ANIMAL WELFARE ASSESSMENT ON INTENSIVE AND EXTENSIVE PIG FARMS

Memòria presentada per Déborah Temple

Per accedir al grau de doctora dins del programa de doctorat de producció animal
del Departament de Ciència Animal i del Aliments
de la Universitat Autònoma de Barcelona

Bellaterra, Abril del 2012

FACULTAT DE VETERINÀRIA
DE BARCELONA

XAVIER MANTECA VILANOVA, Catedràtic del Departament de Ciència Animal i dels Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona,
ANTONI DALMAU BUENO, Investigador al Institut de Recerca i Tecnologia Agroalimentàries i
ANTONIO VELARDE CALVO, Cap del Subprograma de Benestar al Institut de Recerca i Tecnologia Agroalimentàries

Certifiquen:

Que la memòria titulada "Animal welfare assessment on intensive and extensive pig farms", presentada per Déborah Temple per optar al grau de Doctor en Veterinària, ha estat realitzada sota llur direcció i, considerant-la acabada, autoritzen la seva presentació per tal que sigui jutjada per la comissió corresponent.

I perquè consti als efectes oportuns, signen el present certificat a Bellaterra,

el 12 d'Abril del 2012

Dr. Antonio Velarde Calvo   Dr. Antoni Dalmau Bueno   Dr. Xavier Manteca Vilanova

The present studies are part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508.

La autora de esta tesis ha disfrutado de una beca FPU del Ministerio de Educación de España

AGRAÏMENTS

Avui estic convençuda que el camí més recte no sempre és el millor, que els desitjos més forts es poden realitzar, que persones molt especials són les que et fan avançar.

Maman, Papa - vous êtes tout simplement exceptionnels. Merci d'avoir toujours été optimistes, visionnaires, et pour m'avoir incitée à prendre des décisions parfois difficiles. Vous aviez raison me disant que l'année au Mexique resterait une expérience inoubliable, que la "selectividad y su nota de corte" ne serait pas insurmontable, que la Facultat de Veterinària de la UAB m'ouvrirait ses portes. Vous m'avez donné le gout des voyages, de l'effort et de l'endurance. Sans vous, cette grande aventure n'aurait même pas commencé.

Vivre dans différent pays, découvrir différentes cultures, apprendre des langues est une expérience fantastique. Seulement, à long terme, pour la savourer pleinement il est bon de ne pas oublier sa terre. Rien ne vaut de chaleureuses retrouvailles dans l'Aveyron pour que cette région ne cesse de me fasciner. Emeline, Annelyse, Emilie – merci pour ces grands moments.

Un bon tros del camí l'hem fet junts, Família dels Veterinaris Collonuts: Jaumet, Super, Marina, Laura, Edu, Hombre, Mikel, Carlos, Eli, Maritxus, Cris, Iona, Júlia, Núria, Muns. Gràcies per fer-me sentir com a casa, pel vostre companyerisme durant tota la carrera, per ensenyar-me les meves primeres paraules en català, per haver respectat i donat suport als objectius professionals de cadascun de nosaltres. Sense cap mena de dubte us dec aquesta beca i l'orgull de ser veterinari . I dò: "som veterinaris i som collonuts".

La meva estima per Catalunya també me l'ha transmés la meva família de les terres de ponent. Miquel, Lolita, Miquel-Àngel, Meritxell, Andrès i Blai – gràcies pel gran suport i el carinyo que heu aportat al desenvolupament no tant sols d'una tesi, sinó de dues.

He gaudit dels estudis i això gràcies a alguns professors a qui els hi importa de debò l'exerci de la docència universitària. Gràcies per haver-me encoratjat a seguir endavant amb la recerca. També vull agrair al personal del Servei de granges de la UAB amb qui vaig gaudir treballant amb animals, i a tots els companys de producció i nutrició per compartir els seus coneixements i per fomentar l'intercanvi d'idees i cultural entre el tercer i el primer pis.

Escrivint el punt final de la tesi et dones compte que en el fons, el doctorat no tindria cap sentit sense l'equip humà que l'envolta. El cos d'aquest doctorat el formen uns individus apassionats per l'etologia i residents entre el Despatx V0-135 i la Polivalent. Tres d'ells són directament responsables que hagi posat un peu en una consulta d'etologia clínica, cosa que em semblava inicialment impensable. Marta – gràcies per ser una persona directa, per trobar la paraula que toca al moment adequat, i per prendre decisions de forma ràpida i coherent sense donar-hi mil voltes. Susana – a parte que todo se vuelve rebonito y relindo trabajando contigo, te agradezco tu paciencia, tu buen humor y por disfrutar de las cosas sencillas. Tomàs – gràcies per dir les coses com són, per queixar-te quan t'has de queixar i per acabar rient-te de situacions que semblarien muntanyes per altres. Gràcies als tres per fer que aquest despatx desbordi de positivisme i d'històries extravagants. Sempre va bé compartir amb un company proper les diferents etapes d'un doctorat al mateix temps. Pol – sempre has trobat el moment òptim per "picar-me" i engendrar baralles de nens petits, fins i tot al mig d'un congres. Gràcies pel teu sentit del humor, per les teves conviccions i les teves reflexions. En algunes ocasions, un doctorant pot arribar a dubtar de sí mateix, del que està fent, de les seves prioritats. Aquestes preocupacions no arriben a durar més d'un segon quan estàs rodejat per persones com el Pepe i l'Eva. Tots dos preneu el temps necessari per preguntar i per escoltar. Pepe – la teva alegria i empatia fan que els dubtes es solucionin en un moment. Eva – has estat pacient amb mi, amb les meves preguntes, amb els meus perquès. Em sento molt afortunada que una persona com tu m'hagi guiat i aconsellat durant tot aquest doctorat. Amb altres persones, hem compartit moments més puntuals però no menys importants: gràcies a Arantxa y sus pollos, Ceci y sus ovejas, Núria i la seva estadística, Maider y sus pollitos, Pedro y sus quejas, Gemma i els seus PT1r1/pT1r3, Vale y sus historias, Elia i els seus diagnòstics. Finalmente, agradezco a Alessandro, JP, Greta, Victor y Javi por el enriquecimiento ambiental y el buen humor que nos llevaron durante sus respectivas visitas.

Les estàncies a centres estrangers posen la cirera al pastís. They allowed me to have a broader view of my PhD and work with different domestic species. But, above all, those stays gave me the huge opportunity to collaborate with amazing researchers whose immense work, perseverance and strong character will provide undoubtedly a new impulse to the understanding of animal behaviour and welfare. Ainsi donc, je tiens à remercier à Alain Boissy pour consolider ma motivation à continuer dans le monde de l'éthologie et du bien-être animal. Un grand merci aussi à Véronique Deiss pour sa grande amabilité, son dynamisme sur le terrain et pour m'avoir totalement intégrée à l'étude. A huge thanks to Dominique Blache, Joy and Kids for their hospitality, friendship, sense of humor and for keeping and sharing their culture in an isolated Australia. Gostei muito da minha estadia no Brasil, cercada por pessoas agradáveis, sinceras e dispostas a ajudar a qualquer momento. Muito obrigada Mateus Paranhos da Costa, Adriano, Aline, Tiago, Désirée, Fernanda, i Paola. Tenho certeza que algum dia nossos caminhos voltem a se cruzar.

Quiero pensar que, en el seno del Ministerio de Educación, quedan personas que creen en la investigación y a las que les queda un mínimo de respeto hacia los becarios que contratan. A esas personas, les envío mi profundo agradecimiento.

A la pràctica, els estudis presentats en aquesta tesi no s'haguessin ni tant sols desenvolupats sense la participació i l'ajuda d'una sèrie de professionals. Vull donar les gràcies a tots els grangers i associacions per la seva cooperació cordial i per facilitar l'aplicació d'aquest protocol. Gracias a Paloma Ureta y al Pepe Pumariño por su disponibilidad y por su apoyo en el conocimiento del cerdo Ibérico y de sus mejores jamones; al Jaume Jaume per fer-me descubrir el món del Porc Negre Mallorquí i de sa sobrassada; al Joan i a la Maite per guiar-me en les visites a granges de porc blanc. Vull agrair al Joel González i a l'Emma Fàbrega per orientar-me en la caracterització d' aquests diferents sistemes de producció. Gràcies a l'equip d'epidemiologia del CReSA per atendre pacientment les meves nombroses preguntes de principiant. Un grand merci à Valérie Courboulay pour sa collaboration et son aide lors de la fusion des données. Finalment, vull agrair a la Júlia Lacuesta per ser tant amable i competent fent que la temible fase final de "papeleo" sigui menys estressant.

Dirigida per tres directors, una tesi podria ser un embolic absolut. Tot i això quan aquests tres directors es tenen una gran admiració, un infinit respecte i fan ús d'un gran sentit del humor, es pot convertir en una experiència d'allò més entretinguda i enriquidora. Antonio Velarde, Antoni Dalmau i Xavier Manteca – seguint els mateixos principis però conservant els vostres criteris respectius heu sapigut formar un triangle ón he trobat un perfecte equilibri. Tots tres sou uns directors excel·lents, eficaços, immensament treballadors, humils i sobretot sou molt bones persones. Toni Velarde – espero haver après de la teva capacitat per fer relacions públiques, per transmetre les teves idees amb entusiasme, per seguir alegre i de bon humor tot i l'acúmul de projectes i més feines que portes. Toni – t'agraeixo el teu pensament crític i recoltzament desde les granges fins al punt final. Aquestes pàgines demostren que dos persones, molt tossudes, sempre poden acabar posant-se d'acord. Xavi – ets una persona que pensa en els seus doctorants abans de pensar en sí mateix. Gràcies per la teva sinceritat, la teva confiança i per donar valor a l'esforç.

Gerard – El que més em complau és haver compartit i gaudit aquesta Gran Aventura amb tu. Amb moltes ganes i il·lusió espero les que encara han d'arribar.

RESUM

L'objectiu d'aquesta tesi fou avaluar i discutir l'aplicació pràctica, la validesa i la repetibilitat d'un sistema d'avaluació del benestar en granges comercials de porcs d'engreix. La capacitat d'aplicació, la sensibilitat, selectivitat, validesa contextual i repetibilitat al llarg del temps de les mesures basades en l'animal incloses al protocol Welfare Quality® es van avaluar mitjançant cinc estudis. A més a més, al llarg de la tesi es va considerar la utilització del protocol com a eina d'assessorament per resoldre problemes de benestar. En el primer estudi es presenta una anàlisi descriptiva de les mesures basades en l'animal incloses al protocol d'avaluació del benestar. En aquest treball es va avaluar tant l'aplicació pràctica del protocol en 30 granges intensives com la seva capacitat per discriminar-les entre sí. Tanmateix, el treball proposa una metodologia pràctica per identificar granges que presenten algun problema de benestar. El protocol Welfare Quality® per porcs d'engreix en granges es pot desenvolupar fàcilment i requereix poca intervenció per part del pagès. Tot i així, requereix una mitjana de 6 hores i 20 minuts per cada visita. La variabilitat entre granges en la presència d'animals amb bursitis moderades i greus, porcs moderadament bruts (<50% del cos brut), expressió de comportaments socials positius i negatius, i l'exploració, permet discriminar granges intensives entre sí. Tot i això, altres mesures basades en l' animal, especialment les que estan relacionades amb el principi de bona salut, presenten una variabilitat massa baixa per poder diferenciar significativament les granges entre sí. Mitjançant models Generalitzats Mixtes es van extrapolar límits de confiança a nivell de població per cadascuna de les mesures estudiades. Aquest mètode pot ser útil per identificar granges que presenten algun problema de benestar determinat en relació a una població establerta. Al segon i tercer estudi, es van avaluar un total de 11,647 porcs allotjats en 91 explotacions comercials provinents de tres sistemes intensius (convencional a França i Espanya, sobre palla a França, i Ibèric intensiu a Espanya) així com dos sistemes extensius (Ibèric en extensiu, Porc Negre Mallorquí) mitjançant el protocol Welfare Quality®. Tot dos estudis descriuen dades sobre la prevalença i distribució de diversos indicadors de benestar en una gran varietat d'explotacions amb la finalitat d'identificar una sèrie de factors causals. Els resultats corresponents als principis de bona alimentació i bon allotjament es presenten en el segon estudi mentre que els de salut es descriuen en el tercer. Basant-se en les mesures d'alimentació i allotjament es poden establir diferències significatives entre els sistemes de producció. En canvi, les mesures de salut no es diferencien de manera significativa entre els cinc sistemes de producció estudiats. En aquests treballs, s'identifiquen una sèrie de factors causals potencials per les bursitis moderades i greus, els porcs bruts, i l'apinyament. El impacte d'aquests factors causals varia en funció del sistema de producció estudiat. En referència al sistema intensiu convencional es mostren factors causals potencials per als animals amb ferides severes, les mossegades de cues i les coixeses. La presa d'informació sobre factors de l'ambient (com ara el tipus de menjadora, tipus de terra, densitats, etc.) així com característiques dels animals (fase d'engreix, estatus de castració)

pot ajudar a la detecció de granges que tenen més probabilitats de presentar algun dels problemes de benestar esmentats. El quart estudi es centra en l'avaluació del benestar mitjançant mesures de comportament en porcs Ibèrics en condicions intensives i extensives. En aquest treball, el quart principi del Welfare Quality®, titulat "comportament adequat", es va avaluar en 21 granges (11 extensives i 10 intensives), les quals allotjaven una totalitat de 25,856 porcs Ibèrics. Els comportaments socials negatius i positius així com l'avaluació qualitativa del comportament són sensibles a canvis a l'allotjament i maneig de porcs Ibèrics. En canvi, no es troben diferències significatives en l'expressió del comportament exploratori i en la relació humà-animal entre els porcs Ibèrics criats en extensiu i en intensiu. La freqüència elevada de comportament social negatiu registrada en condicions intensives és un indicador de falta de benestar. Tot i així, les variacions trobades en la freqüència de comportament social positiu i d'exploració no són tan fàcils d'interpretar. L'avaluació qualitativa del comportament sembla útil per a discriminar les explotacions basant-se en l' expressió del comportament natural. Finalment, el cinquè estudi avalua, en una mostra de 15 granges intensives convencionals, la repetibilitat després d'un temps perllongat de les mesures basades en l'animal incloses en el protocol Welfare Quality®. Els resultats es van discutir des d'un punt de vista discriminatiu i avaluatiu. Des d' una perspectiva discriminativa, tant sols la brutícia moderada al cos i la relació humà-animal són consistents al llarg d'un període de temps perllongat. Tanmateix, s'observen algunes inconsistències entre el coeficient de correlació de Spearman i el coeficient de correlació intra-classe a l'estudiar la repetibilitat d'algunes mesures (ex. animals amb lesions severes o animals molt bruts). L'avaluació qualitativa del comportament no és consistent al llarg del temps excepte si es refereix únicament a la posició relativa d'una granja en una sola visita. Per altra banda, una aproximació avaluativa com és la dels "límits d'acord" pot ser útil per a detectar granges amb problemes de benestar persistents.

Com a comentari general, el protocol Welfare Quality® per porcs d'engreix es pot aplicar fàcilment en una gran varietat d'explotacions comercials. Tot i així, la durada total del protocol pot ser percebuda com a massa llarga per part d'alguns dels agents implicats (com ara ramaders, industria o empreses certificadores). La sensibilitat dels indicadors de salut és aparentment massa baixa. La interpretació d'alguns resultats en termes de benestar animal, especialment els de comportament, s'ha de realitzar amb molta cura sobretot quan es comparen explotacions provinents de sistemes de cria molt diversos. Una interpretació equivocada d'un resultat pot erròniament afavorir o perjudicar un ramader o un sistema. Finalment, el protocol Welfare Quality® pot tenir enfocaments diversos que poden ser pràctics i complementaris com a eina d'assessorament, tot i que es va dissenyar com una eina per auditar.

SUMMARY

The objective of this thesis was to evaluate and discuss the feasibility, validity and repeatability of a welfare assessment system for growing pigs on commercial farms. The feasibility, sensitivity, selectivity, contextual validity and repeatability over time of several animal-based measures of the Welfare Quality® protocol were evaluated in five studies. Furthermore, across the thesis, the use of the protocol as an advisory tool was considered. The first study presents a descriptive analysis of several animal-based measures of the welfare assessment protocol for pigs kept under intensive conventional conditions on 30 farms in Spain. It evaluates the feasibility of the protocol on intensive farms and its capacity to discriminate among intensive conventional farms. It also aims to propose a practical methodology to identify farms with a particular welfare problem. The overall Welfare Quality® protocol for growing pigs on farm is easy to perform on intensive conventional farms and requires little input from the farmers; however, it takes an average time of 6 hours and 20 minutes per visit. The between-farm variability of moderate and severe bursitis, moderately soiled body, expression of social "positive" and negative behaviours and exploration allows discrimination among intensive conventional farms. Nonetheless, other animal-based measures, especially the ones related to the good health principle, have too low variability to differentiate among farms. Confidence limits were inferred at population level for each animal-based measure by means of Generalized Linear Mixed models. This methodology can be useful to identify farms with poor welfare in relation to a given population. In the second and third studies, a total of 11,647 pigs housed on 91 commercial farms of three intensive systems (conventional in France and Spain, straw bedded in France, and Iberian intensive in Spain) as well as two extensive systems (Iberian extensive, Mallorcan Black pig) were assessed applying the overall Welfare Quality® protocol. Both studies yield data about the prevalence and distribution of several welfare outcomes in a wide variety of commercial farms with the aim to identify possible causal factors. The results from the good feeding and housing principles are presented in the second study whereas those of health are given in the third study. Important differences between production systems can be established based on animal-based measures of the feeding and housing. In contrast, health measures do not differ significantly between the five production systems studied. Several possible causal factors of moderate and severe bursitis, moderately and severely soiled body and huddling were identified and their respective impact varied according to the production system. In the intensive conventional system, several possible causal factors for severely wounded animals, tail biting and lameness were identified. The recording of simple environmental-based factors (such as the feeding system, the type of floor, the space allowance, etc.) as well as characteristics of the animals (growing stage, castration status) can be useful to detect farms that are more likely to show one of the previous welfare concerns. The fourth study focuses on the assessment of welfare through behavioural measures in Iberian pigs in intensive and extensive conditions. The fourth principle of the Welfare Quality® protocol,

labelled "Appropriate behaviour", was assessed on 21 farms (11 extensive and 10 intensive) housing a total of 25,856 Iberian pigs. Changes in occurrence of behaviour and qualitative measures are evaluated and discussed when comparing Iberian pigs either in intensive or extensive rearing conditions. Negative and "positive" social behaviours as well as the qualitative behaviour assessment are sensitive to changes in the housing and management of Iberian pigs; in contrast, no differences are found between extensive and intensive reared Iberian pigs in the expression of exploratory behaviour and in human-animal relationship. The high occurrences of negative social behaviour recorded in intensive conditions are clearly an indicator of poor welfare; however, interpretations of variations in the frequencies of "positive" social behaviour and exploration are not straightforward. The qualitative behaviour assessment appears useful to discriminate farms on the basis of the expression of natural behaviour. Finally, the fifth study evaluates the test-retest repeatability over a long period of time of the mean prevalence of several measures from the Welfare Quality® protocol in a sample of 15 intensive conventional farms of growing pigs. The results are discussed using both a discriminative and an evaluative approach. From a discriminative perspective, only moderately soiled body and the human animal relationship showed a high consistency over a long period of time. However, inconsistencies are found between Spearman's correlation coefficient and intraclass correlation coefficient when interpreting the repeatability of several quantitative measures such as severely wounded animals and severely soiled body. Unless the assessment related only to the relative ranking of a farm within the same visit, the qualitative behaviour assessment is not consistent over time. Adopting an evaluative approach such as the limits of agreements can be useful to detect farms that present persistent welfare problems.

As a general comment, the Welfare Quality® protocol for growing pigs is feasible in a wide range of commercial conditions even though the overall duration may be perceived as too long by stakeholders. The sensitivity of health indicators appears to be low. Interpretation of several outcomes, especially behavioural ones, in terms of animal welfare, must be cautious, especially when comparing farm units from diverse rearing systems. An erroneous interpretation of an outcome can over or underrate a farmer. Several approaches of the Welfare Quality® protocol can be practical and complementary to be used for advisory purposes.

INDEX

GENERAL INTRODUCTION

CHAPTER 1. APPLICATION OF THE WELFARE QUALITY® PROTOCOL TO ASSESS GROWING PIGS KEPT UNDER INTENSIVE CONDITIONS IN SPAIN

CHAPTER 2. THE WELFARE OF GROWING PIGS IN FIVE PRODUCTION SYSTEMS: ASSESSMENT OF FEEDING AND HOUSING

CHAPTER 3. THE WELFARE OF GROWING PIGS IN FIVE PRODUCTION SYSTEMS: ASSESSMENT OF HEALTH

CHAPTER 4. ASSESSMENT OF ANIMAL WELFARE THROUGH BEHAVIOURAL PARAMETERS IN IBERIAN PIGS IN INTENSIVE AND EXTENSIVE CONDITIONS

CHAPTER 5. ASSESSMENT OF TEST-RETEST REPEATABILITY OF ANIMAL-BASED MEASURES ON GROWING PIG FARMS

GENERAL DISCUSSION

GENERAL INTRODUCTION

1. ANIMAL WELFARE

1.1. Importance of animal welfare

Animal welfare did not arouse, originally, as a science. Instead, ethical concerns from the society arouse the scientific interest in animal welfare. Since the 1960s, there has been an increasing interest of the effects of intensive production on animal welfare. The publication of *Animal Machines: The New Farming Industry* by Ruth Harrison in 1964 produced a huge public reaction. This social concern has led to initiatives by governments and international bodies, having a great impact both on the consumers/marketing and the farmers (Sandøe and Christiansen, 2008). Besides an increasing preoccupation for the respect of the animals, the improvement of welfare is known to positively affect the health and disease resistance, as well as the food quality and safety (Jones and Boissy, 2011). With the proliferation of societal interest in animal issues has come philosophical literature on animal welfare that makes sometimes the border between science and ethic hazy. Still, the process of welfare assessment and ethical judgement should be well differentiated (Broom, 1991a). Ethical decision on what should be considered as acceptable or not for the animal will be taken after welfare has been scientifically measured.

1.2. Concept of animal welfare

Before measuring animal welfare, an agreement should be reached on how should be actually defined animal welfare. Three main approaches, with their own strengths and weaknesses, are generally accepted.

A first approach links welfare to the absence of negative emotions and the presence of positive ones. Concerns for animal welfare are generally based on the assumption that non-human animals can subjectively experience emotional states and hence can suffer or experience pleasure (Dawkins, 2000; Mendl and Paul, 2004; Boissy et al., 2007a). European legislation refers to animals as ‘sentient beings’ (Treaty of Amsterdam, 1997). The suffering of an individual animal is precisely what the public is concerned about and nowadays the majority of scientists accept that animal suffering, and the degree of animal consciousness that this implies, are essential aspects of animal welfare (Rushen, 2003). Several studies have achieved to measure animals motivations and emotions through behavioural and physiological changes; however, these methods are not free of problems of interpretation (Paul et al., 2005; Mendl et al., 2009). Furthermore, few studies have dealt with the animal's full range of emotional experiences over a long period of time (months or years) (Rushen, 2003).

A second approach assesses an animal's welfare based on the presence of an adequate biological function (Broom and Johnson, 1993). The welfare of the individual is then defined as its state as regards its attempts to cope with its environment (Broom, 1986). To cope with a challenge, the organism will activate the functioning of body repair systems, immunological defences, physiological stress responses and several behavioural responses. The biological cost of these responses can ultimately affect the growth, heath and reproduction of the animal (Barnett et al., 2001). Since there are various coping strategies, welfare can vary over a wide range, from very good to very bad (Broom and Johnson, 1993). Indicators of fitness should, then, be combined together with measures that quantify the effort in coping with a challenge in order to give a more accurate evaluation of welfare (Broom and Johnson, 1993). For example, indicators such as the mortality rate and the occurrence of multifactorial diseases should be combined with behavioural (e.g. redirected injurious behaviours) and physiological (e.g. cortisol levels, immune response) indicators. However, it should be emphasized that, as it happens for the assessment of emotional states, the link between animal welfare state and indicators of biological functioning is not always straightforward (Rushen, 1991). For instance, decreased weight gain of pigs resulting from decreased space allowance (Meunier-Salaün et al., 1987) have been detected in the absence of any effects on plasma corticosteroid levels (Rushen, 1991).

Finally, some definitions are based on the concept of harmony between the animal and its environment and the consequent ability to perform normal patterns of behaviour (Hughes and Duncan, 1988). Normal behavioural pattern is usually associated with the behaviour show by most members of a species under natural conditions. Such behaviour can be compared with the behaviour displayed by animals in more confined conditions and, as it provides the basis for animals' requirements, it can be used as a welfare assessment. However, the definition of what should be considered as normal or not or even prejudicial or not for the animal is not straightforward especially when dealing with animals housed in a variety of housing conditions. It's fair to think that domestic species have undergone genetic adaption to their new niches living among humanity (Price, 1984), making adaptation in captivity/confined conditions less challenging. In addition, animals are usually very flexible and often have several alternative strategies available in their natural behaviour (Spinka, 2006). Then, there is not necessarily a welfare problem if domestic animals or animals kept in captivity do not show the same pattern of behaviour as their wild counterparts (Veasey et al., 1996). Moreover, some behavioural patterns that are clearly natural are in fact detrimental to the welfare of the animal (e.g. flight reactions). Despite the previous limitations, welfare is clearly impaired when the inability to show a particular behaviour result in pain or suffering. Such behaviours are usually known as behavioural (or ethological) needs and should be given a particular emphasis in a welfare assessment.

These apparently divergent definitions are, in fact, three complementary issues. The animal reaches the state of harmony when it is kept in an environment that allows it to satisfy its motivations. When the situation moves away from this ideal state, the animal will use a wide range of physiological mechanisms and behaviours to cope with its environment. The ability to cope successfully or not will depend on the individual itself and on how much it has to cope with. The impossibility to cope with an aversive situation may lead to the appearance of injuries or diseases (Broom and Johnson, 1993) and consequently to pain and suffering. Integrating those different approaches, an agreement is reached on what is needed to achieve animal welfare. It is generally accepted, then, that the welfare of a sentient animal is determined by its capacity to avoid suffering and sustain fitness (Webster, 2005a). In other words, welfare comprises physical and mental health (Dawkins, 2004; Webster et al., 2004) and includes several aspects such as absence of thirst, hunger, discomfort, disease, pain, injuries, stress and the expression of normal behaviour (Five freedoms of the Farm Animal Welfare Council, 1992).

1.3. Welfare in extensively kept animals

Extensive farming has been given little emphasis in animal welfare science compared to confined animals. Extensive conditions can be defined as those where the animal receives infrequent handling and require a low capital investment (Manteca and Ruiz de la Torre, 1996). The general opinion of extensive animals leaving in a “natural” environment less controlled by humans (Lynch et al., 1992) and the perception of less stress and health problems in extensively kept animals (Hemsworth et al., 1995) are two main reasons for this lack of welfare concerns in extensive systems. Extensive animals have a greater opportunity to express their normal behavioural patterns than intensively reared animals. However, despite being free to show natural behaviour, an animal may experience negative emotional states and altered biological functioning if the environmental challenges overwhelm its evolved coping strategies (Dwyer, 2009). For example, extensive animals may experience hunger in poor pastures and show changes in biological functioning associated with malnutrition. Furthermore, several studies suggest that the behavioural needs of animals are not necessarily met in all extensive environments and that some environments may be suboptimal for welfare (Dwyer, 2009). As commented previously, welfare is much more than the absence of behavioural restriction. Extensive rearing systems present particular challenges to welfare linked to the large seasonal variations (food supply, climate), predation and infrequent supervision (Petherick, 2005). Animals in extensive systems can be grazing across large areas and there is limited opportunity to frequently observe animals. Extensive systems present threats to animal welfare that can differ markedly from those in intensive systems. Therefore, despite the lack of social concern, the welfare of extensive animals has to be measured and assessed in a scientific way.

2. MEASUREMENT OF ANIMAL WELFARE

Welfare is multidimensional and it cannot be measured directly and by a single measure, rather animal welfare science is multidisciplinary (Fraser, 1995) and makes use of a great variety of parameters for the overall welfare assessment. These parameters are included into two main types: resource/management-based and animal-based measures.

2.1. Resource and management-based measures

It is possible to assess welfare looking at the environment of the animals. Resource and management-based measures can indicate if the environment is acceptable for the animals. Much legislation is based on environmental-based measures such as space allowance, type of floor, group housing, presence of perches, etc. Those indirect measures are based on assumptions concerning the relationships between aspects of the environment and the actual welfare state of the animals. Environmental-based factors are fundamental for the provision of advice on the prevention of a welfare problem and for the detection of a risk of deficient welfare. Information on the risk of welfare problems is particularly important to detect problems whose incidences are rare. Short questionnaires on management routines such as feeding or procedures such as castration, de-horning, ear tagging etc. may also be very useful in practice. According to a study on expert opinion, Whay et al. (2003a) reported that the most relevant environmental measures for pigs were: assessment of barren conditions, water and food facilities, bedding (cleanliness and thermal comfort), space allowance, ventilation, lighting, and absence of sharp edges and availability of group housing.

2.2. Animal-based measures

Instead of measuring the provision of good husbandry, welfare can be measured by observing the animal more directly. Animal-based measures involve both animal observations and the use of farm records of the animals (Whay et al., 2003b). Animal-based measures fall into four main categories: Performance (including *post-mortem* measurements), health, physiology and behaviour.

2.2.1. Performance measures

The accumulation of all attempts to cope with several challenging situations will ultimately affect the animal and thus its performance. Examination of records can give a first general approximation of the welfare state of the animals on a farm. For example, Hurnik (1988) proposed that longevity could serve as an overall assessment of welfare, since all serious challenges to animal welfare should ultimately hasten death. Performance parameters give an overview of the problems that may experience a herd over time. However, their unspecificity makes them difficult to interpret and varying widely among studies and among countries. As well, it has to be emphasised that good performances do not guarantee an optimise welfare. For example, broilers show a high growth rates but also many welfare deficiencies. Some examples of usual performance parameters are given next. Mortality, excluding culled or active euthanized animals, merely reflects the number of animals becoming so sick that they die prior to slaughter. If two animal husbandry systems are compared and the mortality rate is significantly higher in the first than in the second then we can say that the welfare is less good in the first (Hurnik and Lehman 1988, Broom 1991a). Mortality would then be a health indicator. Still, mortality rate depends of many factors such as the housing system, group size and stockmanship (Losinger and Heinrichs, 1997) resulting in an unspecific measure unless the causes of death are well known. A large list of reproductive performance parameters can be elaborated (still birth rate, litter size, abortions, farrowing rate, culling rate, weaning to oestrus interval, interval between calving, etc.). They are usually used by vets to detect significant welfare or management problems. Poor reproductive performance can be linked to stress situations as well as many other reasons that are not directly related to welfare (oestrus detection, timing of insemination, storage of semen, differences between genetic strains, etc.). Methods of calculation of these parameters can vary between different national recording schemes or commercial software packages (Almond et al., 2006). For that reason their reliability is questioned for farm assessments. Feed intake and growth rate in growing pigs (Meunier-Salaün et al., 1987), egg/milk/flee depressed production, and abnormal eggs are other performance parameters that can be useful to detect a welfare or management problem. Finally, several measures of “meat quality” can be indicators of acute stress (lairage, management at the slaughterhouse) or chronic stress (handling, transport etc). The prevalence of PSE and DFD meats are indicators of those welfare problems in pigs (Dalmau et al., 2009).

