomas Kuczek and the Statistical Consulting Service, in the Department of Statistical at Purdue for their advice on data analysis.

REFERENCES

Barbut, S. 2010. Past and future of poultry meat harvesting technologies. World's Poultry Sci J. 66:399–410.

Botreau, R., M. Bonde, A. Butterworth, P. Perny, M. B. M. Bracke, J. Capdeville, and I. Veissier. 2007. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: A review of existing methods. Animal 1:1179–1187.

Buchwalder, T., and B. Huber-Eicher. 2003. A brief report on aggressive interactions within and between groups of domestic turkeys (meleagris gallopavo). Appl. Anim. Behav. Sci. 84:75–80.

Burnham, K. P., D. R. Anderson, and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. Wildl. Monogr. 72:3–202.

Butler, M. J., W. B. Ballard, M. C. Wallace, and S. J. Demoaso. 2006. Road-based surveys for estimating wild turkey density in the Texas rolling plains. J. Wildlife Manage. 71:1646–1653.

Dawkins, M. S., C. Donnelly, and T. Jones. 2004. Chicken welfare is influenced more by housing conditions than by stocking density. Nature 427:342–344.

De Jong, I. C., J. Van Harn, H. Gunnink, A. Lourens, and J. W. Van Riel. 2012. Measuring foot-pad lesions in commercial broiler houses. Some aspects of methodology. Anim. Welfare 21: 325–330.

Duncan, I. J. H., M. Park, and A. E. Malleau. 2012. Global animal partnership's 5-step^{TM} animal welfare rating standards: A welfare-labelling scheme that allows for continuous improvement. Anim. Welfare 21(Suppl. 1): 113–116.

Huff, G. R., W. E. Huff, J. M. Balog, and N. C. Rath. 2003. The effects of behavior and environmental enrichment on disease resistance of turkeys. Brain Behav. Immun. 17:339–349.

Huff, G. R., W. E. Huff, N. Rath, A. Donoghue, N. Anthony, and K. Nestor. 2007. Differential effects of sex and genetics on behavior and stress response of turkeys. Poult. Sci. 86: 1294–1303.

Krautwald-Junghanns, M.-E., R. Ellerich, H. Mitterer-Istyagin, M. Ludewig, K. Fehlhaber, E. Schuster, J. Berk, S. Petermann, and T. Bartels. 2011. Examinations on the prevalence of footpad lesions and breast skin lesions in British united turkeys big 6 fattening turkeys in Germany. Part I: Prevalence of footpad lesions. Poult. Sci. 90:555–560.

Marchewka, J., T. T. N. Watanabe, V. Ferrante, and I. Estevez. 2013a. Welfare assessment in broiler farms: Transect walks versus individual scoring. Poult. Sci. 92:2588–2599.

Marchewka, J., T. T. N. Watanabe, V. Ferrante, and I. Estevez. 2013b. Review of the social and environmental factors affecting the behavior and welfare of turkeys (meleagris gallopavo). Poult. Sci. 92:1467–1473.

Martrenchar, A., D. Huonnic, and J. P. Cotte. 2001. Influence of environmental enrichment on injurious pecking and perching behaviour in young turkeys. Brit. Poult. Sci. 42:161–170.

Rigdon, R. H., T. M. Ferguson, and J. R. Couch. 1960. Pendulous crops in turkeys-an anatomic and pathologic study. Am. J. Vet. Res. 21:979–986.

St-Hilaire, S., S. Arellano, and C. S. Ribble. 2003. Association between cellulitis (enlarged sternal bursa) and focal ulcerative dermatitis in Ontario turkeys at the time of processing. Avian Dis. 47:531–536.

SAS Institute Inc., SAS Campus Drive, Cary, NC .

Welfare Quality®. 2009. Welfare Quality® assessment protocol for poultry (broilers, laying hens). Welfare Quality® Consortium, Lelystad, The Netherlands.