2.2.2. Health measures

Health problems remain some of the most serious welfare problems for farm animals (Scientific Veterinary Committee, 1997). Health refers to the state of the body and brain in relation to the effects of pathogens, parasites, tissue damage or physiological disorders. Since all of these effects involve pathology (that is the detrimental derangement of molecules, cells and functions that occurs in living organisms in response to injurious agents or deprivations), the health of an animal is its state as regards its attempts to cope with pathology (Broom, 2006). As the welfare of an individual is its state as regards its attempts to cope with its environment (Broom, 1986); then health is an important part of welfare. The advantage of health indicators is that their link to suffering is clearer, and therefore they are more easily validated than other types of indicators.

The appearance of injuries or diseases is an explicit indicator of poor welfare. A disease is a pathological condition characterized by a group of signs or symptoms. Some diseases are produced by a well defined pathological agent who plays by far the most important role in the incidence and severity of the disease. They generally spread easily from farm to farm and their apparition is usually explosive. Swine Fever and Aujeszky's Disease are examples of those monofactorial diseases (Straw et al., 2006). In contrast, other types of diseases are highly environmental-dependent. The impact of those diseases largely depends on the combination of several management, housing and hygienic conditions of the farm. For that reason they are generally referred as multifactorial diseases. Adrenal activity and the related immunological response that follow difficulties in coping with the environment are shown to reduce the fitness of an individual increasing the incidence of diseases (Broom and Johnson, 1993; Broom, 2006). Multifactorial diseases are considered to reflect the overall quality of the environment of the animals; hence, they should be carefully considered within a welfare assessment system. Pododermatis in broilers, mastitis and lameness in dairy cows, swine respiratory complex in fattening pigs, MMA (Metritis-Mastitis-Agalactia) syndrome in sows are some examples of multifactorial diseases.

While clinical signs of a pathology is a clear indicator of poor welfare, the health (and therefore the welfare) of an individual can be also impaired even though it does not apparently suffer from stress, injury or disease (Broom, 1991b). For instance, an agent may be present in the animal's body without having any effect on the animal and pathology is absent. As soon as the agent causes a 'derangement', the animal will have to cope with this pathology. The effort in coping may range from a light difficulty to a harmful effect on the biological functioning causing pain and suffering. In those former cases, health indicators lack of sensitivity and evaluation of welfare should rely on physiological and behavioural changes.

Despite early recognition of the importance of health for animal welfare (Broom, 1986), health problems as a source of animal welfare have been underestimated and underused (Rushen, 2003). Health indices should be carefully taken into account into a farm assessment (Rousing et al., 2001) and a special emphasis should be given to epidemiology in animal welfare research (Rushen, 2003).

2.2.3. Physiological measures

Both failure to cope with the environment and difficulty in coping are indicators of poor welfare (Broom, 1986). Diseases, injuries, impaired growth or reproduction are clear indicators of suffering and therefore of poor welfare; however, welfare can be poor without suffering and pain. The intents to cope with several challenges will induce changes in physiological responses that may be useful to evaluate the motivational state of the animal (Broom, 1991b).

Before considering physiological parameters in a welfare assessment some of their limitations should be highlighted. First of all, physiological changes are difficult to interpret and are not always associated with poor welfare. Similar changes in physiological responses can occur in situations of opposite affective states. Cortisol levels, for example, may rise in both fear inducing situations and in response to sexual activity (Rushen, 1991). Second, there is usually an important variability between individuals and some parameters may exhibit circadian pattern (Blache et al., 2011). Physiological parameters are also limited for their feasibility at field level. Finally, they usually require methods that are more or less invasive and special material for recollection. It has also to be considered that the stress of the animals during the handling may reduce the validity of the data.

Heart rate and stress hormones are main examples of physiological indicators used in animal welfare measurement. Heart rate is a fast response to changes in the environment. An increased in heart rate is associated to an active response (‘fight or flight’) while bradycardia occurs when the animals is immobile (‘freezing’). Sheep heart rate increased on exposure to a strange person/dog (Baldock and Sibly, 1990); and a strong bradycardia response in response to a sudden unpredicted stressor can be observed in low resistant pigs (characterized by a back test) (Hessing et al., 1994). As heart rate can be affected not only by a stress response but also by an increased activity, results have to be readjusted on the basis of activity levels. High environmental temperatures and circadian rhythms are other influencing factors.

Hypothalamic-pituitary-adrenal (HPA) activity with the release of catecholamines and cortisol has been widely use as indicator of stress. There is clear evidence that levels of adrenalin, noradrenalin and glucocorticoids increase as a response to acute stress. Release of glucocorticoids is slower than catecholamines and can be monitored non-invasively via metabolites excreted in urine and faeces (Wasser et al., 2000). Transport (Dalin et al., 1993; Candiani et al., 2008), shearing (Hargreaves and Hutson, 1990), surgical castration (Mellor and Murray, 1989), dehorning (Morisse et al., 1995) and mixing of unfamiliar animals (Merlot et al., 2004) are some examples of short term challenges that have been associated with an increased in HPA activity. However, how chronic stress affects the activity of the HPA axis still remains unclear. There are relevant discrepancies between research groups on how housing methods affect corticosteroid levels (Rushen, 1991).

Changes in the immune response are considered to be more reliable indicators of chronic stress response. Changes in the homeostasis are known to induce an “acute phase response” during which there is a change in the rate of synthesis and release of acute phase proteins (APP) such as haptoglobin, C-reactive protein, and serum amyloid. The quantification of APP concentration has been used to monitor inflammatory processes in farm animals (Eckersall, 1995). Recently it has been recognised that animal APP are not only useful for monitoring inflammatory processes but also for analysing various non-inflammatory conditions such as pregnancy, parturition, metabolic diseases and stress (Murata et al., 2004). The level of APP in the plasma has been demonstrated to be useful in the assessment of animal welfare. Geers et al. (2003) found a relation between haptoglobin level at slaughter on the one hand and management, housing conditions, and health during fattening on the other hand. APP concentration has been used, for example, to identify the presence of diseases (Eckersall, 2000), lesions (Pallarés et al., 2008), stress linked to long transport (Candiani et al., 2008) or to rearing conditions (Amory et al., 2007). However, it may be difficult to distinguish an increase of APP caused by the activation of the HPA axis (stress response) from an increase of APP due to a latent infection or inflammation. To improve the understanding of APP functioning and its interpretation, further studies should establish common criteria for clinical veterinarians and scientists for APP measurements (Murata et al., 2004). Finally, it should be taken into account that, as it happens for stress hormones changes, immune responses may reflect only some, and not all, types of challenge to animal welfare (Rushen, 1991; Moberg, 2000; Porges, 2002).

2.2.4. Behavioural measures

Behaviour can be a sensitive indicator of the animal’s perception of environmental changes. Variations in behavioural patterns often represent the first level of response of an animal to an aversive or stressful environment. Behaviours which differ in pattern, frequency or context from those which are shown by most members of the species in conditions that allow a full range of behaviours are usually considered ‘abnormal’ behaviours (Fraser and Broom, 1990). The development of ethograms of livestock in natural or seminatural environment gives a sight of how the animals operate in a wide variety of natural features. For instance, Graves (1984), Jensen (1986), Stolba and Wood-Gush (1989) reported behaviours of pigs in semi-natural environment. These studies have provided information used in designing new housing and management systems for pigs (Gonyou, 1994). In addition to those basic studies, other authors have compared the behaviours of extensively raised animals with the behaviour displayed by the same species in more restricted conditions (Kerr and Wood-Gush, 1987; Krohn et al., 1994; Shimmura et al., 2008). Nonetheless, interpretation of variations in behaviour is usually not straightforward and originates many discussions on what should be considered as normal or not or even prejudicial or not for the animal, especially when dealing with animals housed in a variety of systems (Rushen and de Passillé, 2009).

Behaviours clearly indicate poor welfare when associated with physical suffering. Behaviour is commonly used in the clinical assessment of health and particularly of pain. The changes in behaviour brought about by disease are used by veterinarians in the diagnosis of disease (Broom, 1987; Fraser and Broom, 1990). Detection of behavioural changes is particularly important in subclinical diagnosis and thereafter in the prevention and control of pathologies. Behaviour is also considered as the most commonly used parameter to assess pain on farm (Viñuela-Fernández et al., 2007). Examples of behaviour indicating pain in pigs include reluctance to move and a lack of normal social behaviour (Fraser and Broom, 1990). These and other signs of pain, whether related to diseases or injuries, have been considered in detail by several authors (e.g. Houpt and Wolski, 1982; Straw et al., 2006). In addition, scales to evaluate more subtle signs of pain based on behavioural changes have been validated (Dobromylskyj et al., 2000). Mainau et al. (2010), for instance, developed an ease of farrowing score in sows kept in farrowing crates based on the behaviour of the sows and their piglets. Behavioural indicators of pain represent, definitively, a practical way to assess welfare at commercial level.

In addition to pain related behaviours, other changes in duration and frequency of normal behaviour are recognized as indicators of mental suffering (e.g. frustration). Possible signs of stress include startle or defence responses, avoidance, suppression of feeding and sexual behaviour, excessive levels of aggression, stereotypic behaviour, lack of responsiveness or apathy, decreased complexity of behaviour, and the time required to resume normal activity after a stressor (Cook et al., 2000). Still, Mason and Latham (2004) suggested that variations in behaviour are not a sign of poor welfare; but only if associated to other specific symptoms, may be a warning sign of potential suffering. The link between behaviour, chronic stress and poor welfare is here again not easily quantificable. Despite the controversy on what should be considered as a ‘need’ or not and whether it is driven by internal or external factors, it is clear that preventing an animal from carrying out a certain behaviour in a given situation might cause signs of suffering (Jensen and Toates, 1993). Redirected injurious behaviours such as feather-pecking in hens or tail biting in pigs partly linked with a lack of exploratory activity are clearly abnormal behaviours that may easily lead to pain. It should be accepted then that exploration is a behavioural need and that the assessment of levels of exploratory activity as well as frequencies of redirected behaviours in hens or pigs can be used as indicators of welfare. The impossibility to display a behavioural need can also derive towards the appearance of stereotypies. An animal performing a stereotypic behaviour repeats a relatively invariant sequence of behaviours, which has no obvious function (Fraser and Broom, 1990). Sows in stalls unable to nest before farrowing that show bar-biting; tong-rolling in cows; crib-biting in horses; bar-biting in sheep are some examples of stereotypic behaviours found in domestic ungulates (Bergeron et al., 2006). Curiously no clear stereotypic behaviour has been described in growing pigs. Assessment of several behaviours can be very useful indicator and predictor of physical and mental suffering. However, before reaching the state of pain, frustration or chronic stress (failure to cope), it should be achievable to evaluate “if the animals actually have what they want” as alleged by Dawkins (2004), and thereafter to avoid more subtle causes of impaired welfare and be aware about less detectable needs for the animal. Several methodologies are focused to find a way to assess animals’ motivation and emotion at farm level.

Choice (or preference) tests and operant conditioning have been used to assess the motivation as well as to improve the comfort of the animals in a given housing (Fraser and Nicol, 2011). Preference testing is one of the most obvious and direct ways of asking what an animal prefers in its environment. A simple choice test is mainly used to ask the animals what they prefer giving them the opportunity to choose between two or more options. In operant conditioning the effort an animal makes to gain access to a resource or to perform a behaviour is measured (Dawkins, 1983). Gradually, the amount of work needed to get different resources is increased (pressing a plate, or opening a door for example). A resource ends up being ‘inelastic’ when the input from the animal still increases; in contrast, a reward is ‘elastic’ if the animal stops working or work less to obtain it. This methodology allows assessing what resources or behaviour the animal considers the most important. However, the animals can still be influenced by early experiences and they may not work for the choice that is better for their long term welfare. For example, dairy cows tend to choose wider stalls even though wider stalls can become dirtier increasing the risk of mastitis (Rushen et al., 2008). Taken these difficulties into consideration, preference tests can provide evidences of the animal’s perception of what conditions are the best for its welfare. The animal’s welfare will then be improved if it is housed under these conditions. Such example is given by Mason et al. (2001) who achieved to link welfare indicator (increased in cortisol) with motivations in minks (through price elasticity of demands). The study suggests that caging mink on fur farms does cause the animal’s frustration, mainly because they are prevented from swimming. Results from preference tests undergone in experimental or commercial conditions (Dawkins, 2004) can be useful to detect which environmental factors are perceived as important by the animal. Once validated an environmental factor can be transferable to a welfare assessment as an environmental-based measure.

Innovative methodologies to assess animal’s affect are in expansion providing a new approach of animal welfare. These methodologies expect to encompass the limitations of behavioural (and physiological) indicators of affective states. In addition to emotions, which are by definition short-lived (Dantzer, 2002), the concept of welfare involves more persistent affective states that influence the way in which the individual perceives and reacts to its environment (Lazarus, 1991). The impact of housing for example on the mental state of the animal is over a long period of time while much emotions that have been assessed such as pain and fear last for a short period of time. The question of how to integrate the animal’s full range of emotional experiences (affect) over such a long period of time represents one of the main limitations of welfare assessment (Rushen, 2003). Recent studies on the cognitive component of emotion in animals look forward to overcome those limitations. Shettleworth (2001) defines the term cognition as the information processing including attention, learning, memory and decision-making. Broom (2010a) proposes a narrower definition of cognition as: having a representation in the brain of an object, event or process in relation to others, where the representation can exist whether or not the object, event or process is directly detectable or actually occurring at the time. This definition clearly differentiates a cognitive process to "abstractions" where representation occurs on something absent. Several recent studies (Harding et al., 2004; Doyle et al., 2010) have investigated the possibility that an animal's interpretation of an ambiguous situation may be altered by its emotional state, those in negative states being the more likely to respond as if the negative outcome will occur. The influence of affect on a range of cognitive processes including attention, memory and judgement has been called cognitive bias (Mendl et al., 2009). The analysis of cognitive predispositions and cognitive biases induced by repeated emotional experiences may provide a way to better understand long-lasting affective disorders in farm animals (Boissy et al., 2007a) and have a deeper approach of positive emotions (Boissy et al., 2007b).

Innovative methodologies to measure animal's motivations and affects can provide an objective evaluation of what the animal 'want' in experimental conditions. Results from those controlled studies may give advices on what type of housing and management conditions should insure more quality of life for a given species or breed. Result may be, in a future, extrapolate to commercial conditions and provide information on which aspects of welfare the animals consider as a priority. At the time, these methodologies are not feasible enough for been used as animal-based indicator in a welfare assessment system.

An alternative approach based on observer ratings has been proposed to assess the quality of an animal's relationship with its environment (Wemelsfelder, 2007). Ratings by human observers have long been used by animal scientists and veterinarians to assess certain physical traits (e.g. body fat: Russel et al., 1969), and can also be applied to the assessment of behaviour and a variety of welfare-relevant variables (e.g. pain: Dobromylskyj et al., 2000). The use of observer rating as a methodology to evaluate the animal's experience or quality of life is referred as "Qualitative Behaviour Assessment" or QBA (Wemelsfelder, 2007). This methodology relies on the ability of human observers to integrate perceived details of behaviour, posture, and context into descriptions of an animal's style of behaving, or 'body language', using descriptors such as 'relaxed', 'tense', 'frustrated' or 'content' (Wemelsfelder, 1997). Judgments on scoring the descriptors rely on the ability of the observer for complex perception and interpretation of behaviours. The QBA is then vulnerable to various forms of bias from a single observer (e.g. influenced by personal view) or from a group of observers (e.g. cultural bias) and may open the door to an anthropomorphic projection. Its apparent 'subjectiveness' may be highly criticised and controversial. Still, several studies have demonstrated the biological validity of the QBA as an integrative methodology (Rousing and Wemelsfelder, 2006; Wemelsfelder, 2008). In a recent review, Meagher (2009) concluded that despite some challenges and limitations, qualitative observer ratings can be a legitimate and very useful scientific tool, especially in an assessment at farm level.

2.3. Animal-based vs. environmental-based measures

Since welfare is a condition of the animal, animal-based measures are likely to be the most direct reflection of its welfare state. Many interactions between environmental factors make the links between the environment and the welfare of the animals not always clearly understood (Capdeville and Veissier, 2001). Housing (floor type, drinkers, feeders, ventilation system, etc.), management (feeding, hospital pens, etc.) and human animal relationship are three main types of influencing factors on the prevalence of a welfare problem (outcome) (Waiblinger et al., 2001). In some cases, such as in pain related procedures (dehorning, tail docking, tooth clipping, etc.), the link between the management of such procedures (use or not of local anaesthetics and long lasting analgesics) and welfare is confident. In other cases, the relationship between housing, management, human animal relationship and the animal itself (age, immune system, reproductive stage, etc.) is complex. The relative importance of each factor on the appearance of a given welfare problem remains difficult to predict especially in a long term situation. For example, housing animals in groups promotes normal social interactions, but if poorly managed this system can lead to an increased in aggressive interactions (Spoolder et al., 2009). Furthermore, those complex interactions may favor one aspect of welfare and be detrimental for another aspect of welfare. As an example, Rodenburg et al. (2008) showed that laying hens in non-cage systems made better use of the resources available, were less fearful and had stronger bones than hens in furnished cages; however, the latter had lower mortality rates and provided better air quality. An epidemiologic approach through risk assessments has not been overlooked in animal welfare (Rushen, 2003; Smulders, 2009). Influencing factors may provide an estimation of the risk of impaired welfare; however the respective impact of those factors is not well understood and is not precise enough to undertake an assessment system based only on those factors. Moreover, all influencing factors have probably not been identified yet. Animal-based measures overcome the previous limitation as they evaluate the actual welfare state of the animal on-farm (Main et al., 2007). Animal-based measures (outcomes) indicate the effect of those indirect environmental-based measures and their interactions on the animal. They are generally considered more valid measures of welfare compared to environmental-based measures and should be applicable across production systems. However, animal-based measures are not free of difficulties, either. The first problem is that they are, sometimes, difficult to interpret. Some animal-based indicators such as injuries (tissue damage) are clearly linked to poor welfare. However, the link between other animal-based indicators - especially behavioural measures - with impaired welfare, is at the time not clearly understood and may require a correct validation through future studies. Second, it has to be considered that only a sample of animals is observed. This sample has to be representative of all the animals on the farm. A third limitation of most animal-based measures is that recording can be difficult whereas recording of environmental-based measures is quite easy and demands less resources (Bartussek, 2001). Environmental-based measures are also more likely to be repeatable. Environmental and animal-based measures present both advantages and disadvantages. Consequently, the combination of both types of parameters will give the most valid assessment of animal welfare (Johnsen et al., 2001) and, hopefully, enable not only to assess the current welfare state of the animals but also to evaluate potential risks to their welfare.

2.4. Challenges in measuring welfare of extensively kept animals

Extensively reared animals may differ from intensively reared animals as a result of the great heterogeneity of the environment (topography, climate, pasture quality etc.). The welfare of animals in any type of housing system will depend upon the details of the system, upon management (Rushen et al., 2008) and upon many interactions between those factors. This complexity is exacerbated in extensive systems where there is a lack of literature regarding the impact of the housing/rearing systems. Some environmental-based measures used in intensive systems are inappropriate for extensive systems; whereas other specific aspects of the environment are of major importance (handling facilities, winter housing, presence of shade, quality of the water, wallowing facilities, risk of predation, quality of the pasture, etc.). The heterogeneity in extensive environments has favoured the development of a wide diversity of local breeds suited to particular conditions. Different genetic backgrounds can affect the response to specific stressors and subsequent adaptation of a breed to its environment. Breeds of ewes differ in the quality of maternal behaviour (Dwyer, 2008) and there are breed differences in neonatal lamb vigour (Dwyer et al., 1996) and in lamb's ability to discriminate its mother (Nowak et al., 1987). Likewise, fear response of ruminants varies between breeds even when they have been reared under the same management system (Boissy et al., 2005). Genetic background of maternal behaviour and fear response has been demonstrated also in pigs (Grandinson, 2005). The lack of knowledge on the impact of detailed environmental factors on a specific breed make that the appearance of a welfare problem may be even more unpredictable in extensive conditions than in intensive ones. Animal-based measures (outcomes) should, in theory, reflect how the animals cope with this complex situation and be valid across systems. Some animal-based measures rely on the ability to closely approach the animals. This may be the main limitation of the application of animal-based measures on extensively managed animals that have generally large flight distance. Nevertheless, examination of health, injuries, body condition and fear of handling could be assessed during specific husbandry tasks, feeding, or during confinement period (Turner and Dwyer, 2007). In addition to feasibility constraints, species that are traditionally extensively managed tend to be subtle in their expression of pain, injury or distress, as they have evolved as predated species as commented by Romeyer and Bouissou (1992). Thereafter, the detection of signs of poor welfare may be more difficult in these animals. Whether the expression of pain varies between pigs kept under intensive and extensive conditions has not been reported. Finally, animal-based measures may not be accurate enough to evaluate the quality of the stockmanship, one of the main concerns in extensive husbandry.

3. COMBINATION OF MEASURES INTO PROTOCOLS: OVERALL ASSESSMENT OF WELFARE

3.1. Overall assessment system

Due to growing public concern for farm animal welfare and the demand for product labelling which assures the consumer of the welfare-friendly origin of food, the need for on-farm welfare assessment systems has significantly increased. Animal welfare is multidimensional and can be measured in many different ways. Still, none of the measures previously mentioned can give the full picture alone. Welfare must be assessed by a combination of measures (Webster, 2005b; Fraser, 2006) that can be environmental-based (resource and management) and animal-based.

It is generally agreed that if an assessment tool is to be successfully used on individual farms it must (Waiblinger et al., 2001):

- Involve measures that are reliable (repeatable), valid and precise
- Be easily operated by trained people
- Require limited time so that repeated measures on many farms are possible
- Reveal the causes of impaired welfare and thus potential improvements of the husbandry and management system

3.2. Applications of a welfare assessment system

The features and development of the welfare assessment system will greatly be influenced by its particular application. There are four broad categories of potential applications for animal welfare assessment systems (Main et al., 2003a): Research tool; Enforcement of legislative requirements; Certification schemes (voluntary); and as an Advisory/management tool. There is though a large degree of interrelationship between these categories.

On-farm welfare assessment in basic research is often designed to quantify the welfare of animals in certain husbandry systems or to compare different treatments. As a research tool, the protocol needs to be sufficiently transparent for peer review in publications but it does not necessarily need to be understood by individual farmers.

On-farm welfare assessment can be used to evaluate compliance with national or European legislation even though most legislation is currently resource-based. For example, assessment of compliance with cruelty legislation ("unnecessary suffering") requires an assessment of the animal's welfare through direct measurements. For legislative requirements, an assessment system must be clear, unambiguous, flexible and feasible to use by experienced persons.

There has been a large increase in voluntary certification schemes in some countries (Bartussek, 1999; Wood 1999). Membership of these schemes is not a legal requirement and is often associated with a marketing claim. Certification schemes often include a basic requirement to comply with relevant welfare legislation but there are often additional welfare standards that have to be assessed in a similar fashion (Main et al., 2001). Farm assurance standards have been developed for most livestock sectors including pigs (Main and Green, 2000; Main et al., 2001). The standards should be clearly definable, understandable and unambiguous for all interested parties, including the consumers. It should include all resources and husbandry practices on a farm that could affect the welfare of any individual of the species covered by the scheme (Main, 2009). As well, it should include all types of stock such as the management of cull animals.

Welfare assessment can also be used in a non-controlling framework as an advisory tool where farmers use welfare assessment to monitor welfare over time and receive advice about suggested solutions to observed welfare problems. Welfare assessment results may be reported back to the farmer with a comparison of his performance with other farmers using similar systems ('benchmarking').

3.3. Choice of measures and scoring scale

The choice of measures for an overall assessment of welfare is determined on the basis of feasibility, validity and repeatability (Winckler et al., 2003). Taking into account those properties, a scoring scale has also to be defined for each measure.

3.3.1. Validity

One of the main concerns regarding welfare assessment is the extent to which an indicator is actually measuring what it is supposed to measure. By definition validity is determined by accuracy (free from systematic errors), specificity, and scientific validity (Martin and Bateson, 2007). For instance, human-animal relationship test (HAR) has to reflect the animal's perception to human; recording of negative social behaviour should be related to situations of fear and stress; evaluation of skin lesions should estimate aggressive behaviour, etc. Validity can be checked comparing measures proposed to be taken on farms to more sophisticated measures used as references (or possible "gold standard") for animal welfare assessment in controlled experiments (criteria validity or concurrent validity). Another form of validity, named construct validity (or predictive validity) is investigated by formulating hypothesis on the relationship between welfare and some other variables (Scott et al., 2001). Drugs such as anxiolytics and consequent changes in behavioural patterns are a method to assess construct validity. For example, Rushen et al. (2007) demonstrated using local anaesthesia that pain tends to be cause of gait aberrations in dairy cows. Construct validity is also achieved when demonstrating the causal relationship between treatment (e.g. housing) and an effect (e.g. outcome welfare measure). A third type of validity, named consensus (or face) validity is a subjective judgement that a measure is thought to be valid, as judged by one or more experts (Scott et al., 2001). Whay et al. (2003a) identified outcome measures for the welfare state of several species through a consensus of expert opinion. The measures of pig welfare most frequently identified as being of importance by an expert panel are represented in **Figure 1**.

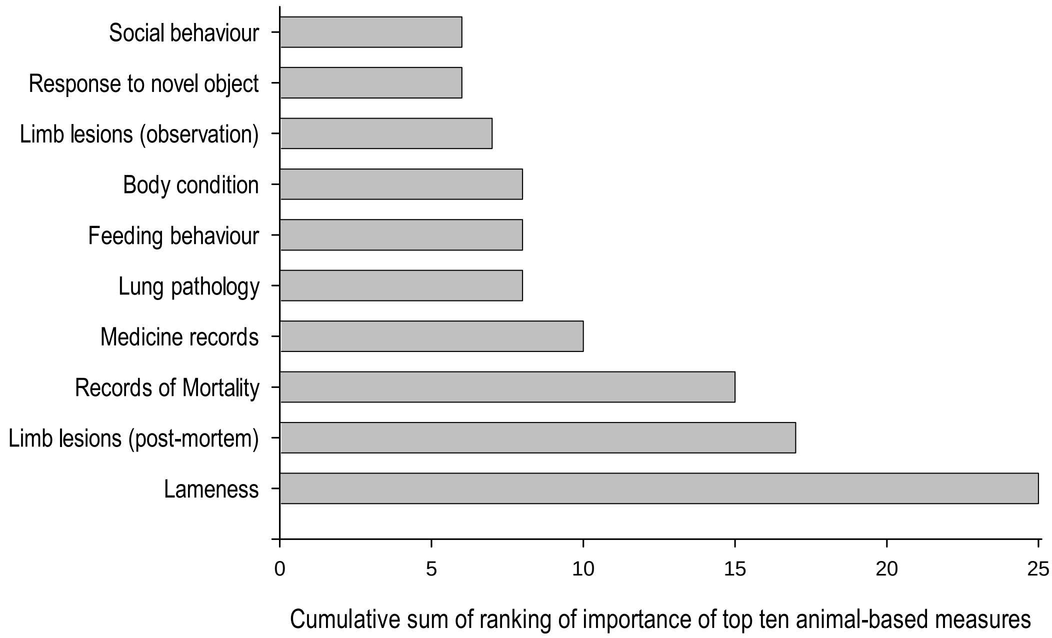


Figure 1 Ten animal-based measures of pig welfare most frequently identified as being of importance by an expert panel (consensus validity) (*Adapted from Whay et al., 2003a*)

3.3.2. Repeatability

Repeatability (or reliability) is defined as the similarity of repeated measurements on one object. Inter-observer repeatability is evaluated if two (or more) persons carrying out an observation of the same animals get similar scores. Intra-observer repeatability analyses if the same person assign similar scores to the same object. Inter and intra-observer repeatability of a measure can be whether assessed under experimental or commercial conditions. Many animal-based measures have been assessed at individual level for intra-observer or between-observer repeatability. On the majority of studies, the robust agreement between different assessors and within assessors on an animal-based measure appears not easy to achieve. In fact, repeatability represents one of the main limitations of the use of animal-based measures (de Rosa et al., 2003; Lensink et al., 2003). A valuable indicator such as the incidence of diseases or the mortality rate for example results unreliable at commercial level unless in countries with well-developed health-recording systems.

3.3.3. Feasibility

For inclusion in a welfare assessment system, a measure should be easy to carry out on commercial conditions and short time consuming. Most of the physiological measures and several behavioural indicators may be excluded for been too invasive or for requiring too much time (Smulders et al., 2006). Non invasive techniques to assess HPA axis activity such as faecal cortisol or more recently cortisol in the hair (Stalder and Kirschbaum, *in press*) may be useful to study chronic stress and welfare at herd level. Still, the cost of cortisol determination should be considered. Plesch et al. (2010) evaluated behavioural measures related to dairy cow behaviour in terms of their suitability for inclusion in an on-farm assessment. Seven of the fifteen potential measures were discarded with respect to limited observation times. According to Spoolder et al. (2003), the feasibility criterion can be particularly restrictive and can act as the main factor for exclusion of animal-based measures. Selection based on feasibility is relatively easy and is often reported (Bartussek, 1999).The identification of a feasible behavioural measure on farm may be particularly challenging.

3.3.4. Scoring scale

Welfare indicators can be quantified using different types of scales. In animal welfare science the choice lies mainly between the ordinal (categorical) and the continuous (interval) scoring scales (Scott et al., 2001). Ordinal scale has been the most frequently created scale for welfare measurement, yet it presents difficulties in analysis and interpretation. Interval measurement is created when the welfare trait is continuous rather than categorical in nature and the differences between a response and a constant is generally known (Scott et al., 2001). If the underlying attribute is continuous then the ordinal scale may lose information and its sensitivity may be compromised. A categorical measure may therefore lose validity to gain repeatability. Continuous scales have been much less studied in animal welfare except for subjective ratings that commonly use continuous Visual Analogue Scale (VAS) (Meagher et al., 2009). VAS is a measurement tool to assess a variable that is believed to range across a continuum of values and can not be measured easily (Wewers and Lowe, 1990). Originally, it was developed for rapid and easy self-assessment of pain in humans for which it has been reported to produce ratio data with high sensitivity to change and good construct validity (Willianson and Hoggart, 2005).

3.4. Aggregation methods and existing methods of welfare assessment

3.4.1. Aggregation: an overview

A protocol requires combining measures to form an overall assessment, either to describe a situation (e.g. housing systems) or to make an absolute assessment on it (checking compliance with welfare standards). The method of aggregation of measures depends greatly of the application that will be given to the welfare assessment system. Several types of evaluation models – descriptive, normative and prescriptive – can be distinguished and they play different roles in a decision process based on several measures (Bell et al., 1988; Botreau et al., 2007a). Descriptive models are use to describe a pre-existing situation that is stable and independent of any observation. This approach provides the ability to describe and compare observed situations. Normative models tell how things should be or how people should act, and aim at providing evaluation procedures to check the adequacy of observed behaviours relative to pre-defined norms. This is, for example, the case for many formal labelling processes (e.g. certification, restaurant rating, etc.). Finally, prescriptive models are used to help people make better decisions and to improve their activity. Prescriptive models aim to gather and organise relevant information so as to facilitate the construction of recommendations to reach a goal. According to Botreau et al. (2007a) the formal evaluation procedure for assessing animal welfare could be seen as both a normative and a prescriptive model. In addition to the choice of measures, a tricky and critical decision has to be taken on the relative weights assigned to each of them (Spoolder et al., 2003). As Sandøe et al. (1996) point out, the weighing of the different measures may depend not only on science but on ethical assumptions as well. Bracke et al. (2002a) proposed weighting coefficients derived from a classification of scientific evidence with an additional validation by comparison of expert opinion (Bracke et al., 2002b). The determination of the relative weights of each measure relies then on expert opinion and ethical judgement on what counts for animals and what societies find acceptable or not.

3.4.2. Existing methods of welfare assessment

Several monitoring systems have been developed in Europe, ranging from simple checklists to more complex index systems. Most of these systems are largely based on environmental-measures presumed to affect animal welfare whereas rare observations are done on the animal itself except for performance indicators. These systems prioritize thus the feasibility and repeatability of the overall methodology rather than its validity. The main differences between those protocols lies in how parameters are aggregated to obtain an overall 'scoring' of welfare. An overview of these systems - which cover welfare assessment in pigs, cattle, and poultry - is presented below and resumed in **table 1** and **figure 2**.

- **Ethical account in Denmark**

The ethical account described by Sorensen et al. (2001) is based on expert judgment and opinion on a set of data collected during experimental investigations and inspections. The prototype method aimed to provide the farmer with detailed information about welfare status on the farm. Assessments of welfare were based on information from four sources: 1) the housing system, 2) the management, 3) records of animal behaviour, and 4) records of health. The results of the welfare assessment were presented to the farmer in an annual welfare report and the information might then be used to guide the farmer in improving welfare but there was no formal aggregation of the measures (Botreau et al., 2007a, 2007b). One limitation of the ethical account is that it is based on expert opinion. Experts, especially the ones from different background, may have different point of views on the interpretation of parameters. For instance, a vet may prioritize health while an animal scientist may focus on behaviour. Furthermore the method does not allow comparing welfare levels from different farms. It does not certify either a welfare standard.

- **Systems based on minimum requirements**

Those systems are based on minimum requirements that a farm or a housing system, for example, must fulfil. A series of parameters are considered and for each of them a threshold is defined. This is the case for the Freedom Food Scheme, which was set up as a farm assurance and food labelling scheme by the Royal Society for the Prevention of Cruelty to Animals (RSPCA) in

1994. A second example is based on the application of the Hazard Analysis and Critical Control Point (HACCP) for pig housing proposed by Von Borell et al. (2001). Control points relate to certain limits that are already defined by the legislation or are based on the current knowledge of good farming practice and might not be scientific-based. The method can be easily explained and it presents a high repeatability between assessors. However, it may be too restrictive as no distinction is made between a farm that fails on only one aspect and a farm that fails on many aspects (Botreau et al., 2007a).

- **Index systems of welfare assessment**

The 'Animal Needs Index' (ANI), also referred as 'Tiergerechtheitsindex' (TGI), was developed in Austria during the 1980s into ANI 35L (Bartussek, 1999). In 1994, it was reworked in Germany as a German version, ANI-200 (Hörning, 2001; Sundrum and Rubelowski, 2001). Both systems assess the impact of the housing system on animal welfare. Index systems have been developed for assessing welfare in cattle, pigs, and laying hens, especially with regard to organic production. Assessments of welfare in very intensive housing systems, such as crates for sows and battery cages for laying hens, cannot be carried out. Environmental and management parameters constitute the main part of the index system, and only a few animal-based parameters are included in the assessment. The parameters are recorded on the farm in about an hour by specially trained inspectors. In general, the ANI system is highly practicable and highly repeatable (Hörning, 1998; Amon et al., 2001). In ANI 35L points are assigned to different areas of the housing and management: 1) possibility of movement, 2) social contact, 3) quality of floor, 4) climate, and 5) stockman care. ANI 35 L was developed as a means of certifying the level of animal welfare on farms. In the ANI 200, scores are assigned to seven different aspects of the housing and management: 1) locomotion, 2) feeding, 3) social behaviour, 4) resting, 5) comfort, 6) hygiene and 7) stockman care. In pigs, the assessment is supplemented with parameters relating to defecation and urination. ANI 200 was developed as a method for on-farm welfare assessment which would allow farms to be compared with certification purposes. It also aims to provide advice and support for farmers (Sundrum and Rubelowski, 2001). The method of aggregation (sum of scores) is discussed by Botreau et al. (2007a). In this scoring method, raw data are transformed into welfare scores on a numeric value scale independently of the sample of farms observed. Then, weights are attributed to the values obtained for the different parameters. Finally, an overall score is calculated as a "weighted mean" of the scores obtained for each parameter. This type of aggregation allows compensations between parameters (Spoolder et al., 2003). As a consequence, a farm that would obtain averages scores for all parameters can obtain the same overall score that a farm with extreme scores for different parameters.

- **Decision support system to assess welfare status in farm animals**

A method proposed by Bracke et al. (2002a, 2002b) in the Netherlands attempts to capture the known relationships between 'attributes' (parameters) in a computerised database and make them available for welfare assessment. This method is based on scientific research and focuses on welfare assessment of housing systems. It has been developed in pregnant sows in crates; however, according to the author, it can be applied to all farm animal species and any housing system. It is not clear whether this method is also sensitive to assess animal welfare at farm level (Johnsen et al., 2001). The aggregation method is an implicit sum of scores and as commented for the index systems, it allows compensation (Botreau et al., 2007a).

- **Increasing interest for assessment systems based on animal-based measures**

When comparing several of those existing protocols with animal-based measures, low correlations have been found (Johnsen et al., 2001; Mollenhorst et al., 2005). One limitation was that variations within housing systems could not be detected. Previous works investigating the impact of farm assurance schemes on dairy cattle welfare have demonstrated that defining resources alone may be ineffective in delivering high welfare standards (Main et al., 2003b). As mentioned previously, many interacting resources affect the welfare state of an animal and it can be difficult to correlate resources and the actual welfare state with any degree of certainty. Limitations of the protocols based mainly on the evaluation of the environment have led animal welfare scientists to have a different approach of welfare and focused on the identification and validation of measures based on the animals. It has been recently accepted that since the animal's welfare state depends of how well that animal is able to survive and to remain fit within the particular constraints of the husbandry system in which it lives, the assessment should be preferentially based on measures taken on the animals (Whay et al., 2003b). Many animal-based measures have been used as tools to assess one specific welfare issue at a time. Examples include plumage scoring in hens (Freire et al., 1999), skin lesion scoring in pigs (Mouttotou et al., 1999; Leeb et a1., 2001) and locomotion and lameness scoring in cattle (Whay and Main, 1999). Since 2000, reliance on animal-based measures in certification schemes has increased. Tosi et al. (2001) described several environmental and animal-based indicators applied on dairy farms in Italy. Capdeville and Veissier (2001) developed in a research project in France an on-farm assessment system of dairy cows' welfare mainly based on animal-based measures. The University of Bristol has also proposed an animal-based welfare assessment protocol for pigs, dairy cattle and laying hens in a research study assessing the welfare impact of the RSPCA Freedom Food scheme (Whay et al., 2003c). Those two innovative systems mainly based on animal-based measures are described below.

The on-farm assessment system for dairy cow proposed by Capdeville and Veissier (2001) uses a multidimensional approach based on the five freedoms (FAWC, 1992). A list of the 49 welfare indicators most commonly measured in animals (behaviours, injuries, social interactions, etc.) was produced. The aggregation method of those measures is based on a sum of scores to which *ad hoc* rules based on expert opinion are considered to limit compensation between measures (Botreau et al., 2007b). Few parameters relevant of the same domain are considered together, before being included in the evaluation of the fulfilment of a need. This required making many small sets of measures and then making many sequential aggregations. This hierarchical process limits problems of overweighting items that are assessed through numerous measures. As in the scoring systems commented before (ANI, Decision support system), weights are introduced to consider the relevance of each parameter for the fulfilment of a need. Finally a score is given to the farm for each basic need and then for the five freedoms together. One of the main limitations of this method is its opacity and its difficulty (Botreau et al., 2007a): only the expert who developed the model knows its weaknesses and strengths.

Whay et al. (2003b) proposed a much easier method of aggregation based on rankings to compare welfare between farms of calves on the basis of 19 parameters corresponding to respiratory health, nutrition and general appearance. The overall rank of a farm was the mean of the 19 partial ranks this farm obtained on each parameter. All parameters (partial ranks) were considered as having the same importance for welfare. Thresholds were established on the average values (ranks) obtained from the population of farms studied. Therefore, the score obtained by a farm depends on how this farm performs in comparison with other farms. This method enables to compare farms. It is clear and easy to understand and to standardise. It can help farmers to rank their farm among others and to become more aware of the conditions in which their animals are living. The main limitation of this method is that the overall ranking of a farm depends on the population observed. Its application as a labelling system is therefore limited (Botreau et al., 2007b).

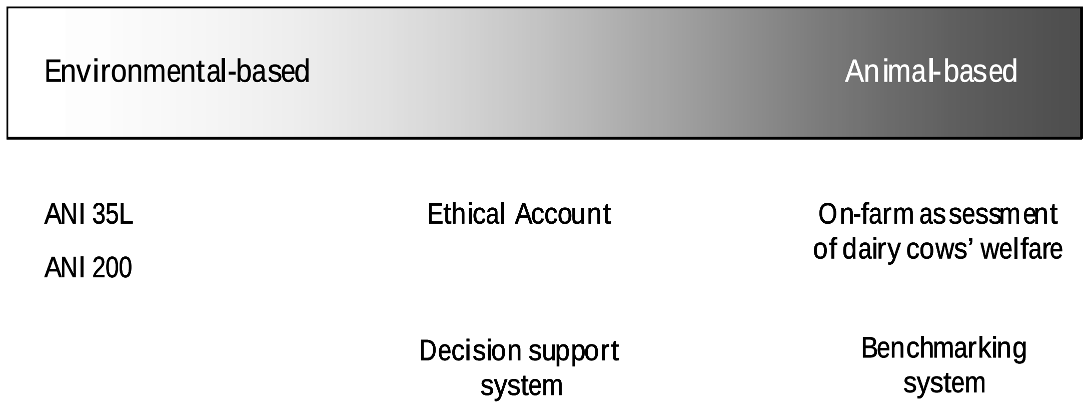


Figure 2 Combination of welfare parameters into welfare assessment system (*Adapted from Johnsen et al., 2001*)

Table 1 Main assessments of farm animal welfare at farm level (*Adapted from Johnsen et al., 2001 and Botreau et al., 2007a*).

Name of assessment	Species	Method of aggregation	Results of assessment	Potential uses	Country
Ethical account	Cattle, pigs	Non-formal aggregation based on expert's opinion	Report	Advisory tool for the farmer	Denmark
Minimum requirement	Cattle, pigs, poultry	Checklist	Yes/no answer on requirements	To check the respect for the current laws or within certification schemes based on the strict fulfilment of a series of specifications	UK Germany
ANI 35L ANI 200	Cattle, pigs, laying hens	Weighted sum of scores	Score	To compare farms. Certification or labelling schemes Implemented in legislation in Austria Used by organic organisations	Austria Germany
Decision support system	Sows (applicable to other species)	Weighted sum of scores	Score	Evaluation of housing systems Certification or labelling shemes	The Netherlands
On-farm assessment of dairy cows' welfare	Cattle	Multidemensional approach. Weihted sum of scores with *ad hoc* rules	Score	To compare farms. Certification or labelling schemes	France
Benchmarking system	Cattle, pigs	Sum of ranks obtained by a farm on a given population	Ranking, Benchmarking	Compare and rank farms on a given set of farms	UK

The five freedoms of the farm animal welfare council provide a basis for a multidimensional approach of welfare. Many measures with their own strengths and limitations are potentially available for a welfare assessment. Further on, one kind of decision that has to be taken in relation to welfare assessment concerns the quality of the evidence (Broom, 2010b). A challenging balance should be found between the feasibility of that measure (and its associated methodology) at field level and its validity and reliability when detecting a given welfare problem. The data obtained from that measure should be methodologically and statistically robust enough once applied at commercial level. Meanwhile another kind of decision has to be taken on the way in which different types of measures can be considered together in order to form an overall evaluation of welfare. The method of aggregation of measures will depend greatly of the application that will be given to this welfare assessment system.

The increased interest in developing a multidimensional and understandable welfare assessment system based mainly on animal-based indicators led the E.U. to fund a large-scale project, Welfare Quality® that aims to produce a European standard for welfare assessment (Blokhuis et al., 2003).

4. WELFARE QUALITY® ASSESSMENT PROTOCOLS

4.1. Overall aim of the Welfare Quality® project

The Welfare Quality® (WQ®) project started in 2004 being a partnership of 44 institutions from 13 European countries. One of the main proposals of this project was to develop an integrated standardised methodology for the on-farm assessment of welfare in cattle, pigs and poultry from farm to slaughter, based mostly on animal measures (Blokhuis et al., 2003). Later on, the project was extended to four Latino American countries: Mexico, Brazil, Uruguay and Chile.

A number of studies have shown that European citizens are becoming increasingly concerned about the way in which animals are treated in the production of animal food products (Bennett, 1996; Miele and Parisi, 2001). However, there is no common standard for assessing animal welfare and for providing consumers with the relevant information. An understandable European standard for animal welfare with appropriate labelling system will help provide consumers with information on the products they buy (Blokhuis et al., 2003). To achieve standardisation and minimise duplication of effort, information should be obtained in a harmonized and multidisciplinary way from different working groups. Furthermore, since welfare problems - as perceived by the animals - do not recognise national boundaries, this uniformity should be achieved at an international level (Blokhuis, 2008).

4.2. Building an understandable framework of aggregation

The WQ® assessment system may be used by certification bodies providing assurances to consumers, by agencies enforcing legislation, or by welfare researchers. For marketing purposes the framework of aggregation should allow an absolute score for any farm to be produced. It should be easy to understand and ease communication with stakeholders. Finally, the overall framework has to be universal and applicable whatever the species and the production system. A multidimensional and multidisciplinary approach of animal welfare was undertaken since the beginning of the project. The five freedoms of the Farm Animal Welfare Council (1992) were considered as a basis to develop such monitoring systems.

A first step was to define a clear set of criteria that followed a set of general requirements (Bouyssou, 1990; Botreau et al., 2007a):

- The set of criteria must be *exhaustive*. For example, containing every important aspects of animal welfare quoted in the literature
- It must be *minimal*. Redundancy between criteria should be avoided. For example, containing only necessary criterion
- Criteria must be *independent* of each other. The interpretation of one criterion should not depend on that from another
- The set of criteria should be *agreed* by all stakeholders
- It should be composed of a limited number of criteria in order to be *legible*

To be selected, a particular welfare criterion had to (Botreau et al., 2007a):

- Be applicable to any farm animal species
- Be mainly based on animal-based measures
- Limit compensation between criteria and may allow compensation within criteria
- Exclude hazards with a low probability (like fire or attacks by predators)

Following all the previous rules, a set of 12 independent but complementary welfare criteria was established. To ease communication with stakeholders, those criteria were grouped into 4 main principles of animal welfare: good feeding; good housing; good health; and appropriate behaviour

(**Table 2).** Those principles and criteria establish a common framework for the development of monitoring systems for the three species concerned and whatever the production system studied.

Table 2 Principles and criteria from the Welfare Quality® evaluation

Principles	Criteria
Good feeding	Absence of prolonged hunger
	Absence of prolonged thirst
Good housing	Comfort around resting
	Thermal comfort
	Ease of movement
Good health	Absence of injuries
	Absence of diseases
	Absence of pain induced by management procedures
Appropriate behaviour	Expression of social behaviours
	Expression of other behaviours
	Good human-animal relationship
	Positive emotional state

The aggregation relied on expert opinion on what counts for animal and what societies find acceptable or not. Experts were asked to give their opinion on situations or data sets and weightings were calculated to match their answer. Rules were based on expert opinion and were designed to limit compensation between criteria (Capdeville and Veissier, 2001; Botreau et al., 2007c). Then the aggregation method was based on a weighted sum of scores. This leads to a hierarchical structure where the information at each level remains available, particularly at the criteria level. A sequential process is followed to construct welfare criteria from measures taken on farms, then to form principles and finally to aggregate principles into an overall score of welfare assessment (Botreau et al., 2007b; **Figure 3**)

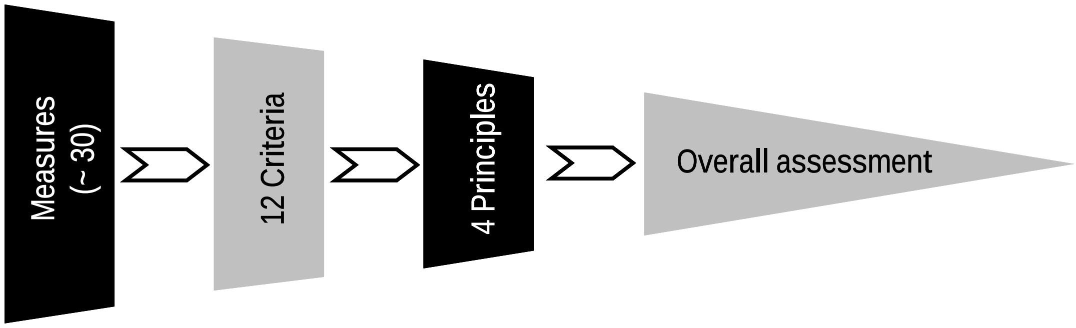


Figure 3 Hierarchical structure designed in Welfare Quality® to integrate several measures into an overall welfare assessment (*Adapted from Botreau et al., 2007b*)

4.3. Selection of measures in the Welfare Quality®

Each welfare criterion was assessed using one or several measures, mostly based on the direct observation of the animals. Selected animal-based measures should evaluate the performance of a husbandry system at a certain moment of the production cycle and in different housing and management conditions. Inclusion into the monitoring system was scientific based and was determined on the basis of validity, repeatability and feasibility (Winckler et al., 2003). Initially, a list of possible measures was identified from the existing literature (consensus validity) for each one of the species and type of production concerned (e.g. sows and piglets; growing pigs). Those measures had to cover the different dimensions of welfare and be practical under commercial conditions. Several animal-based measures of behaviour, physiology, injury, disease, mortality, etc. had been previously developed and validated for assessing welfare under experimental conditions (Archer, 1973; Rushen and de Passillé, 1992; Fraser, 1995; Von Borell et al., 2001). Some of these measures could not be applied under commercial conditions (low feasibility) and were consequently eliminated for further validation. For instance, most physiological measurements that are too invasive and costly or several behavioural observations which are too time consuming were not considered. An additional limitation was the difficulty of interpretation that can give some of these measures on farm (Barnett and Hemsworth, 1990). The measures that were potentially feasible on farm or slaughterhouses needed, then, simplification and an additional study on its validity and repeatability.

It was agreed that all parameters – except behavioural parameters – would be assessed using a three-point scale ranging from 0 to 2 (ordinal scale). A score of zero is indicative of acceptable welfare, a score of 1 indicates some compromise of welfare, and a score of two indicates unacceptable welfare. For feasibility or repeatability reasons two scores could be fused.

Most of the animal-based measures included in the protocol have been tested for their independent validity and repeatability in pilot studies. In the case of sows and growing pigs, those studies are regrouped within the Welfare Quality Reports Nº.10 edited by Forkman and Keeling (2009). The great majority of the measures were agreed for consensus validity and studied for their repeatability. Their feasibility on farm was discussed and evaluated on training sessions. **Table 3** summarizes the assessment of the validity, inter-observer repeatability and feasibility of the animal-based measures included into the WQ® protocol for growing pigs.

Many criteria could be defined by several animal-based measures, however some areas still lacked of any of reliable animal-based indicator. In those cases, a set of resource or management-based measures were included. For example, dehydration parameter which lack of feasible animal-based measure is defined by means of a description of water supply and its management. Other environmental-based measures that were considered easy to carry on were included into the monitoring system in order to be able to diagnose the possible causes of welfare deficiencies.

Table 3 Summary of the validity, inter-observer repeatability and feasibility on farm of the animal-based measures included into the Welfare Quality® protocol for growing pigs (studies from Forkman and Keeling, 2009)

Animal-based measures (*what it is supposed to measure*)	Validity	Inter-observer Repeatability	Feasibility on farm
Body condition (*poor feeding*)	Related to reproductive failure in sows Not studied in growing pigs Specificity: many diseases such as PMWS may also lead to very poor body conditions. Consensus validity	Assessed during training in growing pigs. Good.	Good
Bursitis (*comfort around resting*)	Related to floor quality and floor type. Moderate bursitis doesn't seem to be painful once appeared but may cause pain during development. Eroded bursitis can be painful. Possible confounding factors: age	Three point scale: rs = 0.5 – 0.6	Good
Manure on body (*heat stress*)	Related to thermal stress: Pigs at high ambient temperature were dirtier. Specificity: Animals soiled also showed a higher risk of urinary disorders, mastitis in sows. Possible confounding factors: type of floor; space allowance.	% pigs soiled: rs = 0.7 – 0.9	Good
Shivering (*cold stress*)	Recognised as physiological adaptation to a cold environment. However none of the treatments used in the pilot study induced shivering Specificity: Can also be related to diseases (fever); neurological disorders.	Not evaluated for low incidence	Good

Panting (*heat stress*)	Extreme panting is recognised as a physiological adaptation to a hot environment. However none of the treatments used in the pilot study induced shivering Respiration rate was higher in pigs housed in higher environment temperature	Respiration rate (breaths/min): rs = 0.9 Extreme panting: not evaluated for low incidence	Respiration rate has a low feasibility Panting should be easy to assess.
Huddling (*cold stress*)	Related to thermal stress: pigs maintained at low ambient temperature displayed a higher frequency of huddling. Possible confounding factors: type of floor; space allowance.	rs = 0.6 – 0.7	Good
Lameness (*abnormal gait due to painful lesions*)	Related to physical injury, infection or inherent leg weakness. The relationship between lameness scores and pain is not well known Consensus validity Possible confounding factors: abnormal gait due to a smooth floor	Growing pigs taken out individually to a passage way and scored on a four-point scale: not significant differences between observers Growing pigs assessed within the pen on a three point scale: not assessed	Impossible to take out pigs to a passage way. Lameness has to be assessed within the pen.
Wounds (*injuries of the skin due to physical or social environment*)	Validated by measures of aggression comparing different housing environments or feeding regimes (in short term housings). Validity unknown under long term housing. Possible confounding factors: age; gender; castration	Categories of Severity: rs = 0.2 – 0.7 Total Counts of lesions: r = 0.9	May be time consuming. Difficult to count lesions on dirty pigs
Tail biting lesions (*injuries due to tail biting*)	Severely wounded animals experience pain. Multifactorial problem and episodic. Difficult to predict. Confounding factors: docked/undocked tails; age; weather	Not available. Expected to be high.	Good.
Coughs and sneezes (*respiratory diseases*)	Prevalence of coughing is related to prevalence and severity of lung damage. The quantification of coughing and sneezing can estimate pneumonia and atrophic rhinitis indices respectively. Specificity: Environmental contaminants can produce signs of respiratory disorder Consensus validity	Not assessed	Number of coughs and sneezes per group of animals should be assessed easily.
Pumping (*respiratory diseases*)	Laboured respiration due to parasitic, bacterial, viral invasion of the lungs Specificity: may also be due to anaemia or a great increase in body temperature Consensus validity	Not evaluated for low incidence Expected to be high.	Good

Twisted snouts (*respiratory diseases*)	Severe nasal distortion: Indicator of atrophic rhinitis	Not evaluated for low incidence. Expected to be high when considering severe nasal distortion	The detection of severe nasal distortion should be feasible
Rectal prolapse (*enteric diseases*)	May be due to many factors that increase the intra-abdominal pressure (enteritis, diarrhea, cough, liquid feeding, etc.) Consensus validity	Not assessed Expected to be high when considering the presence of the absence of extrusion	Good
Scouring (*enteric diseases*)	Specificity: may be linked to gastrointestinal or systemic diseases as well as miscellaneous conditions Confounding factor: may vary to some extent according to the diet. Consensus validity	Not assessed.	May be difficult to assess on dirty floors or slatted floor
Skin condition (*diseases of the skin*)	The inflammation or discolouration of the skin can indicate the presence of diseases restricted to the skin or systemic diseases. The skin may be altered by many other factors: nutrition, sunburn, local inflammation due to the impact of the floor etc. Consensus validity	Not assessed	Good
Hernias	May be due to an infection or an abscess. Specificity: strong genetic component (heredity)	Not assessed. Expected to be high when considering the presence or the absence of umbilical or inguinal hernias	Good
Social behaviour (*aggressions; positive social behaviour*)	First level of response to an aversive or stressful environment. Agonistic behaviour can affect the health and the performance of the animals. The link between positive social behaviour and positive emotions is not straightforward.	Negative social behaviour: rs= 0.4 –0.6 Positive social behaviour: rs= 0.6 - 0.7 Total social behaviour: rs= 0.7 – 0.8	Methodology based on repeated scan sampling is feasible.
Exploratory behaviour	Lack of exploration can lead to redirected behaviours and health problems. Exploration is considered as a behavioural need in growing pigs	Investigation of the pen: rs= 0.8 – 0.9	Methodology based on scan sampling is feasible.

HAR (*obvious poor human-animal relationship*)	Predictability with handling treatments (minimum intervention of stockman; mild interaction with a soft stick; aversive interaction with an electric prod) Specificity: May be sensitive to the frequency of human contacts. The motivation of the pigs to explore the new environment (the observer) may confound the results. Confounding factors: pen shape and size may also influence the panic response of the animals.	HAR test: rs= 0.6 – 0.9	May be difficult to evaluate in overcrowded pens.
QBA (*emotional state*)	Coherence with quantitative behavioural and physiological measures at experimental level and using a Free-Choice-Profiling (FCP) methodology. The use of a fixed scale has to be validated in further studies Specificity: a qualitative approach may suffer anthropomorphism, cultural bias, contextual bias etc.	The fixed qualitative rating scale showed a high repeatability between four UK vet students watching video recording of the animals: Kendall's W= 0.6 – 0.7	Good

5. OVERALL EVALUATION OF THE MONITORING PROTOCOL

Aiming to overcome the limitations of the protocols already established, Welfare Quality® has proposed a multidimensional framework to evaluate welfare. A set of measures, mainly based on the direct observation of the animal, has been selected from the literature and each measure has been evaluated in pilot studies for its independent validity, repeatability and feasibility. Further on the final monitoring protocol has to be tested under commercial conditions for: 1) Its feasibility on a routine basis and through many different farms; 2) Its sensitivity to discriminate among farms; 3) Its validity at field level; 4) Its repeatability over time; 5) Its adequacy as an advisory tool.

In total, ten monitoring protocols have been proposed. Three of them concerned the porcine species: sows and piglets on farm, growing pigs on farm, and pigs at slaughter. This thesis will focused on the monitoring protocol of growing pig on-farm.

5.1. Feasibility and sensitivity

Welfare measures may vary in precision, relevance and relative contribution to an overall welfare assessment (Botreau et al., 2007c). The relevance and the sensitivity of the measures included into the protocol have to be assessed at field level. In addition their feasibility as part of a whole monitoring system must be checked in a wide range of commercial conditions. Then, a difficult balance will have to be found between the sensitivity of that measure and its feasibility. Observations in experimental conditions are usually not sufficient to obtain such information and it may be necessary to run observations on a large scale of farms from different rearing systems and local conditions. Information on the range of variation of each measure - such as mean percentage, minimum, maximum, and standard error - once applied at field level is needed. These observations are very demanding and are rarely carried out to identify the distribution of all measures (Botreau et al., 2007c). In order to show significant differences in contribution of various housing systems to sustainable development, it is necessary to quantify the within and between housing systems variation for each indicator. The sensitivity of an overall assessment system refers to its capacity to discriminate between farms (not all farms obtain the same results) as well as its capacity to discriminate between production systems. Variations in observations lead to variations in the assessment so that farms with welfare problems can be identified. They include the effects of variations in the way the farming system is managed as well as specific system-animal interactions. The monitoring systems developed must then enable differentiation between living conditions of animals and their sensitivity must correspond to actual variations between farms (Blokhuis, 2008). To make an evaluation at farm level, when welfare can be considered to have its effect on individuals it is proposed not to recommend the use of average values taken at farm level alone, but rather to try to describe also the variation within the farm (Botreau et al., 2007b). Intra-farm variability describes the extension of a welfare problem within a farm and allows to evaluate if an indicator is representative of the actual welfare state of the overall group of animals present on the farm. Goossens et al. (2008) studied the variability of several animal-based measures (aggression, approach-avoidance, tail and ear biting behaviours, rubbing, dirtiness, scratches, tail and ear lesions and coughing) within a population of 41 farrow-to-finish farms. Further on, a benchmarking system based on a sum of ranks was studied for the measures that presented sufficient variability. To our knowledge, further tests of the sensitivity of an overall monitoring system of animal welfare have not been carried out.

To sum up, information on the relevance and the sensitivity of a measure is needed to:

- Evaluate the incidence of each measure. If the incidence of a welfare problem is rare, information about the housing or the management may be of a more importance (Rousing et al., 2001). If a measures is not sensitive enough it may need refinement.
- Have more objective judgements on which levels of a given measure should be considered as normal or not in a given housing/rearing system. The description of the range of variation of behaviours within a housing system or a between housing system should reveal the pattern of that measure. Further on a benchmarking system may be established.

- Adjust the aggregation model to the characteristics of the measures so that the overall assessment remains sensitive (e.g. not all units obtain the same assessment score) (Botreau et al., 2007b).

The final monitoring systems will be based on the most practicable prototype monitoring systems that still conserve most of the variation between farms and between production systems. The relevance and the sensitivity of each one of the measure included into the protocol will be assessed. The sensitivity will be studied as a function of level of detail and in a context of comprehensiveness. The first version of the complete prototype monitoring system will be run on a sample of 30 farms from a same production system. A second step will be to quantify the variation of each measure as well as their feasibility in different production systems.

5.2. Interpretation of the outcomes: selectivity and contextual validity

The validity of a measure across systems is a key requirement when developing widely applicable welfare assessment systems. A first problem concerns the selectivity of a measure within a given environment – the measure should describe what it is supposed to describe and nothing else (Martin and Bateson, 2007). Measures should then be comparable in different environments and emphasis should be kept on only measures valid across systems. Concerns about contextual validity arise when differences in context potentially threaten the validity of a measure. The same outcome of an indicator may have different meaning in different contexts. The validation (including consensus validity) of the measures included into the WQ® protocol has been mainly estimated for animals kept under intensive conditions. Can this validation be extrapolated to extensive rearing systems? Does the same measure mean the same in different rearing systems? These two questions should be considered when evaluating the validity of the overall protocol.

In fact, to be aware of the limitations in the interpretation of the welfare outcomes, all the parameters included into the welfare assessment system have to be quantified, compared and discussed in a large variety of housing and management conditions.

5.3. Benchmarking and setting thresholds

Some scientists actually argue that instead of attempting to 'measure' animal welfare, the role of science should be seen as identifying, rectifying and preventing welfare problems (Fraser, 1995). For that reason, much attention has been given recently to the study of risk factors in animal welfare (Smulders, 2009). The WQ® protocol has not been thought as a risk assessment tool.

Instead it aims to detect farms or housing and management systems with particular welfare problems leading towards a certification/labelling of those farms. Nevertheless, to be successful, a welfare assessment tool should reveal the causes of impaired welfare and thus potential improvements of the husbandry and management system (Waiblinger et al., 2001). This requirement is fundamental to drive improvement in animal welfare (Whay et al., 2007). The farmers should know how they can prevent a given welfare problem and how much their efforts actually improve the welfare of their animals (intervention programme).

First of all, the assessment should allow the identification of farms with current welfare deficiencies in relation to a given population of farms. As mentioned previously, a benchmarking system (welfare comparisons among group of farms) based on sum of ranks was proposed by Whay et al., 2003b. For each category of welfare measure thresholds or norms are established. Once the limits are placed on particular welfare concerns, the targets farms can be identified and prioritised for further evaluations and intervention programmes. Information on a farm's performance relative to its peers ('benchmarking') is an important encouragement effect as many farmers are motivated to compete with their colleagues (Main et al., 2003a). Benchmarking can be useful to raise awareness and can help to set thresholds (Keeling, 2011).

Secondly, the welfare assessment protocol should be efficient to assess and compare the relative strengths and weaknesses of each production system. Welfare concerns may have more severe consequences in one system that in another. It is essential to identify which factors are the main sources of variation within each system. The detection of causal factors will help to predict which production systems or farms are more likely to show a given welfare deficiency. This prediction will provide a basis to rectify and prevent welfare problems within an existing system. As mentioned before, animal-based observations give the most direct insight into how animals are coping within their own environment. However, as part of the process of stimulating behaviour changes, it is essential that animal owners and carers also understand the significance of these welfare measures (Whay et al., 2007). Besides been fundamental in an intervention programme, a risk assessment may reveal the existence of possible confounding factors. Those factors should be taken into consideration to prevent misinterpretations and guide the aggregation method (Bouyssou, 1990). The epidemiological approach is a well-represented tradition in animal health research; however, it remains underutilised in animal welfare science (Rushen, 2003). To be able to assess risk quantitatively, information on the links between possible causal factors and their consequences on welfare outcomes is needed.

5.4. Test-retest repeatability

Measures taken on animals tend to be subject to variations. The observer can be a source of variation. Inter-observer reliability was evaluated for the great majority of measures included into the WQ® protocol (Forkman and Keeling, 2009). Inter-observer repeatability can be improved by an adequate training and by merging several detailed scores into fewer categories. It should be considered that this last option may lower the sensitivity of the measure. In addition to an observer effect, the variation of a measure may be owing to fluctuations in the environment or in the animals themselves. Test-retest repeatability measures the agreement between observations made on the same subjects on at least two different occasions (Scott et al., 2001). The repeatability of assessments over time is especially important and constitutes a special case where welfare assessment protocols are going to be used for certification purposes with infrequent assessments. In order to be cost effective, such assessments will take place in longer intervals of supposedly more than six months (Knierem and Winckler, 2009). The consistency of an assessment system over time is studied at farm level instead of individual level. Thereafter, changes over a long period of time may occur due to seasonal effects (e.g. climate), to changes in the management or resources of the farm, or be linked to the batch (animals) itself. The results of an assessment tool need to be representative of the long-term situation of a farm instead of being sensitive to changes that are insignificant for the welfare state of the animals (Knierem and Winckler, 2009). Thus, the consistency of a welfare assessment should be studied at different moments of the production cycle when no major changes on the farms occurred. Test-retest repeatability of the approach test in cattle (Lensink et al., 2003; Rousing and Waiblinger, 2004) with a number of days between has been assessed on farm. Winckler et al. (2007) assessed the consistency of several animal-based parameters for dairy cattle during five visits at two-month intervals on eight farms. The results of the previous studies showed a limited consistency of the majority of the measures. Results of test-retest repeatability over a long period of time have not been undergone in growing pigs.

The assessment of welfare should be quite separate from any ethical judgement but once an assessment is completed, it should provide information that can be used to take decisions about the ethics of a situation (Broom, 2010b). This thesis aims to evaluate the robustness of an overall welfare assessment system discussing the protocol from a descriptive and prescriptive point of view. It does not pretend to decide what should be considered as acceptable or not for the welfare of the animals. Either it seeks to provide an aggregation or scoring scale for reflecting the welfare of a farm. Besides depending on the robustness of the overall protocol applied at commercial level, this final step relies on the final application of the protocol and may be, at the end, highly dependent of ethical considerations.

OBJECTIVES

The general objective of this thesis is to evaluate and discuss the feasibility, validity and repeatability of the Welfare Quality® assessment protocol for growing pigs on commercial farms.

The specific objectives are as follows:

1. To evaluate the feasibility of the welfare assessment system in intensive and extensive conditions.

2. To evaluate the sensitivity of the Welfare Quality® assessment protocol to detect welfare variations between farms and between production systems.

3. To assess the selectivity and contextual validity of the results obtained from the application of the welfare assessment system in different production systems.

4. To evaluate the repeatability over a long period of time of the animal-based measures included in the Welfare Quality® protocol.

5. To test the Welfare Quality® protocol as a tool to identify and prevent welfare problems at farm level.

CHAPTER 1

Welfare Quality
Science and society improving animal welfare

APPLICATION OF THE WELFARE QUALITY® PROTOCOL TO ASSESS GROWING PIGS KEPT UNDER INTENSIVE CONDITIONS IN SPAIN

Based on Temple D, Dalmau A, Ruiz de la Torre JL, Manteca X, Velarde A 2011. Application of the Welfare Quality® protocol to assess growing pigs kept under intensive conditions in Spain. *Journal of Veterinary Behavior 6:* 138-149

ABSTRACT

This study is a first step in the validation of the Welfare Quality® protocol for growing pigs on farm. The feasibility and sensitivity of the overall protocol was evaluated on 30 Spanish conventional farms and a methodology to identify farms with particular welfare problems was proposed. A total of 64,496 pigs was assessed based on 12 criteria related to the four welfare principles of good feeding, housing, health and appropriate behaviour. Good feeding, housing and health measures were scored at pen or individual level according to a three-point scale ranging from 0 (good welfare) to 2 (poor or unacceptable). Appropriate behaviour was assessed by means of scan sampling of social and exploratory behaviour, qualitative behaviour assessment and human-animal relationship test. The mean time for carrying out the full protocol was 6h20min ± SD = 51min per visit. The between-farm variability of moderate and severe bursitis, manure on less than 50% of the body (moderately soiled body), expression of social positive and negative behaviours and exploration allowed discrimination among farms. However, other animal-based measures, especially the ones related to the good health principle, presented very little variation that could be used to differentiate among farms. For each animal-based measure, confidence limits were estimated to make statistical inferences at population level. This methodology could be useful to identify farms with poor welfare conditions.

Key words: animal-based measures, animal welfare, farm, growing pig, welfare assessment system, Welfare Quality®

Running title: Welfare assessment on growing pig farms

INTRODUCTION

The Welfare Quality® project aims to provide reliable and standardized information to consumers and European bodies regarding the welfare conditions within which animals are grown. It was suggested that a protocol be created to assess animal welfare by integrating different measurements in an objective and scientific way. The protocols consider three species: cattle, poultry and pigs, although in the present paper, only pigs are considered. It is generally accepted that welfare comprises physical and mental health (Webster et al., 2004; Dawkins, 2004) and that it includes several aspects such as absence of thirst, hunger, discomfort, disease, pain and injuries, stress and the expression of normal behaviour (Farm Animal Welfare Council, 1992). On the basis of this general concept of animal welfare and after an agreement was established among consumers, stakeholders, policy makers and scientists, Welfare Quality® has been defined using four main principles of animal welfare: i) Good feeding, ii) Good housing, iii) Good health and iv) Appropriate behaviour. Each one of these principles is defined by a set of independent but complementary criteria (Botreau et al., 2007) in turn characterized by one or various measures. Since the animal's welfare state depends on how well that animal is able to survive and to remain fit within the particular constraints of the husbandry system in which it lives (Whay et al., 2003), preference is given to animal-based measures in contrast to resource and management based measures. The decision on whether or not to select relevant measures is scientifically based and this choice is determined on the basis of validity, feasibility, and repeatability (Winckler et al., 2003). For that reason, each measure was studied for its independent validity to pig welfare and reliability between observers. Simple procedures to assess behaviour and to score lesions were defined and validated by various research groups through Europe (Courboulay and Foubert, 2007; Meunier-Salaün et al., 2007; Dalmau et al., 2009). Later, several selected measures were proposed and integrated into an overall assessment system for sows and piglets on farm, growing pigs on farm and finishing pigs at the slaughterhouse. In the present article, only those for growing pigs on farm are assessed.

For the overall protocol to be feasible, it should require little input from the farmers, should be easy to perform in practical conditions and should require little time. The assessment systems used by the certification schemes already in place in Europe are mostly based on the checking of resources and management whereas the Welfare Quality® assessment emphases animal-based measures. Therefore, a first step of this study was to estimate the time required to perform the overall assessment. It also presents a descriptive analysis of several animal-based measures of the welfare assessment of intensively kept pigs on 30 Spanish farms as well as an evaluation of the sensitivity of the protocol, that is, its capacity to discriminate among intensive conventional farms.

Finally, it aims to propose a practical methodology to identify farms that are more likely to present poor welfare for a given animal-based measure.

MATERIAL AND METHODS

Farm sample

The welfare assessment system was tested on growing pigs kept under intensive conventional conditions on 30 farms in Spain. Selection of farms was based on management practices, farm size and veterinary records. The selected farm sample was considered representative of the conditions of pigs kept under intensive conditions in Spain. All farms were conventional with an indoor production system on concrete floor, including six farms with fully slatted floor; pigs entered the farms with an average weight of 19 ± SD =1.8 kg and left to be slaughtered at an average weight of 105 ± 2.9 kg. Two farms included a post-weaning period, where piglets entered the farms weighing 5 kg. Age of the growing pigs during the survey ranged from 8 to 24 weeks old (mean ± SD = 16 ± 3.9; median= 15). The size of the farms ranged from 300 pigs to 9 400 pigs (**Table 1**). Pigs were kept in groups of an average of 14 animals (14 ± 4.5; 14), but on one farm some pens contained up to 42 animals. During the days of assessment, environment temperatures ranged from 16ºC to 28ºC. The Welfare Quality® protocol was assessed on nine farms by one observer and on 21 by another. To minimize differences between observers and to standardize the scores from the visits, observers were given identical training before the assessment.

Table 1 Farm size relative to number of intensive pigs.

Average number of pigs	Number of farms
300 - 999	7
1 000 - 1 999	10
2 000 - 4 999	12
> 5 000	1

Protocol for growing pigs

The protocol consisted of four main principles of animal welfare (Good feeding, good housing, good health and appropriate behaviour) that were subdivided into 12 independent criteria (Keeling and Veissier, 2005). To assess each one of these criteria, a set of measures were used, selected according to their validity, feasibility and repeatability in previous studies (**Table 2)** (Velarde et al., 2007).

At the beginning of the visit, general information related to the farm was recorded by means of a questionnaire that was answered by the farmer as well as by visual inspection. The questionnaire recorded information on management, prevention of diseases, feeding, hygiene management, temperature regulation, castration routine, euthanasia criteria, and production and mortality records. After the general information had been collected, the visit consisted of behavioural observations by means of a qualitative assessment and scan samplings of social and exploratory behaviour. Finally, the visit ended with the evaluation of animal-based measures related to the good feeding, housing and health principles. The time needed to perform the different parts of the visit was recorded.

Table 2 Recorded measures for the welfare assessment of growing pigs on farm.

	Welfare Criteria	Measures
Good feeding	1 Absence of prolonged hunger	Body condition score
	2 Absence of prolonged thirst	*Water supply*[a]
Good housing	3 Comfort around resting	Bursitis, absence of manure on the body
	4 Thermal comfort	Shivering, panting, huddling
	5 Ease of movement	*Space allowance*[a]
Good health	6 Absence of injuries	Lameness, wounds on body, tail biting
	7 Absence of disease	Coughing, sneezing, pumping, twisted snouts, rectal prolapse, scouring, skin condition, ruptures and hernias
	8 Absence of pain induced by management procedures	*Castration, tail docking*[a]
Appropriate behaviour	9 Expression of social behaviours	Social behaviour
	10 Expression of other behaviours	Exploratory behaviour
	11 Good human-animal relationship	Fear of humans
	12 Positive emotional state	Qualitative Behaviour Assessment

[a]*Resource-based measures.*

Animal-based measures of feeding, housing and health

Feeding, housing and health measures were scored at pen or individual level according to a three-point scale ranging from 0 to 2. The assessment scales were selected so that a score of 0 was awarded where welfare was good; a score of 1, as when applicable and feasible, where there was some compromise on welfare; and a score of 2 where welfare was poor and unacceptable. For each measure, the number of animals or pens scored as either 1 or 2 was noted down. In some cases, where a condition was either present or absent, a binary (0: absent / 2: present) scale was used (**Table 3).** The pigs of ten selected pens were assessed throughout the farm. Pen locations were spread fairly within the farm so as to have the largest view of it. As far as possible, all rooms were assessed and the number of animals sampled in a room was in proportion of the total number of pigs per room. Hospital pens were not considered in this sampling. Pigs were individually scored for shivering, panting, huddling, body condition, bursitis, pig dirtiness (or manure on the body), wounds on the body, tail biting, lameness, pumping (heavy and laboured breathing), twisted snouts, rectal prolapse, skin condition, and hernias. Shivering, panting and huddling measures were assessed before the observer entered the pen and by observing the whole group. Huddling was assessed only in resting animals, being assessed first, just before the animals began to stand up as a response to the observer presence. Coughs and sneezes were evaluated in 6 different points of the farm considering all the animals that could be seen and controlled from the corridor. To assess them, pigs were forced to stand up and subsequently observed for 5 minutes (Geers et al., 1989). The number of coughs and sneezes and the number of animal coughing and sneezing were counted; coughs and sneezes were considered separately.

Then, all the other measures were scored from inside the pen because observations taken from the corridor resulted in under-evaluation of the number of lesions (Coubourlay and Foubert, 2007). All these measures, except scouring, were assessed individually. Scouring was assessed directly at pen level, as the observer looked for areas with diarrhea in zones with fresh faeces inside the pen (**Table 3**)**.** Wounds on the body, skin condition, pig dirtiness (moderately and severely soiled body) and bursitis were evaluated from only one side of each pig as there was no statistical difference in the amount of lesions and dirtiness scored on the right and left side of the animal (Coubourlay and Foubert, 2007)**.**

Table 3 Scoring scale for animal based-measures of feeding, housing and health.

Animal-based measures	Score	Description
Body Condition	0	Animal with a good body condition
	2	Poor body condition: Animal with visible spine, hip and pin bones
Bursitis	0	No evidence of bursae
	1	Moderate bursitis: one or several small bursae on the same leg or one large bursa
	2	Severe bursitis: several large bursae on the same leg or one extremely large bursa or any eroded bursae
Manure on the body	0	Less than 20% of the body surface is soiled
	1	Moderately soiled body: more than 20% but less than 50% of the body surface is soiled with faeces
	2	Severely soiled body: over 50% of the body surface is soiled with faeces.
Huddling	0	Pig lying with less than half of its body lying on top of another pig
	2	Pig lying with more than half of its body lying on top of another pig
Panting	0	Normal breathing.
	2	Rapid breath in short gasps.
Shivering	0	No vibration of any body part
	2	Slow and irregular vibration of any body part, or the body as a whole
Wounds on the body	0	If all regions of its body have up to 4 lesions
	2	Severely wounded: when more than 10 lesions are observed on at least two body zones or if any zone has more than 15 lesions
Tail biting	0	No evidence of tail biting; superficial biting but no evidence of fresh blood or of any swelling
	2	Bleeding tail and/or swollen infected tail lesion, and/or part of tail tissue missing and presence of crust
Lameness	0	Normal gait or difficulty in walking, but still using all legs; swagger of caudal body while walking; shortened stride
	1	Severely lame, minimum weight-bearing on the affected limb
	2	No weight-bearing on the affected limb, or not able to walk
Pumping	0	No evidence of laboured breathing
	2	Evidence of laboured breathing
Scouring	0	No liquid manure visible in the pen
	1	Areas in the pen with some liquid manure visible
	2	All faeces visible inside the pen is liquid manure
Skin Condition	0	No evidence of skin inflammation or discoloration
	1	Localized skin condition: more than zero, but less than 10% of the skin is inflamed, discoloured or spotted
	2	Widespread skin condition: more than 10% of the skin has an abnormal colour or texture.
Hernias	0	No hernia/rupture
	2	Bleeding lesions, hernias/ruptures and/or hernias/ruptures touching the floor
Twisted snouts	0	No evidence of twisted snouts
	2	Evidence of twisted snouts
Rectal prolapse	0	No evidence of rectal prolapse
	2	Evidence of rectal prolapse

Behavioural measures

Appropriate behaviour was assessed by means of social and exploratory behaviour, qualitative behaviour assessment (QBA) and human-animal relationship (HAR) test.

Social and exploratory behaviour was assessed by means of scan samplings. Pigs were scored as either active or inactive. The behaviours recorded from active pigs were as follows: positive social behaviour, negative social behaviour, exploratory behaviour and others (eating, drinking, etc.) **(Figure 1).** Exploratory behaviour was divided into investigation of the pen and investigation of enrichment material **(Figure 1).** Before beginning the scan, the assessor entered the room to ensure that all animals were standing up. The observation of all the animals was recorded five minutes later from outside the pen. Each pen was observed five consecutive times with a two and a half minutes interval between two scans. This method, validated by Courboulay and Foubert (2007), was performed at three different observation points of the farm so as to give a good overall representation of the farm. Approximately, 40 - 80 animals were observed from each point of observation (120-180 animals in total) and the scans were started on an average at 12:00 noon.

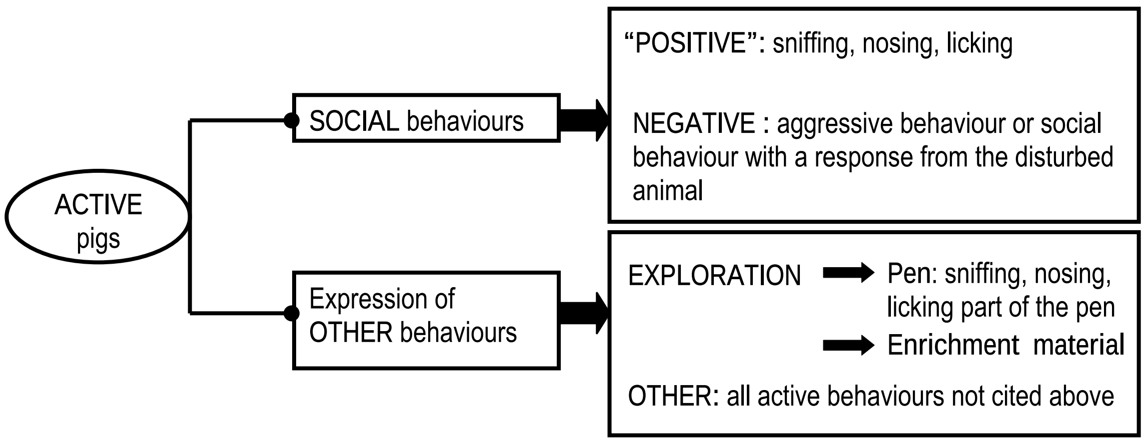


Figure 1 Recorded behaviours during the scan observations of pigs.

The qualitative behaviour assessment (QBA) was carried out at six observation points per farm during a total of 15 minutes. The QBA uses descriptive terms with an expressive connotation to reflect animals' experiences of a situation (Wemelsfelder, 2007). A rating scale was used to score pigs at group level based on 20 different terms: 1.Active, 2.Relaxed, 3.Fearful, 4.Agitated, 5.Calm, 6.Content, 7.Tense, 8.Enjoying, 9.Frustrated, 10.Sociable, 11.Bored, 12.Playful, 13.Positively occupied, 14.Listless, 15.Lively, 16.Indifferent, 17.Irritable, 18.Aimless, 19.Happy and 20.Distressed. The scale was of 125 millimetre long whose left side (or minimum) meant that the expressive quality indicated by the term was entirely absent in any of the pigs observed, and right side (maximum) indicated that the descriptors was dominant across all pigs. Scoring was conducted using this scale on the basis of the number of animals showing each one of the terms used and the intensity of the behaviour registered in those animals.

Human Animal Relationship (HAR) was evaluated according to the fear of human test validated by Courboulay and Foubert (2007). For this test, ten selected pens were tested throughout the farm. Each pen was assessed as a whole. Only two possibilities were considered for this test: **0** – No panic response to human presence; **2** – More than 60% of the animals in the pen showing panic to the human presence. Panic was defined as animals faced away from the observer or huddled in the corner of the pen.

More detailed information about the overall methodology of assessment can be found in the Welfare Quality® Assessment protocol for pigs (2009).

Statistical analysis

Statistical analyses were performed by means of the Statistical Analysis System (SAS® 9.1; software SAS Institute Inc., Cary, NC; 1991-2001). Between-farm variability for each animal-based measure was evaluated by means of non-parametric models (PROC GENMOD) considering the variable farm as a fixed effect. A binomial distribution was used for body condition, housing and health animal-based measures. For the count data, such as behaviours recorded in the social scan, and cough and sneezes, a poisson or negative binomial distribution was used on the basis of the value of the deviance (Cameron and Trivedi, 1998). Since the social scan included five consecutive scans for each observation point, the variability between the three observation points assessed (within–farm variability) was evaluated for each behaviour recorded. The residual maximum likelihood was used as a method of estimation and the least square means of fixed effects (LSMEANS) were used when analysis of variance indicated differences ($P < 0.05$).

Considering that the farm sample was representative of the Spanish farm providing intensive conditions, confidence limits of the population were estimated for each measure by applying pseudo-liquelihood techniques similar to that introduced by Breslow and Clayton (1993). A logistic regression was used in a Generalized Linear Mixed Model (PROC GLIMMIX), where farm was considered as a random classification variable. Thresholds were defined for each measure establishing a confidence level of 0.80. Farms that presented prevalences higher than the upper confidence limit were detected.

RESULTS

Duration of the protocol

The mean time required to apply the Welfare Quality® assessment was 381 ± 51 min (6 h 21 min ± 51 min) ranging from 315 min to 580 min. **Table 4** summarizes the average time to accomplish the principal parts of the assessment.

Table 4 Average time in minutes to record the different parts of the Welfare Quality® protocol.

	Mean	SD	Median	Min	Max
General information recording	72	33.0	60	30	200
Questionnaire alone	37	10.3	35	25	60
Qualitative Behaviour Assessment	30	5.5	30	20	45
Scan sampling of behaviours	86	14.7	85	70	120
Feeding, Housing, Health measures	130	22.9	125	100	170

Animal-based measures of feeding, housing and health

The average prevalence of the feeding, housing and health measures with the minimum and maximum value are shown in **Table 5.** The highest prevalence was found for moderate bursitis. Panting, shivering, huddling, wounds on the body, tail biting, lameness, pumping, hernia and scouring scored as 2 presented prevalences below 2%. Moreover, none of the pigs assessed on the 30 intensive conventional farms from this study presented twisted snout or rectal prolapse. **Table 5** also illustrates within-farm variability showing the mean number of pens affected across farms. Some measures such as bursitis, cough and sneezes, skin condition and moderately soiled body could affect almost 100% of the pens assessed when present on a farm, resulting in low within-farm variability. Conversely, other measures such as tail biting or wounds on the body affected a lower number of pens (a maximum of 40% of the pens assessed) when present on a farm, thereby resulting in higher within-farm variability

Moderate bursitis (Chi-square = 727.6; df = 29; P <0.0001), severe bursitis (218.7; 29; P < 0.0001), moderately soiled body (537.4; 29; P < 0.0001), coughs (50; 29; P = 0.0094) and sneezes

(106.8; 29; $P < 0.0001$) demonstrated significant variability to allow for discrimination between farms using the logistic model. Comparison between farms helped identify different prevalences for each of these measures. An illustration is given in **Figure 2** for moderately soiled body. Conversely, body condition, severely soiled body, huddling, panting, shivering, wounds on the body, tail biting, moderate and severe lameness, pumping, localized and widespread skin condition, both scouring scores and hernia presented a prevalence and variability too low to allow differentiation among farms by means of the logistic model.

Confidence limits were calculated for each measure to allow a practical detection of farms that are more likely to present a welfare deficiency for a particular measure, and to extrapolate at population level (**Table 6).** Pseudo-liquelihood estimates perform poorly in cases with binary repeated observations, with a relatively small number of observations available (Molenberghs and Verbeke, 2005). Estimations were seriously biased when applied to the binary measure scouring scored as 2. Therefore, confidence limits for this measure were not estimated. Given that an average of 130 animals were assessed by farm, one case of poor body condition, panting, lameness, pumping, or hernia observed within a farm would exceed the upper limit of normality established in this study. Using wounds on the body as an example and inferring at population level, four conventional farms from the survey would be detected as farms that were more likely to present poor welfare with regard to this measure, with a confidence level of 0.80 **(Figure 3)**.

Table 5 Descriptive statistics obtained from the farm sample, of each animal-based measure from the principles of good feeding, housing and health. Results are expressed in percentage of animals affected in relation to the number of pigs assessed on each farm.

	Mean	SD	Median	Min	Max	Proportion of pens affected across farms
Poor Body Condition	0.1	0.37	0.0	0.0	1.5	0-20
Moderate bursitis	45.6	21.04	51.2	7.2	83.1	30-100
Severe bursitis	4.4	5.66	2.0	0.0	22.9	0-90
Moderately soiled body	16.6	13.43	17.2	0.0	43.4	0-100
Severely soiled body	3.7	6.31	2.0	0.0	32.1	0-50
Panting	0.1	0.22	0.0	0.0	1.0	0-10
Shivering	0.0	0.00	0.0	0.0	0.0	0
Huddling	1.3	1.78	0.4	0.0	7.3	0-40
Severely wounded	0.9	1.38	0.3	0.0	5.8	0-40
Tail biting	0.9	2.02	0.0	0.0	8.1	0-30
Lameness 1 (severe)	0.2	0.43	0.0	0.0	1.6	0-20
Lameness 2 (no weight bearing)	0.2	0.45	0.0	0.0	1.8	0-20
Pumping	0.1	0.25	0.0	0.0	0.8	0-10
Twisted snout	0.0	0.00	0.0	0.0	0.0	0
Coughs[a]	15.8	0.25	7.8	0	106.2	0-83
Sneezes[a]	19.7	0.19	12.8	2.2	84.3	17-100
Prolapse	0.0	0.00	0.0	0.0	0.0	0
Pens with Scouring 1[b]	12.3	18.70	5.0	0.0	70.0	0-70
Pens with Scouring 2[b]	0.3	1.83	0.0	0.0	10.0	0-10
Localized skin condition	2.8	4.64	0.9	0.0	17.3	0-90
Widespread skin condition	3.6	5.97	0.0	0.0	20.1	0-90
Hernia	0.1	0.27	0.0	0.0	1.1	0-10

[a]Expressed in percentage of coughs or sneezes emitted in relation to the number of pigs assessed.

[b]Expressed in percentage of pens affected in relation to the number of pens assessed

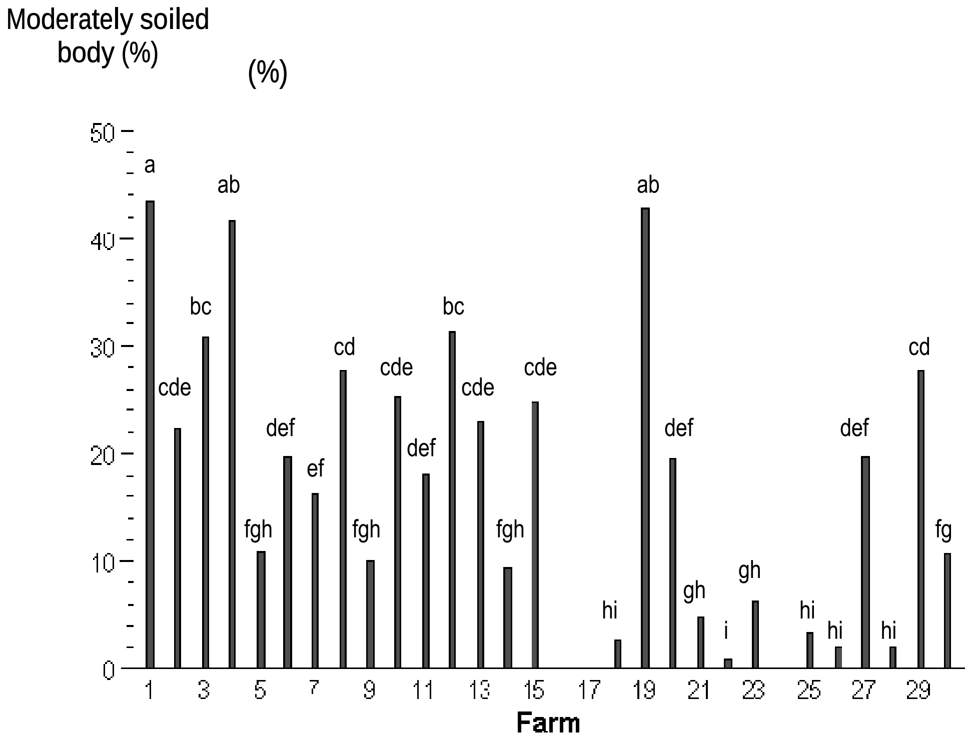


Figure 2 Distribution of the mean prevalence of moderately soiled body in percentage of pigs affected per farm. Different letters indicate significant difference between farms (P < 0.05).

Table 6 Lower and upper Confidence Limits of the percentage of animals affected for each animal-based measure assessed within the four welfare principles. A confidence level of 0.80 was assumed.

Animal-based measure	lower CL	upper CL
Good feeding		
Poor Body Condition	0.02	1.07
Good housing		
Moderate bursitis	25.03	64.52
Severe bursitis	0.97	7.64
Moderately soiled body	3.72	29.32
Severely soiled body	0.60	6.61
Panting	0.01	1.02
Huddling	0.45	4.25
Good Health		
Severely wounded animal	0.22	2.42
Tail biting	0.05	2.35
Lameness scored as 1	0.02	1.17
Lameness scored as 2	0.02	1.15
Pumping	0.02	1.00
Coughs[a]	4.25	29.76
Sneezes[a]	8.15	30.29
Pens with Scouring 1[b]	2.49	37.39
Localized skin condition	0.32	5.12
Widespread skin condition	0.14	6.28
Hernia	0.01	1.09
Appropriate Behaviour		
Active	56.56	78.98
Positive social	8.01	15.88
Negative social	3.55	7.14
Exploration	24.80	39.01
Other	41.25	59.89
Pens with panic	2.34	32.89

[a]Expressed in percentage of coughs or sneezes emitted in relation to the number of pigs assessed.

[b]Expressed in percentage of pens affected in relation to the number of pens assessed.

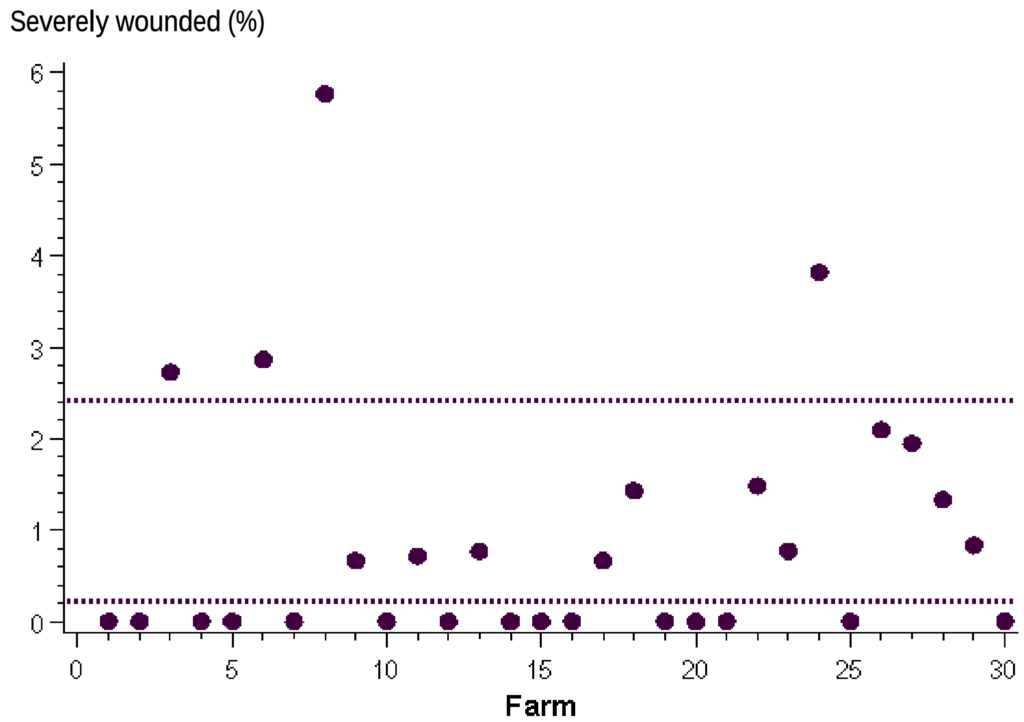


Figure 3 Lower and upper limit of the mean percentage of severely wounded animals per farm. A confidence level of 0.80 was used to allow inference at population level.

Behavioural measures

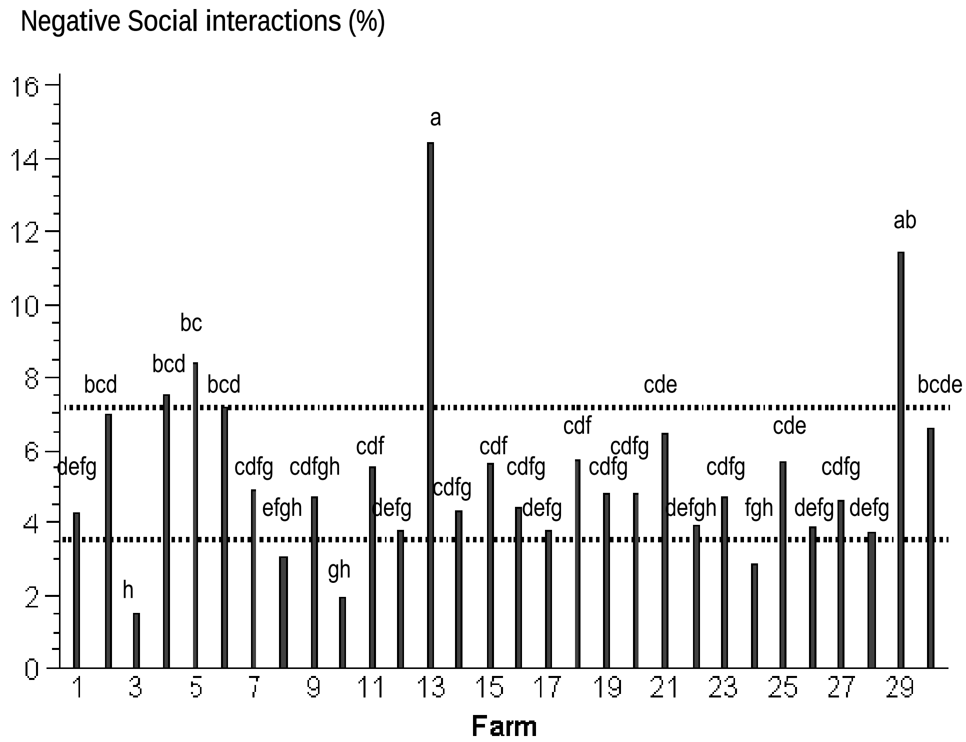
Descriptive statistics of active, social and exploratory behaviours recorded are presented in **Table 7**. The average percentage of active pigs was 67.7%. Approximately 18% of active intensive pigs performed social behaviour and negative social interactions ranged from 1.5% to 14.6% of all active behaviours.

Table 7 Percentage of animals performing the behaviours recorded during the scan sampling in relation to the number of active pigs observed in each farm, and percentage of pens with more than 60% of the pigs showing a panic response, over the total of pens assessed.

	Mean	SD	Median	Min	Max
Active	67.7	12.51	68.9	47.3	89.1
Positive social	12.2	5.12	10.8	3.2	26.9
Negative social	5.4	2.66	4.8	1.5	14.6
Exploration	31.3	8.00	31.3	14.9	53.4
Other	50.7	10.80	52.5	28.8	66.3
Pens with panic	13.7	19.91	5.0	0.0	80.0

Active behaviour (Chi-square = 200.81; = 29; $P < 0.0001$), social positive (172.0; 29; $P < 0.0001$) and negative (97.0; 29; $P < 0.0001$) interactions as well as exploratory (190; 29; $P < 0.0001$) and other behaviours (212; 29; $P < 0.0001$) demonstrated significant variability to allow for discrimination between conventional farms applying the logistic model. Moreover, there was a significant difference ($P < 0.05$) between the three observation points of the scan sampling on 20% of the farms assessed with respect to positive behaviour; on 30% of the farms with respect to negative behaviour; on 33% of the farms with respect to exploratory behaviour; on 37% of the farms with respect to "other" behaviours and on 70% of the farms with respect to the total of active behaviours. Confidence limits were estimated for the measures recorded on the social scan as well as for the human animal relationship test (**Table 6**). A representation of the sensitivity of a measure combined with the lower and upper inferred limit established by the threshold methodology is exposed for social negative behaviour in **Figure 4.** Farms that presented significantly higher occurrences of social negative behaviour applying the logistic model were also identified as farms with a higher probability of presenting poor welfare with regard to this measure, using the threshold methodology based on confidence limits.

The mean length of the scale scored by QBA is shown in **Table 8**. Standard deviations of descriptors indicated certain homogeneity between farms using the same production system. However, ranges of descriptors – except for "distressed" - were wide; indicating that some farms differed from the others on the basis of the qualitative assessment. The evaluation of the variability between farms for the QBA could not be verified because of the complexity of the statistical model.



Different letters indicate significant difference between farms (P < 0.05).

Figure 4 Distribution of occurrence of social negative interactions among farms. Broken lines illustrate the lower and upper inferred limit for the occurrence of negative social behaviour, with a confidence level of 0.80.

Table 8 Descriptive statistics of the qualitative behaviour assessment expressed in millimetres of the qualitative assessment scale established for each descriptor across farms.

Descriptors	Mean	SD	Median	Min	Max
Active	74.4	8.53	75.1	54.7	90.7
Relaxed	56.8	12.67	60.7	25.5	73.3
Fearful	17.9	13.00	19.4	0.0	69.4
Agitated	66.4	13.58	65.3	34.0	86.8
Calm	58.8	14.11	58.3	25.9	83.3
Content	43.1	11.57	45.4	24.3	63.9
Tense	38.9	25.92	26.0	6.9	93.5
Enjoying	36.2	9.43	34.7	20.8	62.5
Frustrated	29.4	12.29	26.6	11.1	63.9
Sociable	66.4	19.04	63.9	29.4	95.8
Bored	53.6	12.40	55.6	27.1	75.0
Playful	35.6	13.31	37.5	15.3	66.0
Positively Occupied	41.7	11.42	42.8	19.4	63.9
Listless	17.7	8.93	16.6	1.4	38.9
Lively	55.5	20.16	50.7	24.3	90.3
Indifferent	68.4	19.76	69.4	22.9	97.2
Irritable	43.2	20.85	37.7	9.7	94.9
Aimless	50.1	15.81	49.5	24.5	98.8
Happy	42.3	14.01	43.1	19.7	65.3
Distressed	2.9	5.83	1.4	0.0	31.7

DISCUSSION

The main objective of the present study was to evaluate the feasibility of the assessment protocol for assessing the welfare of growing pigs as well as its capacity to discriminate among conventional farms. Also, it aimed to propose a practical methodology that would allow for the detection of farms with poor welfare for a given welfare measure.

Feasibility of the protocol

To be feasible, the monitoring system should involve a short duration, be relatively easy to perform on commercial conditions, and it should require little input from the farmers. The average time required for conducting the entire protocol on growing pigs present on farm was 6 hours 20 minutes. This may be considered to be an efficient system of welfare evaluation at farm level as it gives an accurate view of the compliance of a farm to the 12 criteria that constitute animal welfare. However, it may also be considered as excessively long in comparison with stakeholders' requirements (producers, retailers, certification agencies etc.). In the future, so as to be compatible with the expectations of the stakeholders, strategies to reduce the time required to conduct the farm visit should be undertaken. The interview, the only part of the protocol that required the farmer participation, takes roughly 40 minutes ranging from 25 to 60 minutes depending on the interest shown by the farmers. In addition, there was a variation in the time taken to record general information because the time spent making schemes of each room was much longer on big farms, taking 200 minutes on the biggest one. Having a general design of these types of farms before the visit would probably shorten this part of the protocol. Farm size and the distance between buildings also affected the time required for the QBA, scan sampling of behaviours, and the recording of feeding, housing and health measures. The number of animals per pen and the density, the behaviour of the animals (frightened or not), the dirtiness of the pigs and the light intensity within the buildings are other examples of factors that might have influenced the recording time of feeding, housing and health measures.

Feeding, housing and health animal-based measures

Poor body condition is the sole animal-based measure indicator of the good feeding principle considered by the overall Welfare Quality® protocol. Its low sensitivity within the logistic model in detecting farms with a welfare deficiency for prolonged hunger was expected under an intensive production system within which pigs are usually fed *ad libitum*. When reliant on natural foraging, pigs raised on unfavourable environmental conditions would probably present a higher incidence of poor body condition. In those conditions, the variability and sensitivity of the protocol about the good feeding principle is expected to increase.

Both scores of bursitis (moderate and severe bursitis) as well as moderately soiled body (from 20 to 50% of the body soiled with faeces) allowed discrimination among intensive conventional farms making “comfort around resting” one of the most sensitive criteria. Moderate and severe bursitis as well as moderately soiled body can be considered more frequent problems compared to all other measures. Moreover, when present on a farm, these measures tended to affect a large number of pens, resulting in low within-farm variability while between-farm variability was high. Even if on all farms pigs were housed on concrete floor, other factors as the quality of the concrete, stocking densities (Mouttotou et al., 1999), and pig age (Gillman et al., 2008) can explain the high between-farm variability in the apparition of bursitis. The prevalence of pigs with moderately soiled body (16.6 ± 13.4%) and severely soiled body (3.7 ± 6.3%) are similar to the prevalence reported by a survey of French farms with concrete floor which reported prevalence of 13.6 ± 14.8% and 3.6 ± 8.9% for each score of pig dirtiness, respectively (Courboulay et al., 2009). This similarity demonstrates that the mean prevalence of pig dirtiness tended to be consistent within a same production system independent of the geographic localization of the farm. It might be possible that factors affecting resting as the proportion of slatted area within a pen would have a greater impact on the apparition of pig dirtiness than fluctuation of environmental temperatures. However, further research on the impact of risk factors on the prevalence of pig dirtiness is needed.

Shivering, panting and huddling, which are animal-based indicators of thermal comfort, were either not present or shown prevalence lower than or close to 1%. On the days of assessment, temperatures ranged from 16ºC to 28ºC and appeared to be insufficiently extreme to induce shivering or panting in response to thermal environment. Huddling was more sensitive to thermal changes as compared with panting and shivering in the conditions found in the present study. It presented little dispersion of the data but its variability was increased by the presence of extreme values. According to the confidence limits established from the farms studied in this survey, a farm would be more likely to present a welfare deficiency when more than 4.2% of the animals assessed show huddling behaviour.

With the exception of coughs and sneezes, all other measures of health presented very low prevalence and little variability. Interpretations of this result depend on the type of measure studied.

For example, some measures from the Welfare Quality® protocol reflect a condition that is either present or absent, while others such as severely wounded animals depend on a threshold that had been established all through their validation in previous studies. Concerning measures that reflect a condition that is either present or absent, such as twisted snouts, pumping, rectal prolapse, or hernia, low prevalences indicate that there seems to be no major health problem among the visited farms. Twisted snouts are characteristic of atrophic rhinitis and low prevalence as the 0.10% registered in a study of Danish herds (Petersen et al., 2008), are expected. Pumping results from changes in the respiratory rate caused by a respiratory disorder, a dysfunction of other organs or an increased body temperature. Therefore, the occurrence of pumping in pigs should be considered as a serious respiratory sign. Twisted snouts and pumping should be considered as indicators of a severe deterioration of respiratory health. In contrast with these latter two uncommon measures, most swine production units experience cases of rectal prolapse (Straw et al., 2006). However, no cases of rectal prolapse were seen among the pigs assessed in this study. Intensive pig farmers tend to move to hospital pen pigs with rectal prolapse as well as pigs suffering severe hernia. However, these pens were not included with the sample of pens assessed. Consequently, the presence of a case of rectal prolapse or hernia on a farm assessed by the Welfare Quality® protocol reflects not only an important health problem but also a severe management problem, as affected pigs should be housed in hospital pens. Concerning health measures for which an applicable and feasible threshold had been established all through their validation in previous studies, such as severely wounded animals or lameness, the lack of sensitivity found in this survey might be unexpected as they are considered relatively frequent welfare concerns on intensive farms. The prevalence of wounded animals (0.9%) was consistent with the mean prevalence (0.5%) found by Courboulay et al. (2009) applying the same methodology. Using less restrictive scoring scales, other studies found much higher prevalences (Mullan et al., 2009, Whay et al., 2007). The number, severity and distribution of lesions considered in the scoring scale cause such differences. Therefore, to increase the sensitivity of the assessment of wounded animals without affecting the feasibility of its assessment, a scoring scale with a reduced number of scratches could be defined. The average prevalence of growing pigs with lameness (0.4%) was lower than the average prevalences found in other studies (Petersen et al., 2004, van den Berg et al., 2007). In comparison with these studies, moderate lameness was not recorded. Lameness was considered a reliable measure when pigs were individually taken out to a passage way and scored on a four-point scale in which moderate lameness was registered (Geverink et al., 2009). Nevertheless, as letting pigs out of the pen was not feasible in commercial conditions; moderate lameness was not taken into account because of the difficulty in evaluation from inside the pen. In the slaughterhouse, the prevalence of moderate lameness was 2.8% (Dalmau et al., 2009), comparable to results obtained by van den Berg et al. (2007), and 0.2% severe lameness, that were comparable to those encountered in the present study.

Thereafter, the assessment of lameness seems to be accurate but not very sensitive for feasibility reasons.

Behavioural measures

Emphasis must be given to the fact that all behaviours recorded on the social and exploratory scan were expressed as the total of active behaviours, thus excluding resting pigs. The percentage of active behaviours showed a high variability between farms and a significant variability between observation points of the same farm. This observation highlights the importance in expressing behaviours as the total of active behaviours and not as a percentage of total behaviours.

Occurrences of social behaviours recorded in this study were very similar to the values reported by Courboulay et al. (2009). According to these authors, positive and negative social behaviours accounted for 12.7% and 2.8% of all active behaviour, respectively, whereas corresponding occurrence of 12.2% and 5.4% were recorded in the present survey. This similarity emphasizes the reliability of the scan sampling methodology.

A significant variability among intensive conventional farms was found for positive and negative social behaviours. This variability between farms may result from the interaction of several factors. The extent of negative behaviours such as aggressions is mainly affected by management methods and design of the housing system. Potential sources of aggression include crowding on limited available space, access to a limited resource, mixing of unacquainted animals (Ewbank and Bryant, 1972; Kelley et al., 1980; Erhard et al., 1997). However, several other factors inherent to the animals, such as the age (Courboulay et al., 2009), the castration status (Fredriksen et al., 2008), and the breed (Breuer et al., 2003) may also contribute to differences among farms with respect to the expression of negative behaviour. Therefore, a combination of environmental-based factors and characteristics of the animals assessed contribute to the outcome of social behaviour. The respective impact of each factor on social behaviour is difficult to identify. The quantification of the variability between scan points should reveal the effect of environmental factors on the expression of social behaviours as influencing factors linked to the animals tended to be the same within a farm. For example, the majority of Spanish conventional farms follow an “all in all out” management; therefore, the age of the pigs during the day of survey was the same for all the animals. Also, on a same farm all the pigs were either castrated or not and of the same breed. Significant differences between scan points of negative social behaviour were seen on 30% of the farms. Consequently, each room within a farm represents a unit with specific combination of environmental factors that affected the expression of social behaviours. The methodology of assessment of social behaviours is not only sensitive to between-farm differences but also to within-farm influencing factors.

Therefore, possible causes of variability should be studied not only at farm level but also at room level.

Inferring at population level by means of confidence limits from the farms surveyed, a farm presenting more than 7.1% of negative interactions would be considered as more likely to present a welfare deficiency with regard to social negative behaviour. The threshold established by means of confidence limits allows the detection of the most significantly different farms of this survey.

There were also significant differences between intensive conventional farms in the frequencies of exploration. Providing manipulative materials increases the occurrence of explorations (Studnitz et al., 2007). However, in this study, only 8% of the pens surveyed provided tires, fixed chains or other material. Other environmental variables as the type of floor, feeding management, novelty (Studnitz et al., 2007), and the presence of suitable rooting material in the pens may stimulate the pigs to explore. All these factors may explain the variability in the occurrence of exploratory behaviour.

The human-animal relationship (HAR) test aims to detect the fear response of the animals toward human beings. The variability of this measure was too low to fit with the logit model proposed in this study. However, inferences can be made by means of confidence limits at population level. Using this method, a farm with more than around 30% of pens housing pigs showing a panic response should be considered as more likely to obtain a welfare deficiency for the fear of human criteria. Fear of humans is largely determined by the way stockpeople interact with their animals (Hemsworth et al., 1989). However, there are other important factors linked to the animal itself such as age (Dalmau et al., 2009) or genetic background (Boissy, 1995), and even factors linked to the environment such as the space allowance, which might affect the fear response of the animals toward the observer and therefore these factors should be taken into consideration when interpreting the results from the HAR test.

The qualitative behaviour assessment (QBA) may serve as an integrative methodology that could guide the interpretation of the quantitative assessment of behaviours of the scan observations. This approach deals with the complex notion of "quality of life" that reflects a dynamic approach, which inquires as to what animals like or prefer doing and what opportunities they have to fulfil these interests (Wemelsfelder, 2007). Several descriptors presented wide ranges indicating that differences between conventional farms might be identified. Nonetheless, these observations could not be verified because of the complexity of the statistical model.

CONCLUSIONS

Many animal-based measures assessed within the Welfare Quality® protocol applied to growing pigs on-farm present high variability and allow for discrimination among intensive conventional farms. Still, causes of this variability are difficult to interpret, especially for behavioural measures, and further studies on risk factors identification should be undertaken. Next, the methodology proposed to identify farms with a particular welfare problem appears to be useful, especially for animal-based measures that present low prevalence. This methodology does not pretend to define what should be considered as good or bad welfare but allows a practical detection of farms which are more likely to present a given welfare deficiency. Finally, the overall protocol is easy to carry out under commercial conditions and requires little input on the farmers. However, it involves an average time of 6 hours and 20 minutes per visit.

Assessment of farms from different production systems is necessary to increase the sensitivity of animal based measures from the protocol as well as to establish thresholds on a farm sample more representative at European level, and to evaluate the feasibility of the overall protocol, especially in extensive conditions.

ACKNOWLEDGEMENTS

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. The authors are grateful to the farmers, veterinarians and pig associations for their cordial collaboration. Déborah Temple was supported by an FPU scholarship from the Ministerio de Ciencia e Innovación of Spain.

REFERENCES

Botreau, R., Veissier, I., Butterworth, A., Bracke, M.B.M., Keeling, L.J., 2007. Definition of criteria for overall assessment of animal welfare. Anim. Welf. 16, 225-228.

Boissy, A., 1995. Fear and fearfulness in animals. Q. R. Biol. 70, 165-191.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in Generalized Linear Mixed Models. J. Am. Stat. Assoc. 88, 9-25.

Breuer, K., Sutcliffe, M.E.M., Mercer, J.T., Rance, K.A., Beattie, V.E., Sneddon, I.A., Edwards, S.A., 2003. The effect of breed on the development of adverse social behaviours in pigs. Appl. Anim. Behav. Sci. 84, 59-74.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Courboulay, V., Foubert, C., 2007. Testing different methods to evaluate pig welfare on farm. Anim. Welf. 16, 193-196.

Courboulay, V., Eugène, A., Delarue, E., 2009. Welfare assessment in 82 pig farms: effect of animal age and floor type on behaviour and injuries in fattening pigs. Anim. Welf. 18, 515-521.

Dalmau, A., Fabrega, E., Velarde, A., 2009. Fear assessment in pigs exposed to a novel object test. Appl. Anim. Behav. Sci. 117, 173-180.

Dalmau, A., Temple, D., Rodriguez, P., Llonch, P. Velarde, A., 2009. Application of the Welfare Quality® protocol at pig slaughterhouses. Anim. Welf. 18, 497-505.

Dawkins, M.S., 2004. Using behaviour to assess animal welfare. Anim Welf. 13, 3-7.

Erhard, H.W., Mendl, M., Ashley, D.D., 1997. Individual aggressiveness of pigs can be measured and used to reduce aggression after mixing. Appl. Anim. Behav. Sci. 54, 137-151.

Ewbank, R., Bryant, M.J., 1972. Aggressive behaviour amongst groups of domesticated pigs kept at various stocking rates. Anim. Behav. 20, 21-28.

Farm Animal Welfare Council, 1992. FAWC updates the five freedoms. Vet. Rec. 17, 357.

Fredriksen, B., Lium, B.M., Marka, C.H., Mosveen, B., Nafstad, O., 2008. Entire male pigs in farrow-to-finish pens - Effects on animal welfare. Appl. Anim. Behav. Sci. 110, 258-268.

Gillman, C.E.A., KilBride, L., Ossent, P., Green L.E., 2008. A cross-sectional study of the prevalence and associated risk factors for bursitis in weaner, grower and finisher pigs from 93 commercial farms in England. Prev. Vet. Med. 83, 308-322.

Geers, R., Dellaert, B., Goedseels, V., Hoogerbrugge, A., Vranken, E., Maes, F., Berckmans, D., 1989. An assessment of optimal air temperatures in pig houses by the quantification of behavioural and health-related problems. Anim. Prod. 48, 571-578.

Geverink, N.A., Meuleman, M., Van Nuffel, A., Van Steenbergen, L., Hautekiet, V., Vermeulen, K., Lammens, V., Geers, R., van Reenen, C.G., Tuyttens, F.A.M., 2009. Repeatability of a lameness score measured on farm. *Assessment of animal welfare measures for sows, piglets and fattening pigs*. In: Forkman, B., Keeling, L., (Eds.), Welfare Quality Reports, Nº. 10.

Hemsworth, P.H., Barnett, J.L., Coleman, G.J., Hansen, C., 1989. A study of the relationships between the attitudinal and behavioural profiles of stockpersons and the level of fear of humans and reproductive performance of commercial pigs. Appl. Anim. Behav. Sci. 23, 301-314.

Keeling, L., Veissier, I., 2005. Developing a monitoring system to assess welfare quality in cattle, pigs and chickens. *Science and society improving animal welfare*. In: Butterworth, A., (Ed), Welfare Quality conference proceedings. Brussels, pp 46-50.

Kelley, K.W., McGlone, J.J., Gaskins, C.T., 1980. Porcine aggression: Measurement and effects of crowding and fasting. J. Anim. Sci. 50, 336-341.

Meunier-Salaün, M.C.S., Courboulay, V., Scott, K., Guy, J., Edwards, S., 2007. Mesure de la relation homme-animal chez le porc en croissance: validation d'un test applicable en élevage. Journées de la Recherche Porcine 39, 85-86. [Title translation; The human-animal interaction: a validation test]

Molenberghs, G., Verbeke, G., 2005. Models for discrete longitudinal data. In: Spinger Science (Eds.), Springer Series in Statistics, New York.

Mouttotou, N., Hatchell, F.M., Green L.E., 1999. Prevalence and risk factors associated with adventitious bursitis in live growing and finishing pigs in south-west england. Prev. Vet. Med. 39, 39-52.

Mullan, S., Edwards, S., Butterworth, A., Whay, H.R., Main, D.C.J., 2009. Interdependence of welfare outcome measures and potential confounding factors on finishing pig farms. Appl. Anim. Behav. Sci. 121, 25-31.

Petersen, H.H., Enøe, C., Nielsen, E.O., 2004. Observer agreement on pen level prevalence of clinical signs in finishing pigs. Prev. Vet. Med. 64, 147-156.

Petersen, H.H., Nielsen, E.O., Hassing, A., Ersboll, A.K., Nielsen J.P., 2008. Prevalence of clinical signs of disease in danish finisher pigs. Vet. Rec. 162, 377-382.

Studnitz, M., Jensen, M.B. Pedersen, L.J., 2007. Why do pigs root and in what will they root?: A review on the exploratory behaviour of pigs in relation to environmental enrichment. App. Anim. Behav. Sci. 107, 183-197.

Straw, B.E., Zimmerman, J.J., D'Allaire, S., Taylor, D.J., 2006. Diseases of swine. Blackwell Publishing, Iowa.

van den Berg, A., Danuser, J., Frey, J., Regula, G., 2007. Evaluation of the acute phase protein haptoglobin as an indicator of herd health in slaughter pigs. Anim. Welf. 16, 157-159.

Velarde, A., Algers, B., Bracke, M.B.M., Courboulay, V., D'Eath, R., Edwards, S.A., Fàbrega, E., Forkman, B., Geers, R., Geverink, N.A., Gispert, M., Guy, J.H., Hautekiet, V., Keeling, L., Lammens, V., Lenskens, P., Meuleman, M., Meunier-Salaün, M.C., Millard, F., Nordensten, L., van Nuffel, A., van Reenen, C.G., Scott, K., Spoolder, H.A.M., van Steenbergen, L., Turner, S., Tuyttens, F.A.M., Vermeulen, K., Wemelsfelder, F., Dalmau, A., 2007. Full monitoring system in sows and piglets. In: Veissier I, Forkman B and Jones RB (eds.) *Assuring animal welfare: form societal concerns to implementation*. Second Welfare Quality stakeholder conference. Berlin, Ger.

Webster, A.J.F., Main, D.C.J., Whay, H.R., 2004. Welfare assessment: Indices from clinical observation. Anim. Welf. 13, 93-98.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25-31.

Whay, H.R., Main, D.C.J., Greent, L.E., Webster, A.J.F., 2003. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. Anim. Welf. 12, 205- 217.

Whay, H.R., Leeb, C., Maint, D.C.J., Green, L.E., Webster, A.J.F., 2007. Preliminary assessment of finishing pig welfare using animal-based measurements. Anim. Welf. 16, 209-211.

Winckler, C., Capdeville, J., Gebresenget, G., Hörning, B., Roiha, U., 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Anim. Welf. 12, 619-624.

Welfare Quality®, 2009. Welfare Quality® Assessment protocol for pigs (sows and piglets, growing and finishing pigs). In: Dalmau, A., Velarde, A., Scott, K., Edwards, S., Veissier, I., Keeling, L., Butterworth, A., (Eds.), Welfare Quality® Consortium, Lelystad, Neth.

CHAPTER 2

THE WELFARE OF GROWING PIGS IN FIVE PRODUCTION SYSTEMS: ASSESSMENT OF FEEDING AND HOUSING

Based on Temple D, Courboulay V, Manteca X, Velarde A, Dalmau A 2012. The welfare of growing pigs in five different production systems: assessment of feeding and housing. *Animal 6:* 656-667

ABSTRACT

Ninety one farms were visited over a two-year period to assess the welfare of growing pigs in five different production systems found whether in France or in Spain using the Welfare Quality® protocol. This study focused on animal-based measures indicators of “good feeding” and “good housing”. Multiple Generalized Linear Mixed Models were performed for each measure to evaluate differences between production systems and to detect possible causal factors. Pigs in the conventional system presented the lowest prevalence of poor body condition while extensive Mallorcan Black pigs and extensive Iberian pigs were associated with a decreased prevalence of bursitis and pig dirtiness. The straw bedded system presented a lower prevalence of bursitis but poorer hygiene and more susceptibility of poor body condition than the conventional system. The age of the animals had a significant effect on the appearance of bursitis in the three intensive systems studied. The type of floor was a significant causal factor of bursitis and pig dirtiness in the conventional system and among intensive Iberian pigs. The feeding system was another causal factor of pig dirtiness on more than 50% of the body in the conventional system whereas pig dirtiness on less than 50% of the body was influenced by the age of the animals. The prevalence of huddling animals in the conventional system was associated with the highest stocking densities and the lowest environmental temperatures. The results indicate that there were important differences between production systems based on animal-based indicators of the good feeding and housing principles. The recording of the age of the animals, type of floor, feeding system, stocking density and environmental temperature can be useful to predict the appearance of a given welfare measure of “good housing” on a farm.

Keywords: animal welfare, housing systems, growing pig, causal factors

IMPLICATIONS

Animal-based observations give the most direct insight into how animals are coping within their own environment, but it is essential that farmers understand their significance. A welfare assessment system should provide information to assess and compare the relative strengths and weaknesses of a production system, as well as to identify which factors are the main sources of variation within a given production system. Besides being fundamental in an intervention programme, a risk assessment can reveal the existence of possible confounding factors that should be taken into consideration to prevent misinterpretations. This study presents a descriptive epidemiological approach in animal welfare science across countries.

INTRODUCTION

Many studies seek to compare the welfare of pigs in different production systems. Since the Green Revolution, conventional farms with concrete floors and relatively high stocking densities displaced more traditional systems and greatly predominate throughout Europe. Recently, consumer concerns about animal welfare lead to an increasing interest for alternative production systems. Special attention has been given to the straw bedded system as straw is considered as one of the most appropriate ways in intensive conditions to meet the behavioural need of the pigs (Directive 91/630/EEC). Several studies have been conducted to compare the welfare of growing-finishing pigs in either conventional system or straw bedded accommodation (Lyons et al., 1995; Scott et al., 2006; Courboulay et al., 2009). Both systems show advantages and disadvantages that should be focused specifically in each production system. The proportion of farms using housing with straw bedding varies widely between countries. At present, around 7 % of growing pigs in France are housed on straw bedding while very few Spanish farms use deep litter systems for its a priori incompatibility with high environmental temperatures and limited availability of the material. Parallely, an increasing attention is given to outdoor production systems, which are usually found in specific locations. Mallorcan Black pigs and Iberian pigs represent approximately 15% of the overall population of growing pigs in Spain and are found in specific areas of the country. In both systems, pigs are reared extensively; still, noticeable differences between those two systems are seen in the way the animals are fed and managed (Jaume and Alfonso, 2000; Rodríguez-Estévez et al., 2009). Currently, Iberian pig system has greatly intensified leading to a reduction of the production cycle and the systematic use of fodder. Intensive Iberian pigs are housed in a wide variety of accommodations ranging from indoor slatted pens to outdoor paddocks. Such intensification is generally associated to a decrease in animal welfare; however much knowledge still remains to be transferred, especially in the Iberian pig sector (Aparicio Tovar and Vargas Giraldo, 2006).

Despite a general interest to compare production systems, few studies have been developed on commercial farms across countries. One of the main reasons is the lack of harmonized information that makes reliable the comparison of data from different sources. Welfare Quality® protocols provide the tool that makes possible the assessment of farms through Europe in a standardized way by observers that received an identical training. Moreover, the assessment system should be applicable and valid in a wide variety of farming systems (Blokhuis, 2008). Welfare Quality® contemplates four main principles of animal welfare: i) Good feeding, ii) Good housing, iii) Good health and iv) Appropriate behaviour. Each of these four principles comprises several independent but complementary criteria (Botreau et al., 2007) in turn characterized by one or various measures. Preference was given to animal-based measures for being more valid when measuring the state of the animal as regards to its attempts to cope with its environment (Capdeville and Veissier, 2001; Whay et al., 2003). Animal-based measures give an indication of the performance of a husbandry system at a certain moment of the production cycle. Nonetheless, resource-based factors are still essential for the provision of advice on the prevention of a welfare problem and on the evaluation of the risk of deficient welfare. It is accepted that the most valid assessment of welfare is obtained combining both types of measures (Johnsen et al., 2001).

In addition to the comparison between production systems, much attention has been given recently to the study of risk factors in animal welfare (Smulders, 2009). Within a same housing system, farms can differ for many welfare indicators and intra-farm variability may also be high (Courboulay et al., 2009; Temple et al., 2011). Even though a given housing system may constitute a risk factor for the welfare of the animal by itself; possible explanations should be given to understand differences between farms from a given production system. Some variables may represent a risk in all production systems while other ones simply make no sense in a specific system. Those causal factors can be useful to offer proper advice to farmers to improve the welfare of their animals. Key causal factors can also give an estimation of which type of welfare problem are more likely to appear on a given farm. The visit could then be focused on the target animal-based measures. Finally, identification of possible causal factors may reveal possible confounding factors.

"Good feeding" and "housing" are the first two principles contemplated by the Welfare Quality® protocol. "Good feeding" includes the absence of prolonged hunger and absence of prolonged thirst while "good housing" refers to three criteria: comfort around resting, thermal comfort and ease of movement. Poor body condition represents a valid and feasible indicator for malnutrition or under nutrition. Dehydration could not be defined in terms of animal-based measures into the Welfare Quality® project for its lack of feasibility. Several animal-based measures indicators of housing have been selected for their validity and repeatability in pilot studies. Bursitis has been widely studied either in epidemiological surveys (Mouttotou et al., 1999; Gillman et al., 2008) or in experimental conditions (Lyons et al., 1995; Scott et al., 2006). Prevalences of pig dirtiness have been compared between different types of flooring (Courboulay et al., 2003; Scott et al., 2006) and factors that affect excretory behaviour as well as lying behaviour have been identified under controlled conditions (Geers et al., 1986; Hacker et al., 1994). However, as it happens for huddling behaviour, pig dirtiness has been little studied under commercial conditions.

This paper reports part of a welfare assessment study carried out on a sample of commercial farms from five different production systems of growing pigs by a team of French and Spanish researchers. It does not pretend to define what should be considered as good or bad welfare. Instead, it presents a description of several welfare outcomes and possible causal factors on a wide variety of farms. Such a descriptive approach may form the basis of hypotheses that could be tested in future explanatory studies. The overall Welfare Quality® protocol was applied on 91 farms and the results from the “good feeding” and “housing” principles are presented in this paper.

MATERIAL AND METHODS

Farm selection and housing systems

A total of 91 farms of growing pigs were assessed applying the Welfare Quality® protocol over a two year period (2007-09). Sixty one of these farms were evaluated in Spain and 30 in France. Selection of farms was based on management practices, farm size and veterinary records. Farm sampling concerned the main production systems of growing pigs found whether in France or in Spain. Intensive systems included 52 Spanish and French conventional farms with concrete floor, eight straw bedded farms in France and 10 intensive farms of Iberian pigs in Spain. Extensive systems included 11 Iberian extensive units and 10 extensive Mallorcan Black pig exploitations. Data were collected on the feeding system (liquid feed in trough, dry hopper, wet feed hopper, turbomat, dry food on the ground, on pasture), average environmental temperature and on the type of floor (fully slatted floor, partly slatted floor, solid concrete floor, straw or sand). All fully slatted floors were concrete slatted floors. Space allowance was calculated taking into account the length and the width of the pen or paddock in relation to the number of animals housed in that pen or paddock.

Sampling of animals

On each intensive farm, 10 pens of growing pigs were randomly selected. If there were less than 15 pigs in a pen all pigs were assessed, whereas if there were more than 15 pigs in the pen, 15 were arbitrarily selected. Pigs were categorized into three growing stages:

- First growing stage: at the beginning of the growing process (less than 90 days old).
- Second growing stage: mid growing period from 90 to 130 days old.
- Third growing stage: at the end of the growing process and before leaving to the slaughterhouse (from 130 to 180 days old).

Among intensive Iberian pigs as well as extensive pigs, a fourth category included pigs older than 180 days old. Information on whether male pigs were surgically castrated or not were also taken into account.

Pig and pen characteristics

A total of 11 647 pigs from 719 pens or paddocks were evaluated.

In the conventional system, 7 030 pigs from 530 pens were assessed. On the day of assessment, pig age ranged from 29 days old to 180 days old (SD = 34). Among pigs kept on concrete floor, 64% were housed on fully slatted floors and 36% on partly slatted floor without bedding. Space allowance ranged from 0.20 to 1.56 m^2 per pig (median = 0.68, Q1 = 0.58, Q3 = 0.75). As for the feeding system, 22% of the pigs were liquid fed in trough two to four times a day, 38% were fed by dry hopper *ad libitum*, 37% by wet feed hopper *ad libitum* and 3% by turbomat once or twice a day. Finally, environmental temperatures ranged from 16 to 29ºC (median = 23, Q1 = 21, Q3 = 25). French and Spanish conventional systems were considered as the same production system; however some differences in the distribution of several environmental variables should be considered. More than 90% of concrete floors on French farms were fully slatted floors; liquid feeding system in trough was only seen in France; and all the temperatures were recorded above 21ºC in France.

In the straw bedded system, 1 110 pigs of an average of 112 days old (SD = 42.0) ranging from 41 days old to 194 days old, were assessed in 74 pens. Space allowance ranged from 0.3 to 3.0 m^2 per pig (median = 1.2, Q1 = 0.7, Q3 = 2.9). 19% of the animals were liquid fed in trough, 59% were fed by dry hopper and 22% by wet feed hopper. Feed allowance was the same as in the previous production system. In straw bedded pens, environmental temperatures ranged from 15 to 40ºC (median= 23, Q1 = 20, Q3 = 24). This system was assessed only in France.

A total of 1 255 Iberian pigs kept in intensive conditions were evaluated in 58 pens or paddocks. Pig age ranged from 37 days old to 280 days old (SD = 80). 16% of intensive Iberian pigs were housed on fully slatted floors, 18 % on partly slatted floors without bedding, and 47% on concrete floor with whether a resting area on straw or an outdoor access, and 19 % housed in outdoor paddocks on deep sand or straw. Space allowance ranged from 0.30 to 5.4 m^2 per pig (median = 1.4, Q1 = 0.8, Q3 = 2.6). All the pigs were dry fed *ad libitum* by dry hopper (86%) or by trough (14%). Environmental temperatures ranged from 12 to 24ºC (median = 22, Q1 = 19, Q3 = 23). This system was assessed only in Spain.

Under extensive conditions, 1 428 Iberian pigs of an average 250 days old (SD = 97) ranging from 42 days old to 420 days old were assessed in 31 different paddocks. The average space allowance per pig was 430 m^2 (SD = 578). On the days of assessment pigs were supplemented with fodder as the availability of natural resources was insufficient. Supplementation was given manually once or twice a day whether on the floor (55%) in a trough (26%) or dry hopper (19%). The Mallorcan Black pig production system was the second extensive system assessed. A total of 824 Mallorcan Black pigs of an average 220 days old (SD = 100) ranging from 75 days old to 405 days old were evaluated in 24 paddocks. The average space allowance per pig was 692 m^2 (SD = 818). All pigs were on pasture and were supplemented with household refusals, bran, legume seeds, cereals and figs. Supplementation was done manually, once or twice a day, whether on the floor (18%) or in trough (30%) and given *ad libitum* in dry hopper (47%). The remaining 5% were entirely reliant on pasture. During the days of assessment, the average environment temperature was 18ºC and 25ºC in extensive Iberian and Mallorcan Black pigs, respectively.

Measurements

Assessments were performed by four observers, two of them in Spain and the other two in France. In order to minimize difference between observers and to standardize the scores from the visits, observers followed the same training prior beginning the assessment.

Poor body condition

Pigs were visually examined in the pen and any pig with a poor condition (visible hip bones and backbone) was counted. The assessment was done as follows: 0 – No visible hip bones and backbone; 2 – Visible hip bones and backbone.

Bursitis

Bursitis was visually examined from a maximum distance of one meter from the animal. Only one side of the animal was inspected. Each animal was considered using the following individual scale: 0 – no evidence of bursa/swelling; 1 (moderate) – one or several small bursa (comparable in size to a grape; 1.5 - 2.0 cm) or one large bursa (comparable in size to a walnut; 3.0 - 5.0 cm diameter); 2 (severe) – several large bursa, or one extremely large bursa (comparable to a tangerine; 5.0 - 7.0 cm diameter), or any eroded bursa.

Pig dirtiness

Pig dirtiness (manure on the body) was defined as how soiled the pig was with faeces. Only the presence of manure was assessed; pigs soiled with mud were not considered as dirty. Only one side of the body of individual pig was assessed. The three categories were: 0 – less than 20% of the body surface was soiled; 1 (moderately soiled body) – more than 20% but less than 50% of the body surface was soiled; 2 (severely soiled body) – more than 50% of the body surface was soiled.

Huddling, shivering and panting

Huddling was the first measure done, only in resting animals, just before the animals began to stand up as a response to the observer presence. This measure was assessed visually from the corridor. Huddling was considered when a pig was lying with more than half of its body in contact with another pig and any part of the body was in top of the other. It was not considered huddling when an individual was just side by side besides another animal.

Shivering was defined as the slow and irregular vibration of any body part, or the body as a whole. This measure was assessed visually from the corridor and considered for each animal assessed as present or absent.

Panting was defined as a rapid breath in short gasps. This parameter was assessed visually from the corridor and considered for each animal assessed as present or absent.

More detailed information about the overall methodology of assessment can be found in the Welfare Quality® Assessment protocol for pigs (2009).

Statistical Analysis

The pen or paddock was the experimental unit. Data were expressed as the number of animals affected out of the number of animals assessed in each unit. To account for possible dependence between observations on pens from the same farm, random farm effects were included into the model. Data were also clustered at pen level and modelled for over-dispersion. Multiple Generalized Linear Mixed Models for binomial data were performed separately for each welfare measure using the GLIMMIX procedure of the SAS statistical package Version 9.1 (SAS Inst.Inc., 2002, Cary, NC, USA). Residual pseudo-likelihood was used as estimation technique (Wolfinger and O'Connell, 1993).

The general form of the model was:

$$\text{Logit}\,(y/n) = \beta_0 + \beta_1 X_{1\,i} + \ldots + \beta_k X_{k\,i} + u_{\text{farm}\,(i)} + v_{\text{pen}\,(i)}$$

where y is the number of animals affected, n the number of animals assessed, β_0 the intercept, β_k the regression coefficient predictors, $X_{k\,i}$ are the causal factors, and u_{farm} and v_{pen} are the residual variances of the random effects.

First, models were built to compare the five production systems of growing pigs: conventional, straw bedded, intensive Iberian, extensive Mallorcan Black pig, extensive Iberian. Then, to avoid misinterpretations, separate models were developed for each intensive system to identify possible causal factors such as the type of floor, the feeding system, the growing stage, stocking density and environment temperature. At the expense of throwing away information, these two last continuous variables (density and temperature) were converted into categorical variables by means of quartiles to avoid strong assumptions about the linear relation between the outcome measure and the causal factor (Altman et al., 1994). Causal variables were entered into the model as fixed effects.

Statistical analyses consisted of initial univariable screening to determine variables statistically associated with each welfare indicator. Variables were taken forward for multivariable analysis when significant at $P < 0.2$ (Dohoo et al., 2009). Where variables were highly correlated (PROC SPEARMAN) or showed a strong association (PROC FREQ with the CHISQ option), the most biologically plausible variable was selected for inclusion in the final model. Stepwise backward selection was performed to identify the variables that had a significant association ($P < 0.05$) with the outcome measure. The country effect was considered into each model of the conventional system. When the country effect vanished in the analysis while other factors were added to the model, "country" ceased to be an interest as a separate factor. Observer effect was also included into each final model and taken out when it did not alter the interpretation of fixed effect. The ratio of the generalized chi-square statistic and its degrees of freedom was considered to check how well the variability in the data had been properly modelled and that there was no residual over-dispersion. Goodness of fit was also assessed visually by means of standardized Pearson residuals (Lee et al., 2006). When data did not allow a proper modelling, identification of causal factors was not done.

Spearman correlation coefficients were calculated to investigate the association between the different animal-based measures from the good housing principle.

RESULTS

Table 1 Prevalences and standard errors according to the production system studied, for each animal-based measure of the good feeding and housing principles of Welfare Quality®

Animal-based measure	Conventional	Straw bedding	Intensive Iberian	Extensive Mallorcan Black pig	Extensive Iberian
Poor body condition	0.4+/- 0.14^{b}	0.9+/- 0.31^{a}	0.9+/- 0.38^{ab}	1.9+/- 0.75^{a}	1.0+/- 0.58^{ab}
Moderate bursitis	43.5+/- 1.01^{a}	3.8+/- 0.62^{bc}	14.4+/- 2.20^{b}	2.1+/- 0.74^{c}	7.3+/- 0.98^{bc}
Severe bursitis	7.8+/- 0.53^{a}	0.2+/- 0.13^{b}	1.7+/- 0.68^{b}	0.0+/- 0.00	0.3+/- 0.22^{b}
Moderately soiled body	23.5+/- 1.06^{a}	30.8+/- 2.62^{a}	9.6+/- 2.16^{a}	0.3+/- 0.20^{b}	0.9+/- 0.62^{b}
Severely soiled body	5.3+/- 0.60^{b}	15.6+/- 2.65^{a}	10.7+/- 3.76^{ab}	0.1+/- 0.04^{c}	0.7+/- 0.56^{c}
Huddling	3.5+/- 0.60	1.9+/- 0.84	0.0+/- 0.00	0.0+/- 0.00	0.0+/- 0.00
Shivering	0.0+/- 0.00	0.1+/- 0.11	0.0+/- 0.00	0.0+/- 0.00	0.0+/- 0.00
Panting	0.2+/- 0.07	0.7+/- 0.27	0.0+/- 0.00	0.9+/- 0.54	0.0+/- 0.00

Values within rows with different letters are significantly different

Poor body condition

Differences in the prevalence of poor body condition between production systems

Prevalences of poor body condition among pigs from different production systems are shown in **Table 1**. There was an increased prevalence of poor body condition in pigs in the straw bedded system and extensive Mallorcan Black pigs compared with pigs in the conventional system (**Table 2**).

Table 2 Logistic-regression model of poor body condition of growing pigs for production system

System	OR	CI
Conventional	1	
Straw bedded	6.3	1.8, 22.3
Intensive Iberian	2.3	0.6, 8.6
Extensive Iberian	1.7	0.4, 6.4
Extensive Mallorcan Black pig	4.3	1.2, 15.9
Random effect	*Estimate*	*SE*
Farm	2	0.5
Pen	0.75	0.04

Intercept coefficient= - 6.2

Odds ratios (OR); 95% confidence intervals (CI); standard errors (SE)

Prevalences associated with poor body condition

Among extensive Mallorcan Black pigs, the highest prevalence of poor body condition was recorded in pigs entirely dependent on pasture (3%) while pigs supplemented with cereals in trough showed the lowest prevalence (0.4%). In the straw bedded system, pigs that finalized their growing period presented the lowest prevalence of poor body condition (0.1%). Prevalence and variability of poor body condition were considered insufficient to undergo a correct modelling for this measure to detect possible causal factors.

Bursitis

Differences in the prevalence of bursitis between production systems

Prevalences of moderate and severe bursitis among pigs from different production systems are shown in **Table 1**. Pigs on conventional farms presented the highest prevalence of moderate bursitis while extensive Mallorcan Black pigs presented the lowest prevalence. Considering severe bursitis, pigs housed on conventional farms presented the highest prevalence. None severe bursitis was registered among extensive Mallorcan Black pigs. The odds of having a severe bursitis was estimated as being 42.1 times greater in pigs housed in the conventional system than in extensive Iberian pigs (**Table 3**).

Table 3 Logistic-regression models of moderate and severe bursitis as well as for both types of pig dirtiness (moderate and severe soiled body) for production system of growing pigs

	Moderate bursitis [a]		Severe bursitis [b]		Moderately soiled body [c]		Severely soiled body [d]	
System	OR	CI	OR	CI	OR	CI	OR	CI
Extensive Iberian	1		1		1		1	
Conventional	10.6	4.8, 23.3	42.1	4.3, 413.0	56.5	7.9, 403.9	11.7	1.2, 112.6
Straw bedded	0.6	0.3, 1.8	1.0	0.1, 25.5	67.0	8.1, 549.2	39.6	3.3, 472.9
Intensive Iberian	1.4	0.5, 4.1	6.2	0.5, 82.9	21.1	2.4, 185.6	20.7	1.7, 247.2
Extensive Mallorcan Black pig	0.3	0.1, 1.1	-	-	1.0	0.1, 23.8	0.4	0.01, 62.6
Random effect	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
Farm	0.6	0.1	1.2	0.3	1.2	0.3	1.0	0.2
Pen	1.5	0.1	1.2	0.1	2.7	0.1	2.7	0.1

[a] Intercept coefficient= -2.7
[b] Intercept coefficient= - 2.9
[c] Intercept coefficient= - 5.4
[d] Intercept coefficient= -5.7
Odds ratios (OR); 95% confidence intervals (CI); standard errors (SE)

Prevalences and possible causal factors associated with moderate and severe bursitis

In the conventional system, prevalence of moderate and severe bursitis varied by growing stage, stocking densities and type of floor as detailed in **Table 4**. There was an increased risk of moderate and severe bursitis (**Table 5**) in pigs in the mid growing period (average 107 days old) and finishing pigs (average 150 days old) compared with pigs that initiate their growing period (average 68 days old). Moreover, pigs housed in concrete pens with the higher stocking densities (1^{st} Q: average 0.58 m^2/animal) presented a lower risk of moderate bursitis whereas space allowance did not have a significant effect on the appearance of severe bursitis. There was a trend for increased bursitis among pigs on fully slatted floors; however, the confidence interval contained the unity.

The prevalence of moderate bursitis among intensive Iberian pigs varied by growing stage and floor type (**Table 4**). The risk of moderate bursitis (**Table 5**) increased among Iberian pigs housed on fully slatted floor. The odd of moderate bursitis decreased as the pigs get older (average 267 days old).

In the straw bedded system, the highest prevalence of moderate bursitis was registered in the mid growing period (6.0%) and finishing period (4.5%) while pigs that entered the growing period showed the lowest prevalence (0.9%). Pigs on straw in their mid and final growing stages presented an increased risk of moderate bursitis compared with pigs that initiate their growing stage (OR 6.9; 2.1, 22.1).

Table 4 Number and prevalence of growing pigs in the conventional system and intensive Iberian pigs with moderate and severe bursitis by variable selected for multivariate analysis

	Conventional				Intensive Iberian	
	Moderate bursitis		Severe bursitis		Moderate bursitis	
Variable	n	%	n	%	n	%
Growing stage (days old)						
category 1: < 90 days old	616	30.8	60	3	83	18.0
category 2: 90 – 130 days old	1200	42.2	215	7.6	17	29.5
category 3: 131 – 180 days old	1161	58.5	239	12.1	45	15.8
category 4: > 180 days old					35	7.8
Space allowance (m^2 per pig)						
Q1	590	28.9	99	4.9		
Q2	691	45.9	144	9.6		
Q3	886	52.8	162	9.6		
Q4	801	49.8	114	7.1		
Floor type						
Partly slatted	1187	42.1	90	3.2	22	9.6
Fully slatted	1816	45.2	432	10.8	62	29.8
Concrete with resting area					56	9.7
Outdoor on deep bedding					7	3.1

Table 5 Multivariate logistic binomial mixed models of possible causal factors associated with moderate and severe bursitis in the conventional system and intensive Iberian pigs

	Conventional				Intensive Iberian	
	Moderate bursitis [a]		Severe bursitis [b]		Moderate bursitis [c]	
Variable	OR	CI	OR	CI	OR	CI
Growing stage (days old)						
category 1: < 90 days old	1		1		1	
category 2: 90 – 130 days old	1.6	1.3, 2.0	3.2	2.1, 4.8	0.6	0.2, 1.6
category 3: 131 – 180 days old	2.2	1.7, 2.8	5.1	3.4, 7.8	0.9	0.5, 1.9
category 4: > 180 days old					0.3	0.1, 0.6
Space allowance (m^2 per pig)						
Q1	1		1			
Q2	1.3	1.1, 1.7	1.5	0.9, 2.3		
Q3	1.4	1.1, 1.7	1.2	0.8, 1.8		
Q4	1.3	1.1, 1.7	1.2	0.8, 1.9		
Floor type						
Partly slatted	1		1		1	
Fully slatted	1.3	1.0, 1.6	1.5	1.0, 2.3	3.4	1.7, 6.9
Concrete with resting area					0.5	0.1, 1.7
Outdoor on deep bedding					0.6	0.1, 3.1
Random effect	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
Farm	0.5	0.1	0.7	0.2	0.5	0.4
Pen	1.3	0.1	1.2	0.1	1.3	0.3

[a] Intercept coefficient= 0.002
[b] Intercept coefficient= -3.1
[c] Intercept coefficient= - 3.6
Odds ratios (OR); 95% confidence intervals (CI); standard errors (SE)

Pig dirtiness

Differences in the prevalence of manure on the body (pig dirtiness) between production systems

The prevalence of manure on less than 50% of the body (moderately soiled body) was significantly higher in intensive systems compared with extensive systems (**Table 1**). Considering manure on more than 50% of the body (severely soiled body), extensive Mallorcan Black pigs and extensive Iberian pigs presented also the lowest prevalences. Moreover, among intensive systems, the straw bedded system presented significantly a higher prevalence of manure on more than 50% of the body compared with the conventional system. The odd of being moderately soiled was estimated as being 67 times greater in pigs on-straw than in extensive Iberian pigs (**Table 3**).

Prevalences and possible causal factors associated with pig dirtiness

The prevalence of manure scored as 1 (moderately soiled body) and scored as 2 (severely soiled body) in the conventional system varied according to the type of floor, the feeding system and environmental temperature levels as detailed in **Table 6**. Moderately soiled body also varied depending on the growing stage whereas it was not identified as a possible causal factor for severely soiled body. In the conventional system, there was an increased risk of both scores of manure on the body (**Table 7**) in partly slatted floor compared with pigs housed on fully slatted floor. Pigs in their final growing stage presented higher odds of moderately soiled body compared with pigs that entered the growing period. There was a significant increased risk of severely soiled body among liquid fed pigs in trough compared with pigs fed by wet feed hopper. Environmental temperatures did not have a significant effect on the apparition of pig dirtiness. The factor country was included into the final models of both types of pig dirtiness as it had a relevant effect on the final results of the analysis.

Among intensive Iberian pigs, the highest prevalences of moderately soiled body and severely soiled body (16% and 32%, respectively) were found in pigs housed on partly slatted floor while the lowest prevalences were observed in pigs on fully slatted floor (0.4% and 0.0%, respectively). The type of floor was the sole significant causal variable. Iberian pigs on partly slatted floor presented a significant higher odd of moderately soiled body compared with pigs on fully slatted pens (OR 36.1; 2.9, 448.8).

In the straw bedded system, the highest prevalence of severe soiled body (27%) was registered among finishing pigs compared to pigs that initiate their growing period (8%). The growing stage was the sole variable identified as significant causal variable of the prevalence of pig dirtiness in pigs on straw bedding. The odd of moderately soiled body (OR 2.1; 1.2, 3.7) and severely soiled body (OR 4.6; 1.8, 11.8) increased among older pigs compared with younger ones.

Table 6 Number and prevalence of growing-finishing pigs in the conventional system with moderate soiled body and severe soiled body by variable selected for multivariate analysis

	Conventional			
	Moderately soiled body		Severely soiled body	
Variable	n	%	n	%
Country				
Spain	643	16.6	145	3.8
France	955	32.3	227	7.6
Growing stage (days old)				
category 1: < 90 days old	428	21.4		
category 2: 90 – 130 days old	705	24.8		
category 3: 131 – 180 days old	447	22.5		
Floor type				
Partly slatted	992	24.7	171	6.1
Fully slatted	590	20.9	197	4.9
Feeding system				
Liquid feed in trough	514	37.4	141	10.3
Dry hopper	484	18.2	129	4.8
Wet feed hopper	550	21	94	3.6
Turbomat	29	16.1	1	0.6
Temperature (ºC)				
Q1: 16 – 20.9	237	14.8	44	2.7
Q2: 21 – 22.9	388	21.7	73	4.1
Q3: 23 – 24.9	331	29.3	109	9.7
Q4: 25 - 29	631	27.2	143	6.2

Table 7 Multivariate logistic binomial mixed models of possible causal factors associated with both types of pig dirtiness (moderate and severe soiled body) in the conventional system

	Conventional			
	Moderately soiled body [a]		Severely soiled body [b]	
Variable	OR	CI	OR	CI
Country				
Spain	1		1	
France	6.6	3.0, 14.3	6.5	2.2, 19.8
Growing stage (days old)				
category 1: < 90 days old	1			
category 2: 90 – 130 days old	1.3	0.9, 1.9		
category 3: 131 – 180 days old	1.6	1.1, 2.4		
Floor type				
Partly slatted	1		1	
Fully slatted	0.3	0.2, 0.6	0.1	0.02, 0.1
Feeding system				
Liquid feeding by trough	1		1	
Dry hopper	0.6	0.4, 1.0	0.6	0.3, 1.1
Wet feed hopper	1	0.5, 2.1	0.2	0.1, 0.6
Turbomat	0.4	0.1, 1.1		
Temperature (ºC)				
Q1: 16 – 20.9	1		1	1
Q2: 21 – 22.9	1	0.6, 1.5	0.8	0.4, 1.8
Q3: 23 – 24.9	0.9	0.5, 1.6	2.1	0.9, 5.3
Q4: 25 - 29	1.2	0.7, 2.0	0.8	0.3, 2.0
Random effect	*Estimate*	*SE*	*Estimate*	*SE*
Farm	0.8	0.2	1.1	0.3
Pen	2.8	0.2	2.2	0.1

[a] Intercept coefficient= -2.1
[b] Intercept coefficient= -3.9
Odds ratios (OR); 95% confidence intervals (CI); standard errors (SE)

Huddling

Differences in the prevalence of huddling between production systems

No huddling animal was registered among extensive Iberian pigs, intensive Iberian pigs and extensive Mallorcan Black pigs (**Table 1**). The prevalence of huddling animals housed on conventional farms was not significantly different from the prevalence found in pigs in the straw bedded system (OR 1.7, 0.4, 7.0).

Prevalences and possible causal factors associated with huddling behaviour

The prevalence of huddling animals in the conventional system varied by growing stage, stocking densities, type of floor, and temperature levels (**Table 8**). Prevalences of huddling animals by causal variable are shown for each country to have a better view of the patterns of prevalences. There was an increased risk of huddling behaviour (**Table 9**) associated with the highest stocking densities. Moreover, there was an increased risk of huddling among pigs housed in the lowest environmental temperatures compared with pigs housed in the highest environmental temperatures. There were no significant differences in the risk of huddling between growing stages and between types of floors as the confidence intervals (CI) included unity ($P > 0.05$). The factor country was included into the final model of huddling behaviour as it had a relevant effect on the final results of the risk analysis of this measure presenting a much higher prevalence in France. None significant causal factor could be identified among pigs in the straw bedded system.

Table 8 Number and prevalence of growing-finishing pigs huddling in the concrete conventional system by variable selected for multivariate analysis

	Conventional					
	France		Spain		Total	
Variable	n	%	n	%	n	%
Country						
Spain					49	1.4
France					194	6.5
Growing stage (days old)						
category 1: < 90 days old	123	10.6	11	1.4	130	6.6
category 2: 90 – 130 days old	50	4.8	38	2.3	88	3.3
category 3: 131 – 180 days old	24	3.1	1	0.1	24	1.3
Space allowance (m^2 per pig)						
Q1	154	11.8	20	2.8	175	8.7
Q2	20	3.1	16	2.0	36	2.5
Q3	24	3.4	6	0.7	29	1.8
Q4	0	0	10	0.8	10	0.7
Floor type						
Partly slatted	0	0	11	1.0	38	1.5
Fully slatted	195	6.8	38	1.5	199	5.1
Temperature (ºC)						
Q1: 16 – 20.9	0	0	28	1.9	28	1.9
Q2: 21 – 22.9	42	11.8	18	1.4	58	3.5
Q3: 23 – 24.9	35	3.9	0	0	34	3.1
Q4: 25 - 29	117	6.9	3	0.6	119	3.2

Table 9 Logistic binomial mixed models of possible causal factors associated with huddling behaviour in the conventional system

	Conventional	
	Huddling	
Variable	OR	CI
Country		
Spain	1	
France	7.5	2.1, 27.4
Growing stage (days old)		
category 1: < 90 days old	1	
category 2: 90 – 130 days old	0.8	0.4, 1.6
category 3: 131 – 180 days old	0.7	0.3, 1.5
Space allowance (m^2 per pig)		
Q1	1	
Q2	0.3	0.1, 0.6
Q3	0.2	0.1, 0.5
Q4	0.1	0.01, 0.2
Floor type		
Partly slatted	1	
Fully slatted	2.1	0.7, 6.5
Temperature (ºC)		
Q1: 16 – 20.9	1	
Q2: 21 – 22.9	0.7	0.2, 2.1
Q3: 23 – 24.9	0.3	0.1, 1.1
Q4: 25 - 29	0.2	0.1, 0.7
Random effect	*Estimate*	*SE*
Farm	0.9	0.4
Pen	2.4	0.2

Intercept coefficient= - 7.8

Odds ratios (OR); 95% confidence intervals (CI); standard errors (SE)

Shivering and panting

Prevalences of shivering and panting were under 1% and no significant differences between systems were found for any of these measures (**Table 1**). Identification of possible causal variables was not possible for these measures.

Associations between animal-based measures of housing

In the conventional system, the strongest statistical correlation was found between both scores of pig dirtiness (r = 0.5; **Table 10**). Other correlated variables were statistically significant at low values (r = 0.02 – 0.30). In intensive Iberian pigs, the highest statistical correlation was also found between both types of pig dirtiness (r = 0.6; **Table 11**). Additionally, both scores of manure on the body were negatively correlated with moderate bursitis in intensive Iberian pigs (r = -0.5). In the straw bedded system, the highest correlation was found between moderate soiled body and huddling (r = - 0.2, P < 0.05).

Table 10 Spearman correlation coefficients between the different animal-based measures from the good housing principle evaluated on 52 intensive conventional farms in France and in Spain

	Moderate bursitis	Severe bursitis	Moderately soiled body	Severely soiled body	Huddling
Moderate bursitis	1				
Severe bursitis	0.3***	1			
Moderately soiled body	-0.2**	0.1**	1		
Severely soiled body	-0.02	0.2**	0.5***	1	
Huddling	-0.04	0.01	-0.1	0.1	1

** correlated at P < 0.01
*** correlated at P < 0.001

Table 11 Spearman correlation coefficients between the different animal-based measures from the good housing principle evaluated on 10 intensive Iberian pig farms

	Moderate bursitis	Severe bursitis	Moderately soiled body	Severely soiled body
Moderate bursitis	1			
Severe bursitis	0.2	1		
Moderately soiled body	-0.5***	0.1	1	
Severely soiled body	-0.5***	0.2	0.6***	1

*** correlated at P <0.001

Associations between possible causal factors

In the straw bedded system, the growing stage and the space allowance were highly correlated (r = 0.7, P < 0.001). This correlation was weak in the conventional system (r = 0.3, P < 0.001). Among intensive Iberian pigs, there was an association between the type of floor and the space allowance. Fully slatted and partly slatted floors were associated to lower space allowance while concrete floor with bedding or outdoor access and outdoor deep bedding floors housed pigs in higher space allowance (chi-square = 63, df = 9, phi-coefficient = 0.9, P < 0.001). Inclusion of the observer effect did not alter the interpretation of any of the fixed effect.

DISCUSSION

Poor Body Condition

Poor body condition has been little studied as a welfare indicator of growing pigs. In the Welfare Quality® protocol, poor body condition is the sole animal-based measure indicator of “good feeding”. Extensive Mallorcan Black pigs presented a higher risk of poor body condition compared with pigs in the conventional system. Mallorcan Black pigs are reared following traditional methods based on profitability of resources. Five percent of the pigs were entirely dependent on pasture which might have increased the risk of poor body condition. On the other side, pigs in the straw bedded system that were fed in more controlled conditions were also associated with an increased risk of poor body condition. Scott et al. (2006) reported more prevalent loss of body condition in the straw bedded system than in the fully slatted system and associated it with poorer hygiene and wasted animals, generally associated with Postweaning Multisystemic Wasting Syndrome (PMWS). Knowing that PMWS most commonly affects pigs at 2- 4 months of age (Segalés et al., 2006), this hypothesis is reinforced in the present study by the fact that few cases of poor body conditions were recorded at the end of the growing period.

Bursitis

A bursa is a fluid filled sac which arises in the subcutaneous connective tissue as a result of exudation of fluid from traumatized capillaries and lymphatic vessels after pressure on the skin over a bony prominence (Smith, 1993; Mouttotou et. al, 1998). Moderate and severe bursitis are indicators of comfort around resting; however, each type of bursitis was considered separately for reflecting different aspects of discomfort. Pigs in the conventional system presented the highest prevalence of moderate bursitis while extensive Mallorcan Black pigs showed the lowest one. Intensive Iberian pigs had an intermediate prevalence of moderate bursitis between the conventional and the straw bedded systems. The highest prevalence of severe bursitis was also recorded in the conventional system; however no significant differences were seen between pigs in straw bedding, intensive Iberian pigs and pigs in extensive conditions. Severity of bursitis is associated with a hard environment that increases pressure and impact on pigs' limbs (Lyons et al., 1995; Mouttotou et al., 1998). Severe bursitis is therefore expected to be less sensitive than moderate bursitis when comparing production systems with smooth flooring.

The differences between production systems found in the present study were in accordance with previous studies on the conventional system, straw bedded accommodation and outdoor housing (Guy et al., 2002; Gillman et al., 2008). Still, the variability within intensive systems was high even in the straw bedded system and should be explained by several causal factors. Among intensive Iberian pigs, fully slatted floor was clearly a causal factor for moderate bursitis; however no significant differences were observed between partly slatted floor, concrete floor with resting area and outdoor access, and outdoor paddocks with deep sand bedding. Solid concrete floor without bedding was identified as a risk factor of bursitis in previous studies (Gillman et al., 2008). Also Mouttotou et al. (1999), showed that bedding that covered the lying area was the most important factor that reduced the risk of bursitis. Therefore, a significantly higher risk of bursitis would have been expected in partly slatted floors compared to soil or bedded floors. Moderate bursitis and manure on the body were negatively correlated among intensive Iberian pigs. Detection of moderate bursitis may have been underestimated in floors where pigs were found dirtier. Contrary to intensive Iberian pigs and to the results found by Gillman et al. (2008), in the conventional system, fully slatted floor was not identified as a significant causal factor of bursitis compared to partly slatted floor. As commented by Mouttotou et al. (1999), it is possible that the floor quality account for such differences instead of the type of floor per se. The growing stage was the predominant confounding factor of moderate and severe bursitis in the conventional system. As commented by Guy et al. (2002) and Gillman et al. (2008) older animals housed on concrete floors presented a higher risk of bursitis as greater body weights exert additional pressure of the concrete on the limbs. Opposite results were observed in intensive Iberian pigs, where the risk of moderate bursitis decreased in pigs that finalized their growing period (average 267 days old). Iberian pigs are housed in different types of floor through their growing period going from one type of paddock to the other and in the present study older pigs were more likely to be kept on smooth surfaces. It appears, then, that when pigs are housed until high body weight on smooth surfaces the risk of bursitis is more likely to decrease. Smith (1993) also noted that as the amount of time spent on bedded floors increased, the prevalence of bursitis decreased. In the conventional system, pigs housed at the highest stocking densities presented the lowest prevalence of moderate bursitis. On the other hand, stocking density had no effect on the apparition of severe bursitis among conventional pigs and was no detected as a causal factor in intensive Iberian pigs and pigs in the straw bedded system. Smith (1993) reported that pigs kept at high stocking densities tended to have an increased prevalence and severity of bursa lesions whereas Gillman et al. (2008) did not find any effect of the density on bursitis. The effect of the density on the apparition of bursitis is unclear. Even though the relation between the density and the age of the pigs was weak in the conventional system, the age might have a confounding effect on the stocking density.

Pig dirtiness

Faeces can be a source of infectious agents; therefore, reducing the contact between pigs and their excrements is important in terms of prevention of diseases. If possible, pigs excrete in a localized part of their pen and prefer to separate their dunging area from their lying area (Meunier-Salaün, 1989; Hacker et al., 1994). If pigs are dirty with faeces, it is because some factors have temporarily forced them to lie in their excrement. Both types of manure on the body were correlated moderately; however this correlation was not strong enough to decide to analyse them as a sole measure. Although no differences were seen in the prevalence of manure on less than 50% of the body (moderate soiled body) between intensive production systems, pigs in straw bedding presented a significantly higher risk of manure on more than half of the body (severe soiled body) than pigs in the conventional system. Scott et al. (2006) also reported a lower hygiene in pigs in the straw bedded system compared to pigs in conventional housing. Excretions are absorbed by the bedding material and require much more cleanliness management than slatted floors. Moreover, as straw maintains a temperature warmer than concrete, upper critical temperature is reached before. Pigs are then more likely to dissipate heat wallowing in their own excrements. Intensive Iberian pigs showed an intermediate prevalence of severe soiled body, between the conventional and the straw bedded systems. In overall, severe soiled body appeared to be more sensitive to differences between production systems than moderate soiled body. However, when studying the sensitivity of pig dirtiness between conventional farms (Temple et al., 2011), moderate soiled body allowed a better discrimination between farms than severe soiled body.

Effective temperature results from the interaction of many factors such as the type of floor, the relative humidity, air velocity and the weight of the pigs with the room temperature (Geers et al., 1989). Those interactions make difficult the prediction of pig dirtiness, especially under commercial conditions. The fact that manure on the body presented over-dispersion and the country appeared to have a significant effect reflect this complexity. Still, several possible causal factors that would contribute to pigs lying in their dunging area could be identified for each intensive production system. The type of flooring was the predominant causal factor for the apparition of both scores of manure on the body in the conventional system. It was also identified as a significant causal factor of moderate soiled body among intensive Iberian pigs. In both production systems, pigs on fully slatted floor presented a decreased risk of dirtiness than pigs on partly slatted floor. High effective temperatures increase lying in the slatted area and pigs use the foul area of the solid floor as a dunging area (Courboulay et al., 2003; Huynh et al., 2005). They may use this area to cool themselves by wallowing in excrement (Huynh et al., 2005). The age of the pigs was a second possible causal variable of moderate soiled body in the conventional system as well as for both types of dirtiness in the straw bedded system. Older pigs presented higher risk of poor hygiene than younger ones. Older pigs spend more time lying and they are more likely to adopt a fully recumbent position (Ekkel et al., 2003). Moreover, several studies on pigs in conventional conditions have found that dunging on the solid floor increases when pigs are heavier (Hacker et al., 1994). Thereafter, a higher evaporative heat loss in heavy pigs in addition to a greater space occupational area may increase the probability of an older pig to lie on a dunging area. Severe soiled body was not influenced significantly by the growing stage in conventional pigs; instead, the feeding system appeared to be a significant causal factor of severe soiled body. In the conventional system, liquid fed pigs in trough presented a higher prevalence of severe soiled body than pigs fed by wet feed hopper, being in accordance with the results of Scott et al. (2007). The increase in water excretion originated by the ingestion of large volumes of water can make worse the accumulation of excretion especially on the solid zones of the concrete.

Huddling, shivering and panting

Huddling, shivering and panting are animal-based indicators of thermal comfort. Shivering and panting showed low prevalences (<1%) in all production systems and none huddling animals were detected among intensive Iberian pigs and among pigs in extensive conditions. Moreover, pigs in the conventional and straw bedded systems did not display a significant difference in the prevalence of huddling behaviour. On the days of assessment, pigs were not subjected to extremely cold environments as temperatures ranged from 12ºC to 24ºC in intensive Iberian pigs, from 15ºC to 40ºC in the straw bedded system and from 16ºC to 29ºC in the conventional system. As well, observations in extensive conditions were not done during winter months. Estimations of the thermal comfort zone range from 15 to 24ºC and from 14 to 21ºC in growing and finishing pigs, respectively (Gonyou et al., 2006). Thereafter, the ranges of temperatures recorded in the present study appeared to be insufficiently low to observe a high variability of huddling behaviours and possible differences between production systems.

In the conventional system, the prevalence of huddling animals was highly dependent of the country and, as it happens for pig dirtiness, an important variability in the prevalence of this measure could not be explained by the model. It is possible that the type of floor account for differences between countries; however it is more likely that a combination of other factors such as the air velocity, the floor temperature and the variation in environmental conditions (Geers et al., 1986) originate such differences. The degree of huddling occurring in a group was also affected by the stocking density that represents an important confounding factor. Still, the risk of huddling in the present study decreased at high room temperatures (from 25 to 29º).

CONCLUSIONS

Important differences between production systems can be established based on animal-based measures of the good feeding and housing principles of Welfare Quality® protocols. Pigs in the conventional system presented the lowest prevalence of poor body condition while extensive Mallorcan Black pigs and extensive Iberian pigs were associated with a decreased prevalence of bursitis and pig dirtiness. The straw bedded system presented a lower prevalence of bursitis but poorer hygiene and more susceptibility of poor body condition than the conventional system. Several possible causal factors of animal-based measures of good housing were identified and their respective impact should be interpreted in based of the production system studied. Some prevalences were largely affected by the age of the animals being a confounding variable that has to be taken into account when assessing a farm. The recording of resource-based measurements could be useful to identify farms that are more likely to show a welfare deficiency for a given measure and to interpret levels of prevalences for advisement purposes.

ACKNOWLEDGEMENTS

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. The authors are grateful to the farmers, veterinarians and pig associations for their cordial collaboration. Déborah Temple was supported by an FPU scholarship from the Ministerio de Educación of Spain.

REFERENCES

Altman DG, Lausen B, Sauerbrei W and Schumacher M 1994. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. Journal of the National Cancer Institute 86, 829 – 835.

Aparicio Tovar MA and Vargas Giraldo JD 2006. Considerations on ethics and animal welfare in extensive pig production: Breeding and fattening Iberian pigs. Livestock Science 103, 237–242.

Blokhuis HJ 2008. International cooperation in animal welfare: the Welfare Quality® project. Acta Veterinaria Scandinavica 50, 1-5.

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007. Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, 225-228.

Capdeville J and Veissier I 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Veterinaria Scandinavica 51, 62-68.

Courboulay V, Bregeon A, Massabie P and Meunier-Salaün MC 2003. Incidence du type de sol (caillebotis partiel/caillebotis intégral) et de la taille de la case sur le bien-être des porcs charcutiers. Journées de la Recherche Porcine 35,163-170.

Courboulay V, Eugène A and Delarue E 2009. Welfare assessment in 82 pig farms: effect of animal age and floor type on behaviour and injuries in fattening pigs. Animal Welfare 18, 515-521.

Dohoo I, Martin W and Stryhn H 2009. Model-buiding strategies. In Veterinary Epidemiological Reseach (eds I Dohoo, W Martin and H Stryhn), pp 368–374. AVC Inc: Charlottetown, Price Edward Island, Canada.

Ekkel ED, Spoolder HAM, Hulsegge I and Hopster H 2003. Lying characteristics as determinants for space requirements in pigs. Applied Animal Behaviour Science 80, 19–30.

Geers R, Goedseels V, Parduyns G and Vercruysse G 1986. The group postural behaviour of growing pigs in relation to air velocity, air and floor temperature. Applied Animal Behaviour Science 16, 353-362.

Geers R, Dellaert B, Goedseels V, Hoogerbrugge A, Vranken E, Maes F and Berckmans D 1989. An assessment of optimal air temperatures in pig houses by the quantification of behavioural and health-related problems. Animal Production 48, 571-578.

Gillman, CE, KilBride AL, Ossent P and Green LE 2008. A cross-sectional study of the prevalence and associated risk factors for bursitis in weaner, grower and finisher pigs from 93 commercial farms in England. Preventive Veterinary Medicine 83, 308-322.

Gonyou HW, Lemay SP and Zhang Y 2006. Effects of the environment on productivity and disease. In Diseases of swine (eds B Straw, JJ Zimmerman, S D'Allaire, D Taylor), pp 1027-1038. Blackwell Publishing, Ames, IA, USA.

Guy JH, Rowlinson P, Chadwick JP and Ellis M 2002. Health conditions of two genotypes of growing-finishing pig in three different housing systems: implications for welfare. Livestock Production Science 75, 233–243.

Hacker RR, Ogilvie JR, Morrison WD and Kains F 1994. Factors affecting excretory behaviour of pigs. Journal of Animal Science 72, 1455–1460.

Huynh TTT, Aarnink AJA, Gerrits WJJ, Heetkamp MJH, Canh TT and Spoolder HAM 2005. Thermal behaviour of growing pigs in response to high temperature and humidity. Applied Animal Behaviour Science 91, 1-16.

Jaume J and Alfonso L 2000. The Mallorcan Black pig. Agri 2000 27, 53-58.

Johnsen PF, Johannesson T and Sandøe P 2001. Assessment of farm animal welfare at herd level: Many goals, many methods. Acta Veterinaria Scandinavica 51, 26-33.

Lee Y, Nelder JA and Pawitan Y 2006. Generalized linear models. In Generalized Linear Models with Random effects (eds Y Lee, JA Nelder and Y Pawitan), pp 54–64. Chapman and Hall/ CRC, Boca Raton, FL, USA.

Lyons CAP, Bruce JM, Fowler VR and English PR 1995. A comparison of productivity and welfare of growing pigs in four intensive systems. Livestock Production Science 43, 265–274.

Meunier-Salaün MC 1989. Relations comportementales du porc avec son environnement: critères d'évaluation des systèmes d'élevage. Journées de la Recherche Porcine 21, 281-296.

Mouttotou N, Hatchell FM and Green LE 1998. Adventitious bursitis of the hock in finishing pigs: prevalence, distribution and association with floor type and foot lesions. Veterinary Records 142, 109–114.

Mouttotou N, Hatchell FM and Green LE 1999. Prevalence and risk factors associated with adventitious bursitis in live growing and finishing pigs in south-west England. Preventive Veterinary Medecine 39, 39–52.

Rodríguez-Estévez V, García A, Peña F and Gómez AG 2009. Foraging of Iberian fattening pigs grazing natural pasture in the dehesa. Livestock Science 120, 135-143.

Scott K, Chennells DJ, Campbell FM, Hunt B, Armstrong D, Taylor L, Gill BP and Edwards SA 2006. The welfare of finishing pigs in two contrasting housing systems: Fully-slatted versus straw-bedded accommodation. Livestock Science 103, 104-115.

Scott K, Chennells DJ, Armstrong D, Taylor L, Gill BP and Edwards SA 2007. The welfare of finishing pigs under different housing and feeding systems: liquid versus dry feeding in fully-slatted and straw-based housing. Animal Welfare 16, 53-62.

Segalés J, Allan GM and Domingo M 2006. Porcine Circovirus Diseases. In Diseases of swine (eds B Straw, JJ Zimmerman, S D'Allaire and D Taylor), pp 299-307. Blackwell Publishing, Iowa.

Smith WJ 1993. A study of adventitious bursitis of the hock. Thesis, University of Edinburgh, UK.

Smulders FJM 2009. A practicable approach to assessing risks for animal welfare – methodological considerations. In Food safety assurance and veterinary public health. Vol. 5. Welfare of production animals: assessment and management of risks (eds FJM Smulders and B Algers), pp 239-274. Wageningen Academic Publishers, The Netherlands.

Temple D, Dalmau A, Ruiz de la Torre JL, Manteca X and Velarde A 2011. Application of the Welfare Quality® protocol to assess growing pigs kept under intensive conditions in Spain. Journal of Veterinary Behavior 6, 138–149.

Welfare Quality® 2009. Welfare Quality® applied to growing and finishing pigs. In Welfare Quality® Assessment protocol for pigs (eds A Dalmau, A Velarde, K Scott, S Edwards, I Veissier, L Keeling, A Butterworth), pp 49-78. Welfare Quality® Consortium, The Netherlands.

Whay HR, Main DCJ, Greent LE and Webster AJF 2003. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. Animal Welfare 12, 205- 217.

Wolfinger R and O'Connell M 1993. Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. Journal of Statistical Computation and Simulation 4, 233–243.

CHAPTER 3

THE WELFARE OF GROWING PIGS IN FIVE PRODUCTION SYSTEMS: ASSESSMENT OF HEALTH

Based on Temple D, Courboulay V, Velarde A, Dalmau A, Manteca X 2012. The welfare of growing pigs in five different production systems in France and Spain: assessment of health. *Animal Welfare 21:* 257-271

ABSTRACT

This study was carried out to compare the health of growing pigs in five different production systems in France and Spain using measures provided by the Welfare Quality® protocol. A total of 11 647 pigs housed on 91 commercial farms were evaluated over a two year period (2007 – 2009). Farms considered as conventional were close to the European dominant production system, rearing "white" pigs (e.g. Large White - Landrace x Pietrain) housed on concrete floors. Systems considered as differentiated had specifications to distinguish them from the conventional one. Farms that housed "white" breeds of pigs on straw were then considered as a different production system. Mallorcan Black pigs managed extensively on family farms at the Balearic Islands represented a third production system. The remaining two systems assessed were represented by the methods used for Iberian pig rearing extensively or intensively. Multiple Generalized Linear Mixed Models were performed for each animal-based measure of health. The straw bedded and the conventional systems did not differ in the prevalence of any animal-based measures. Mallorcan Black pigs and Iberian pigs kept extensively had a lower prevalence of severe wounds than pigs in the conventional system and the lowest prevalence of tail biting. Focusing on pigs housed in the conventional system, several possible causal factors (such as the feeding system and the type of floor) were identified relating to severely wounds, tail biting and lameness. Therefore, the recording of simple environmental-based factors can be useful in detecting farms that are more likely to show these problems.

Keywords: animal welfare, growing pig, health, housing systems, Welfare Quality®, causal factors

Running title: Assessment of health on growing pig farms

INTRODUCTION

Health refers to the state of the body and brain in relation to the effects of pathogens, parasites, tissue damage or physiological disorder. Since all of these effects involve pathology (that is the detrimental derangement of molecules, cells and functions that occurs in living organisms in response to injurious agents or deprivations), the health of an animal is its state as regards its attempts to cope with pathology (Broom, 2006). Health is, therefore, a significant component of welfare (Broom, 2010) and must be properly considered in a welfare assessment. The Farm Animal Welfare Council (1992) provides a narrower definition of health as the absence of pain, injuries or diseases. Injuries and diseases can cause acute or chronic pain which, in turn, is defined as an aversive emotional experience (Molony and Kent, 1997; Rainville, 2002). It should be emphasised that a light difficulty in coping with pathology may not be detected using this previous definition.

Animal health depends on several influencing factors and may vary according to the production system. For growing pigs, farms with concrete floors and relatively high stocking densities greatly predominate throughout Europe and can be considered as conventional. Recently, consumer concerns about animal welfare have led to a growing interest for alternative production systems. Several studies have been conducted to compare the prevalence of health indicators - such as skin lesions, tail biting and lameness – in conventional system and straw bedded accommodation (Lyons et al., 1995; Guy et al., 2002; Scott et al., 2006; Courboulay et al., 2009). At present, around 7% of growing pigs in France are housed on straw bedding while very few Spanish farms use deep litter systems because of its *a priori* incompatibility with high environmental temperatures and limited availability of the material. Pigs housed on conventional and straw bedded farms are usually "white" breeds of pigs selected for their high growth speed or high conversion index and their adaptability for indoor husbandry. At the same time, increasing attention is being given to outdoor production systems, which are usually found in specific geographical areas. Mallorcan Black pigs and Iberian pigs represent approximately 15% of the overall population of growing pigs in Spain. Both autochthonous breeds are rustic animals derived from the Mediterranean line characterised by a dark-grey skin colour. Mallorcan Black pigs are grown in Balearic Islands while Iberian pigs are traditionally found in the southwest of the Iberian Peninsula. Both breeds are traditionally reared under extensive conditions in specific ecosystems, taking advantage of the existing natural resources. There are differences in the way these animals are housed, fed and managed as well as in the nature of the final product (Jaume and Alfonso, 2000; Rodríguez-Estévez et al., 2009). Mallorcan Black pigs are produced on family farms with old social structures where the pig enterprise is never the main activity of the farm. Preservation of Mallorcan Black pig is mainly sustained by the utilization of pure breed animals in the elaboration of the "Sobrassada de Mallorca de Porc Negre Mallorqui", a kind of cured sausage with paprika qualified as a Protected

Geographical Indication. This breed is always reared extensively and the feeding regime is based on pasture, cereals (barley and rye), legume seeds, almonds and several Mediterranean shrubs. The most common practice with Mallorcan Black Pig is slaughtering animals at heavy live weight, around 150 kg. The traditional rearing system of the Iberian pig is linked to the “dehesa” (*Quercus ilex* and *Quercus suber*). The abundance of food provided by acorn ripening is used by the Iberian pigs during the late fattening phase, which is called “montanera” and takes place from early November to late February, when the diet is based only on natural resources. This system has its own legal regulation (the Quality Standards for Iberian Pork and Cured Products), which specifies a stocking rate of less than 2 pigs per hectare and that animals must be slaughtered above 14 months of age (approximately 150 kg live weight) (MAPA, 2007). The growing demand from the consumers for Iberian pig meat products such as the Iberian dry-cured hams and shoulders has recently led to an intensification of the production cycle. Indeed, taking advantage of this market niche, some “white” pig producers started to rear Iberian pure breed or Iberian x Duroc crossbred with the aim of producing high quality meat products. Intensification of the Iberian pig system has led to a reduction in length of the production cycle (slaughtering at 10 months of age, approximately) and the systematic use of fodder resulting in the total independence of seasonality (“montanera period”). Iberian pigs in intensive conditions can, thus, be reared elsewhere in Spain. Intensive Iberian pigs are housed in a wide variety of accommodations ranging from indoor slatted pens to outdoor paddocks. Such intensification is generally associated with a decrease in animal welfare; however much knowledge still remains to be transferred (Aparicio Tovar and Vargas Giraldo, 2006).

Despite a general interest in comparing production systems, few studies have been undertaken on commercial farms across countries. One of the main reasons is the lack of harmonized information that enables reliable comparison of data from different sources. Welfare Quality® protocols provide the tool that makes possible the assessment of farms through Europe in a standardized way by observers that receive an identical training. Moreover, this assessment system should be applicable and valid in a wide variety of farming systems (Blokhuis,2008). Welfare Quality® is based on four main principles of animal welfare: i) Good feeding, ii) Good housing, iii) Good health and iv) Appropriate behaviour. Each of these four principles comprises several independent but complementary criteria (Botreau et al., 2007) in turn characterized by one or various measures. The principle labelled as “Good Health” includes three different criteria: absence of injuries (such as wounds on the body, lameness, tail bitting, etc.), absence of disease (such as respiratory problems, enteric disorders, skin condition, etc.) and absence of pain induced by management procedures (such as tail docking or castration practices). Preference was given to animal-based measures for being more valid when measuring the state of the animal as regards to its attempts to cope with its environment (Capdeville and Veissier, 2001; Whay et al., 2003). Animal-based measures give an indication of the performance of a husbandry system at a certain moment of the production cycle. A Welfare Quality® protocol has not been thought as a risk assessment tool. Instead it aims to detect farms or housing and management systems with particular welfare problems. Welfare concerns may indeed have more severe consequences in one production system than in another. It is essential though to identify which factors are the main sources of variation within each system. Knowledge about the possible causes of impaired welfare is essential to drive improvement in animal welfare (Whay, 2007) and to prevent misinterpretations (Temple et al., 2011a). For that reasons and to ensure proper feed-back to farmers (Blokhuis, 2008), simple design measures were included into the Welfare Quality® protocol.

To sum up, the aim of this study was to compare the health of growing pigs reared in five different production systems using the Welfare Quality® assessment protocol. It provides data about the prevalence and distribution of several welfare outcomes on a wide range of commercial farms with a view to possible causal factors. The overall Welfare Quality® protocol was applied on 91 commercial farms in France or in Spain and this study presents a benchmark description of several health measures.

MATERIAL AND METHODS

Farm selection and production systems

A total of 91 commercial farms of growing pigs (from 28 days old) were assessed applying the Welfare Quality® protocol over a two year period. Sixty one farms were evaluated in Spain and 30 in France. Choice of farms was based on management practices, farm size and veterinary records in order to obtain a large variety of situations. Farm sampling concerned the five main production systems of growing pigs found in France or in Spain.

Farms considered as conventional were close to the Europe-wide dominant production system aiming at the lowest possible production costs. Fifty two Spanish and French farms with concrete floor were categorized as conventional farms. Systems considered as differentiated had specifications to distinguish them from the conventional one with one or several claims among the following: animal welfare, eating quality, nutritional quality, environment, organic production, local production. Four other production systems were differentiated in the present study: the straw bedded system (8 farms assessed in France); the intensive Iberian pig system (10 farms assessed in Spain); the extensive Iberian pig system (11 farms assessed in Spain) and the extensive Mallorcan Black pig system (10 farms assessed in Spain).

Sampling of pigs and pens

On each intensive farm, 10 pens of growing-finishing pigs from different ages were randomly selected. Hospital pens were not sampled; however, their presence/absence was recorded. If there were 15 or fewer pigs in a pen/group, all pigs were assessed, whereas if there were more than 15 pigs in the pen/group, 15 were arbitrarily selected. Pigs were categorized into three growing stages:

- Early growing stage: at the beginning of the growing process (less than 90 days old).
- Mid growing stage: mid growing period from 90 to 130 days old.
- Final growing stage: at the end of the growing process and before leaving to the slaughterhouse (from 130 to 194 days old).

Among intensive Iberian pigs as well as extensive pigs, a fourth category included pigs older than 194 days old. Information on whether pigs were castrated or not were also taken into account. Pigs were assessed at least ten days after arrival at the farm to avoid sensitive period around mixing.

Data regarding the floor type (fully slatted floor, solid concrete floor, straw, or sand) and the feeding system (trough, dry hopper, wet feed hopper, Turbomat, on the ground, on pasture) were collected. Space allowance was calculated and the average environmental temperature was noted down.

Pigs and pens characteristics

A total of 11 647 pigs from 719 pens or paddocks were evaluated.

In the conventional system, 7 030 pigs of an average of 109 days old (SD = 34) ranging from 29 to 180 days old were assessed in 530 pens. Pigs were housed in groups of an average of 16 animals (SD = 7) ranging from 6 to 67 animals per pen. The majority of growing pigs on conventional farms come from a maternal line Large White x Landrace. The paternal line mainly used was the Pietrain . However, in some farms the Duroc breed was used as a paternal line. On any given farm, growing pigs could have different origins and genetics. Overall, 60% of the farms housed castrated males. Pigs were kept on concrete floor whether fully or partly slatted and the average space allowance was 0.66 m^2 per pig (SD = 0.20). The majority of pigs were fed *ad libitum* by means of dry hopper or wet feed hopper. Some animals were liquid fed in trough from two to four times a day or by Turbomat once or twice a day. In these two feeding systems, food was restricted in time but not in quantity. Turbomat (manufactured by Roxell® in Belgium) is a circular trough with four or six drinking nipples above it where feed is available for ten pigs simultaneously. **Table 1** provides more descriptive data on pens characteristics. The mean environmental temperature was 23.4 ºC (SD = 2.3). Hospital pens were present on 70% of the conventional farms. French and Spanish conventional systems were considered as the same production system; however some differences in the distribution of some explanatory variables should be considered. More than 90% of concrete floors on French farms were fully slatted; liquid feeding system in trough and Turbomat system were only seen in France; all the temperatures were recorded above 21ºC in France and all males were castrated in France while 30% of the Spanish farms housed castrated males.

In the straw bedded system, 1 110 pigs of an average of 112 days old (SD = 42) ranging from 41 days old to 194 days old were assessed in 74 pens. Pigs were housed in groups of an average of 38 animals (SD = 15) ranging from 18 to 83 animals per pen. As it happens on conventional farms, pigs on straw bedded systems were "white" pigs from different genetic lines. On the same farm growing pigs could have different origins and genetics. The average space allowance was 1.52 m^2 per pig (SD = 0.96). All the males were castrated. Pigs were fed *ad libitum* by means of dry hopper or wet feed hopper or liquid fed in trough from two to four times a day (**Table 1**). The mean environmental temperature was 23.4 ºC (SD = 4.5). Hospital pens were present on 43% of the farms.

A total of 1 255 Iberian pigs kept in intensive conditions were evaluated in 58 pens or paddocks. Pig age ranged from 37 days old to 280 days old (mean = 250; SD = 97). Pigs were housed in groups of an average of 115 animals (SD = 91) ranging from 7 to 320 animals per pen. On a same farm pigs could be from the Iberian pure breed or Iberian x Duroc crossbred. All the males were castrated. Pigs were housed on a wide range of floor types (Table 1). The average space allowance was 1.9 m^2 per pig ranging from 0.30 to 5.4 m^2 per pig. All the pigs were dry fed *ad libitum* by dry hopper or trough (**Table 1**). The mean environmental temperature was 20.5 ºC (SD = 3.3). Hospital pens were present on 90% of the farms.

Under extensive conditions, 1 428 Iberian pigs of an average 250 days old (SD = 97.3) ranging from 42 to 420 days old were assessed in 31 different paddocks. Pigs were kept in groups of an average of 170 animals (SD = 116) ranging from 12 to 470 animals per paddock. All animals assessed were pure Iberian pigs and all the males were castrated. The average space allowance per pig was 430 m^2 (SD = 578) ranging from 83 to 2 500 m^2 per pig. On the months of assessment pigs were supplemented with fodder as the availability of natural resources was insufficient. Supplementation was given manually once or twice a day on the floor (55% of the paddocks), in a trough (26% of the paddocks) or in dry hoppers (19% of the paddocks).

Mallorcan Black pig reared extensively was the second extensive system assessed. A total of 824 Mallorcan Black pigs of an average 220 days old (SD = 100) ranging from 75 days old to 405 days old were evaluated in 24 paddocks. Pigs were kept in groups of an average of 58 animals (SD = 37) ranging from 11 to 170 animals per paddock. All animals assessed were Mallorcan Black pigs and all the males were castrated. The average space allowance per pig was 692 m^2 (SD = 818) ranging from 130 to 4 000 m^2 per pig. All pigs were on pasture and were supplemented with household refusals (such as watermelon/melon peels, citric peels, vegetables, etc.), bran, legume seeds, cereals and figs. Supplementation was done manually, once or twice a day, on the floor (23% of the paddocks) or in a trough (35% of the paddocks) and given *ad libitum* by dry hopper (37% of the paddocks). In the remaining 5% of paddocks, pigs were entirely reliant on pasture. During the days of assessment, the average environment temperature was 18ºC and 25ºC in extensive Iberian and Mallorcan Black pigs, respectively. Sick animals were usually kept in small provisional areas of a paddock, however the presence/absence of hospital pens as such could not be properly recorded in extensive conditions.

Table 1 Description of variables collected on the growing pig pens of the three intensive production systems studied

Variable	Conventional n = 530 pens	Straw bedded n = 74 pens	Intensive Iberian n = 58 pens
Space allowance (m^2/pig)			
Mean ± SD	0.66 ± 0.20	1.52 ± 0.96	1.86 ± 1.32
Minimum	0.2	0.3	0.3
Q1	0.58	0.7	0.8
Median	0.68	1.2	1.4
Q3	0.75	2.9	2.6
Maximum	1.56	3.0	5.4
Temperature (ºC)			
Mean ± SD	23.4 ± 2.3	23.2 ± 4.5	20.5 ± 3.3
Minimum	16.0	15.0	12.0
Q1	21.0	20.0	19.0
Median	23.0	23.0	22.0
Q3	25.0	24.0	23.0
Maximum	29.0	40.0	24.0
Feeder type (% of pens)			
Trough	23%	19%	11%
Dry hopper	38%	59%	89%
Wet feed hopper	37%	22%	
Turbomat	2%		
Floor type (% of pens)			
Partly slatted	40%		24%
Fully slatted	60%		27%
Concrete with resting area			37%
Outdoor on deep bedding			12%
Hospital pen (% of farms)			
Presence	70%	43%	90%
Absence	30%	57%	10%

Upper limit of the first quartile (Q1) and third quartile (Q1).

Measurements

Assessments were performed by four observers, two of them in Spain and the other two in France. In order to minimize difference between observers and to standardize the scores from the visits, observers followed the same training prior beginning the assessment and repeatability among them was assessed. The training consisted in a set of 40 – 60 video clips and images for each measure included into the protocol as well as training on a commercial farm. If the observer did not reach an acceptable agreement with the golden standard (in the case of images or videos) or with the silver standard (in the case of live animals), a set of extra images and animals were assessed until achieve a good repeatability.

Severely wounded animals

Pigs were encouraged to stand up to make the body clearly visible. One side of the pig's body was inspected visually for the presence of lesions and/or penetration of the muscle tissue, considering five separate regions: i. Ears, ii. Front (head to back of shoulder), iii. Middle (back of shoulder to hind-quarters), iv. Hind-quarters, v. Legs (from the accessory digit upwards). The tail zone was not considered here. A scratch longer than 2 cm as well as a round lesion smaller than 2 cm was given a lesion score of 1. A round lesion from 2 to 5 cm of diameter or more than 5 cm and healed was given a lesion score of 5. A round lesion of more than 5 cm, deep and opened was given a lesion score of 16. Each zone was considered separately. Animals were considered affected (severely wounded animals) when presenting more than 10 lesions on at least two zones of the body or any zone with more than 15 lesions.

Tail biting

Severe tail biting was considered when fresh blood was visible on the tail; when there was evidence of some swelling and infection or when part of the tail tissue was missing and a crust had formed.

Lameness

Animals were observed individually during walking from inside the pen or paddock. A pig was considered lame when it presented a minimum weight-bearing on the affected limb or when it was not able to walk.

Skin condition

Each animal was visually inspected and considered using the following individual scale: 0 – No evidence of skin inflammation or discoloration; 1 (localized) – Up to 10% of the skin was inflamed, discoloured or spotted; 2 (widespread) – More than 10% of the skin had an abnormal colour or texture. The number of animals in each category was considered.

Scouring

Scouring was measured at pen level instead of individual level. The observer looked at areas in the pen/paddock where the dung was visible. When liquid and fresh manure was visible, the pen was considered as a pen with scouring. In extensive conditions, loose faecal material around the anal region helped in the detection of diarrhea. When scouring could not be evaluated properly for feasibility reasons (i.e. dirty pens), the pen was considered as a missing value.

Pumping (laboured breathing), twisted snout, rectal prolapse and hernia

Pigs with heavy and laboured respirations were defined as pumping animals. Pigs that presented a nasal distortion that was characteristic of atrophic rhinitis were considered as pigs with twisted snouts. To detect rectal prolapse, pigs were examined from behind, checking for presence of swellings and extrusion of tissue from the rectum. To detect hernias, the animals were observed from the front, back and side. Hernias with bleeding lesion or hernias that affected the behaviour of the animal were recorded.

More detailed information about the overall methodology of assessment of health measures can be found in the Welfare Quality® Assessment protocol for pigs (2009).

Statistical Analysis

The pen or paddock was the experimental unit. Data – except scouring - were expressed as the number of animals affected out of the number of animals assessed in each unit. Scouring was expressed as presence or absence of diarrhea at pen level. To account for possible dependence between observations on pens from the same farm, random farm effects were included into the model. Data were also clustered at pen level and modelled for over-dispersion. Multiple Generalized Linear Mixed Models for binomial data were performed separately for each welfare measure using the GLIMMIX procedure (SAS statistical package Version 9.1, SAS Inst.Inc., 2002, Cary, NC, USA). Residual pseudo-likelihood was used as estimation technique (Wolfinger and O'Connell, 1993).

The general form of the model was:

$$\text{Logit}(y/n) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_{farm(i)} + v_{pen(i)}$$

where y is the number of animals affected, n the number of animals assessed, β_0 the intercept, β_k the regression coefficient predictors, X_{ki} the predictive factors. u_{farm} and v_{pen} are the residual variances of the random effects.

Models were first built to compare the five production systems: conventional, straw bedded, intensive Iberian, extensive Mallorcan Black pig, extensive Iberian. Then, separate models were developed for each intensive system (conventional, straw bedded, Iberian intensive) to identify possible causal factors such as the type of floor, the feeding system, the growing stage, space allowance and ambient temperature. At the expense of throwing away information, these two last continuous variables (space allowance and temperature) were converted into categorical variables by means of quartiles to avoid strong assumptions about the linear relation between the outcome measure and the predictive factor (Altman et al., 1994). Possible causal variables were entered into the model as fixed effects.

Statistical analyses consisted of initial univariable screening to determine the variables statistically associated with each indicator of health. Variables were taken forward for multivariable analysis when significant at $P < 0.2$ (Dohoo et al., 2009). Where variables were highly correlated (PROC SPEARMAN) or showed a strong association (PROC FREQ with the CHISQ option), the most biologically plausible variable was selected for inclusion in the final model. Stepwise backward selection was performed to identify the variables that had a significant association ($P < 0.05$) with the outcome measure. The country effect (Spanish *vs.* French farms) was considered into each model of the conventional system. When the country effect vanished in the analysis while other factors were added to the model, "country" ceased to be an interest as a separate factor. The observer effect (observer one *vs.* observer two in Spain; and observer three *vs.* observer four in France) was also considered into each final model and taken out when it did not alter the interpretation of fixed effect. The ratio of the generalized chi-square statistic and its degrees of freedom was considered to check how well the variability in the data has been properly modelled and that there was no residual over-dispersion. Goodness of fit was also assessed visually by means of standardized Pearson residuals (Lee et al., 2006). When data did not allow a proper modelling, identification of predictive factors was not undergone.

Spearman correlation coefficients were calculated to investigate the association between the different animal-based measures of health.

RESULTS

Table 2 Mean prevalence (% of pigs affected at pen level) and standard error of animal-based measures of health in five production systems. The number of pigs and pens/paddocks (n) assessed is shown for each production system

Animal-based measure	Conventional 7 030 pigs n = 530 pens	Straw bedded 1 110 pigs n = 74 pens	Intensive Iberian 1 255 pigs n = 58 pens	Extensive Mallorcan Black pig 824 pigs n = 24 paddocks	Extensive Iberian 1 428 pigs n = 31 paddocks
Severely wounded	2.5 ±0.35[a]	1.4 ±0.40[ab]	0.1 ±0.10[b]	0.2 ±0.13[b]	0.4 ±0.29[b]
Tail biting	1.1 ±0.27	1.4 ±1.26	0.2 ±0.30	0.0 ±0.00	0.0 ±0.00
Lameness	1.2 ±0.17	2.3 ±0.31	1.0 ±0.31	0.1 ±0.12	0.4 ±0.21
Localized skin condition	1.6 ±0.24	0.5 ±0.19	0.4 ±0.17	0.2 ±0.11	1.3 ±0.53
Widespread skin condition	2.1 ±0.28	0.1 ±0.09	0.3 ±0.17	1.0 ±0.49	0.4 ±0.33
Scouring *	9.0 ±0.2[b]	15.9 ±0.4[ab]	53.9 ±0.8[a]	23.1 ±0.8[ab]	23.0 ±0.8[ab]
Pumping	0.2 ±0.06	0.2 ±0.13	0.2 ±0.17	0.0 ±0.00	0.0 ±0.03
Twisted snout	0.0 ±0.00	0.0 ±0.0	0.0 ±0.00	0.0 ±0.00	0.0 ±0.00
Rectal prolapse	0.0 ±0.00	0.0 ±0.0	0.0 ±0.00	0.0 ±0.00	0.3 ±0.27
Hernia	0.1 ±0.04	0.2 ±0.13	0.1 ±0.00	0.0 ±0.00	0.0 ±0.00

Different letters within row means significant differences between systems at $P < 0.05$

*Expressed in percentage of pens affected in relation to the number of pens assessed

Severely wounded animals

The prevalence of pigs with severe wounds on the body (**Table 2**) was significantly higher in the conventional system compared with intensive Iberian pigs, extensive Iberian pigs, and extensive Mallorcan Black pigs. There was no significance difference in the prevalence of severely wounded animals between the conventional and the straw bedded systems. The odd of being severely wounded was estimated as being 4.1 times greater in pigs housed in the conventional system than in extensive Iberian pigs (**Table 3**).

Table 3 Logistic-regression model of severely wounded animals and tail biting in growing-finishing pigs for production system. The last two rows display the estimates of the covariance parameters.

	Severely wounded [a]		Tail biting [b]	
System	OR	CI	OR	CI
Extensive Iberian	1		-	-
Conventional	4.1	1.3, 13.8*	1	
Straw bedded	2.9	0.7, 13.1	0.1	0.01, 1.8
Intensive Iberian	0.6	0.1, 3.8	0.9	0.2, 4.4
Extensive Mallorcan Black pig	0.6	0.1, 4.4	-	-
Random effect	*Estimate*	*SE*	*Estimate*	*SE*
Farm	1.3	0.4	2.2	0.6
Pen	1.2	0.1	1.3	0.1

[a] Intercept coefficient= - 5.6
[b] Intercept coefficient= - 5.3

Odds ratios (OR), 95% confidence intervals (CI) and standard errors (SE)

* The confidence interval (CI) does not include unity therefore the factor is significantly different from the reference category ($P < 0.05$)

The prevalence of animals with severe wounds on the body in the conventional system varied by growing stage, feeding system, temperature levels and castration state as detailed in Table 4. In this system, the risk of severe wounds on the body (**Table 5**) was higher in pigs early in the growing period (average 68 days old) compared with older pigs. The risk of wounds on the body also increased when pigs were fed by means of Turbomat compared with the other feeding systems and in liquid fed pigs compared with dry and wet feed hoppers. Finally, there was a trend for increased wounds on the body in pens with castrated pigs; however, the confidence interval (CI) included unity ($P > 0.05$).

In straw bedding, the highest prevalence of wounds (3.4%) was found in pigs within the lowest stocking densities (First quartile: 0.3 – 0.7 m^2 per pig). Model fitness did not allow an appropriate analysis for pigs with severe wounds on the body in the straw bedded system.

Table 4 Number (No.) and prevalence (%) of growing-finishing pigs in the conventional system with severe wounds on the body and tail biting by variable selected for multivariate analysis.

	Conventional (n = 530 pens)			
	Pigs affected with severe wounds		Pigs affected with tail biting	
Variable	No.	%	No.	%
Growing stage (days old)				
Early: < 90 days old	75	4	4	0.2
Mid: 90 – 130 days old	62	2.6	49	1.7
Final: 131 – 180 days old	37	1.8	22	1.1
Feeding system				
Liquid feed in trough	60	4.1	33	2.2
Dry hopper	78	3.4	19	0.7
Wet feed hopper	19	0.8	22	0.9
Turbomat	17	9.4	1	0.6
Temperature (ºC)				
Q1: 16 – 20.9	13	1	18	1.2
Q2: 21 – 22.9	45	2.8	24	1.4
Q3: 23 – 24.9	30	2.7	27	2.4
Q4: 25 - 29	87	4	5	0.2
Castration				
No	13	0.5		
Yes	162	4.1		

Table 5 Multivariate logistic binomial mixed models of possible causal factors associated with severely wounded animals and tail biting in pigs in the conventional system. The last two rows display the estimates of the covariance parameters.

	Conventional			
	Severely wounded [a]		Tail biting [b]	
Variable	OR	CI	OR	CI
Growing stage (days old)				
Early: < 90 days old	1		1	
Mid: 90 – 130 days old	0.4	0.2, 0.7*	5.0	1.4, 17.5*
Final: 131 – 180 days old	0.3	0.2, 0.5*	3.0	0.8, 11.0
Feeding system				
Liquid feed in trough	1		1	
Dry hopper	0.5	0.3, 0.95*	0.4	0.1, 1.5
Wet feed hopper	0.2	0.1, 0.5*	0.2	0.1, 0.8*
Turbomat	4.2	1.2, 14.8*	0.3	0.1, 6.2
Temperature (ºC)				
Q1: 16 – 20.9	1		1	
Q2: 21 – 22.9	1.6	0.7, 4.0	0.4	0.2, 1.0
Q3: 23 – 24.9	0.5	0.2, 1.6	0.3	0.1, 1.2
Q4: 25 - 29	0.6	0.2, 1.7	0.1	0.02, 0.4*
Castration				
No	1			
Yes	3.0	0.9, 10.0		
Random effect	*Estimate*	*SE*	*Estimate*	*SE*
Farm	1.3	0.4	2.3	0.8
Pen	1.2	0.1	1.2	0.1

[a] Intercept coefficient= - 2.5
[b] Intercept coefficient= - 7.1
Odds ratios (OR), 95% confidence intervals (CI) and standard errors (SE)

* The confidence interval (CI) does not include unity therefore the factor is significantly different from the reference category (P < 0.05)

Tail biting

No case of tail biting was registered among extensive Iberian and extensive Mallorcan Black pigs (**Table 2**). The odds of tail biting were not significantly different between the three intensive systems studied as the confidence intervals (CI) included unity (**Table 3**). The amount of variation of tail biting accounted for by "farm" was large (estimate = 2.2 ±SE = 0.6) (**Table 3**).

The prevalence of tail biting in the conventional system varied by growing stage, feeding system and temperature levels as mentioned in **Table 4**. In this production system, the risk of tail biting (**Table 5**) was higher in pigs in the mid growing period (average 107 days old) compared with pigs early in their growing period. Moreover, there was a significant increased risk of tail biting in liquid fed pigs in trough compared with dry hopper. Finally, there was a significantly increased risk of tail biting associated with the lowest environmental temperatures (average 20ºC) compared with pigs housed within the highest environmental temperatures (average 26ºC).

Almost all the cases of tail biting in pigs in the straw bedded system were associated with pigs in the mid growing period (4.6%), housed within the highest densities (5.1%) and under the highest environmental temperatures (5.8%). Model fitness did not allow a risk factor analysis for tail biting among pigs in the straw bedded system

Lameness

There was no significant effect of the production system on the prevalence of lameness (**Table 2**). The odd of being lame was estimated as being around 3 times greater in pigs housed in the three intensive systems than in extensive Iberian pigs, however this difference was not significant as the confidence intervals (CI) included unity (**Table 6**).

Table 6 Logistic-regression model of lameness as well as localized and widespread skin condition of growing-finishing pigs for production system. The last two rows display the estimates of the covariance parameters.

	Lameness [a]		Localized skin condition [b]		Widespread skin condition [c]	
System	OR	CI	OR	CI	OR	CI
Extensive Iberian	1		1		1	
Conventional	3.0	0.7, 13.4	0.8	0.2, 4.9	3.6	0.2, 59.9
Straw bedding	3.8	0.7, 21.1	0.6	0.1, 7.6	0.5	0.1, 44.9
Intensive Iberian	2.9	0.5, 18.0	0.6	0.1, 6.9	1.6	0.1, 60.5
Extensive Mallorcan Black pig	0.8	0.1, 1.7	0.3	0.1, 5.6	3.9	0.1, 139.1
Random effect	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
Farm	1.0	0.3	4.7	1.0	2.7	0.6
Pen	0.7	0.1	0.4	0.1	0.6	0.1

[a] Intercept coefficient= - 4.7
[b] Intercept coefficient= - 5.4
[c] Intercept coefficient= - 6.0
Odds ratios (OR), 95% confidence intervals (CI) and standard errors (SE)

* The confidence interval (CI) does not include unity therefore the factor is significantly different from the reference category (P < 0.05)

In the conventional system, the prevalence of lameness varied by growing stage, stocking density, type of floor and availability of hospital pens (**Table 7**). In this system, the risk of lameness (**Table 8**) increased with the age of the pigs. There was also an increased risk of lameness associated with the most tightly stocked pigs compared with pigs housed in the lowest stocking densities. Moreover, pigs housed on fully slatted floors presented a significantly higher risk of lameness than pigs housed on partly slatted floors. Finally, there was an increased odd for lameness in pigs housed on conventional farms without hospital pens.

Among pigs in the straw bedded system and intensive Iberian pigs no causal variable of lameness could be detected.

Table 7 Number (No.) and prevalence (%) of growing-finishing pigs in the conventional system with lameness and widespread skin condition by variable selected for multivariate analysis.

	Conventional (n = 530 pens)			
	Lame pigs		Pigs affected with widespread skin condition	
Variable	No.	%	No.	%
Growing stage (days old)				
Early: < 90 days old	18	0.9	7	1.2
Mid: 90 – 130 days old	31	1.1	36	4.4
Final: 131 – 180 days old	35	1.7	91	8
Space allowance (m^2 per pig)				
Q1: 0.20 – 0.58	39	1.9	1	0.4
Q2: 0.59 – 0.68	20	1.3	26	5.6
Q3: 0.69 – 0.75	20	1.2	64	8.5
Q4: 0.76 – 1.56	7	0.5	45	4.3
Floor type				
Partly slatted	5	0.2	88	5
Fully slatted	79	1.9	48	6.2
Hospital Pen				
Presence	37	0.7		
Absence	47	2.5		

Table 8 Multivariate logistic binomial mixed models of possible causal factors associated with lameness and widespread skin condition in the conventional system. The last two rows display the estimates of the covariance parameters.

	Conventional			
	Lameness [a]		Widespread skin condition [b]	
Variable	OR	CI	OR	CI
Growing stage (days old)				
Early: < 90 days old	1		1	
Mid: 90 – 130 days old	2	1.1, 3.7*	1.5	0.5, 4.9
Final: 131 – 180 days old	2.7	1.5, 5.0*	3.9	1.2, 12.6*
Space allowance (m^2 per pig)				
Q1: 0.20 – 0.58	1		1	
Q2: 0.59 – 0.68	0.8	0.4, 1.6	8	0.9, 70.0
Q3: 0.69 – 0.75	0.6	0.3, 1.1	7.8	1.0, 62.9
Q4: 0.76 – 1.56	0.4	0.2, 0.9*	3.8	0.4, 33.2
Floor type				
Partly slatted	1		1	1
Fully slatted	6.3	2.6, 15.3*	0.6	0.3, 1.2
Hospital Pen				
Presence	1			
Absence	2.1	1.2, 3.8*		
Random effect	*Estimate*	*SE*	*Estimate*	*SE*
Farm	0.3	0.2	2.6	1.1
Pen	0.8	0.1	0.8	0.1

[a] Intercept coefficient= - 6.4
[b] Intercept coefficient= - 3.4

Odds ratios (OR), 95% confidence intervals (CI) and standard errors (SE)

* The confidence interval (CI) does not include unity therefore the factor is significantly different from the reference category (P < 0.05)

Skin condition (localized and widespread skin discolouration)

There was no significant effect of the production system on the prevalence of localized (score 1) and widespread (score 2) skin conditions (**Table 2**). According to the estimates of the covariance parameters, both types of skin condition presented over-dispersion and a large variability between farms (**Table 6**).

In the conventional system, the prevalence of widespread skin condition was highly associated with the observer. The observer number two evaluated 5.4% of pigs affected while observer one, three, and four found 0.0%, 0.1% and 0.05% of widespread skin condition, respectively. For that reason, the detection of possible causal factors for widespread skin condition was developed only for pigs in the conventional system evaluated by the observer number two. Fixing the observer, prevalences of widespread skin condition are detailed in **Table 7**. In the conventional system, and fixing the observer, there was an increased risk of widespread skin condition (**Table 8**) in pigs in the final growing stage compared with pigs early in the growing stage.

Not any causal variables were identified for localized skin condition in any of the production systems.

Scouring

There was no significant difference in the prevalence of scouring between pigs on conventional farms, pigs in the straw bedded system, extensive Iberian pigs and extensive Mallorcan Black pigs (**Table 2**). The prevalence of scouring was significantly higher in intensive Iberian pigs compared with pigs in the conventional system. For feasibility reason, 33% of the pens could not be evaluated in intensive Iberian pigs, 16% in extensive Iberian pigs, 8% in the conventional system, 7% in the straw bedded system and 0% in extensive Mallorcan Black pigs.

In the conventional system, the highest prevalence of scouring was found in the youngest pigs (15%) while the lowest one was found in finishing pigs (2%). In this system, there was a significant increased odd of scouring in pigs early in the growing period compared with finishing pigs (*OR* 20.1; *CI* 6.1, 70.7).

In the straw bedded system, the highest prevalence of scouring was also registered among youngest pigs (36%) while finishing pigs presented a lower prevalence (5%). Among pigs on straw, the odd of scouring was also significantly higher in pigs early in the growing period compared with finishing pigs (*OR* 15.3; *CI* 2.0, 121.7).

No possible causal variable was identified in intensive Iberian pigs.

Pumping (laboured breathing), twisted snout, rectal prolapse and hernia

Prevalences of pumping, twisted snout, rectal prolapse and hernia were under 1% and no significant difference were found between systems for any of these measures. The highest prevalence was seen in extensive Iberian pigs with rectal prolapse (0.3%).

Identification of possible causal variables was not possible for these measures.

Association between animal-based measures of health

In the conventional system, the strongest statistical correlation was found between both scores of skin condition ($r = 0.2$, $P < 0.001$). In the straw bedded system, localized skin condition and widespread skin condition were also correlated ($r = 0.4$; $P < 0.001$). Not any significant association between measures was found in intensive Iberian pigs.

Associations between possible causal variables, and observer effect

In the straw bedded system, the growing stage and the space allowance were highly correlated ($r = 0.7$, $P < 0.001$). This correlation was weak in pigs in the conventional system ($r = 0.3$, $P < 0.001$). Among intensive Iberian pigs, there was a strong association between the type of floor and the space allowance. Fully slatted and partly slatted floors were associated to lower space allowance while concrete floor with bedding or outdoor access and outdoor deep bedding floors housed pigs with a higher space allowance (*chi-square* = 63, *df* = 9, *phi-coefficient* = 0.9, $P < 0.001$).

Inclusion of the observer effect did not alter the interpretation of any of the fixed effect except widespread skin condition. The country effect (France *vs.* Spain) was included in the final model of scouring.

DISCUSSION

This study formed part of a larger study which was conducted to assess the welfare of growing pigs in five different production systems found in France or in Spain using the Welfare Quality® protocol and focused on the discussion of animal-based measures related to health. Contrary to the previous findings on “good housing” (Temple et al., 2012), health data assessed in the present study differed poorly between production systems. When focusing on the conventional system, the majority of health indicators did not differ significantly between farms studied in France and Spain. Except scouring, all other measures of health presented low prevalences, as the highest one was 2.5% of conventional pigs with severe wounds on the body. As a result of the large sample size and relatively high variability of the data, the detection of possible causal factors for several measures was achieved mainly in the conventional production system.

Severely wounded animals

The prevalence of severely wounded animals differed between production systems. Extensively kept pigs as well as intensive Iberian pigs presented a lower prevalence of severely wounded animals than pigs in the conventional system. Aggression between individuals is infrequent and rarely injurious among pigs housed under semi-extensive conditions (Stolba and Wood-Gush, 1989). However, low prevalences of wounds in intensive Iberian pigs were less expected as aggression was much more frequent in intensive Iberian pigs than in extensive ones (Temple et al., 2011b). Therefore, wounds on the body did not appear to be a sensitive indicator of the level of aggressions in intensive Iberian pigs. Only severely injured pigs (e.g.those with more than 15 lesions on a given body zone), were recorded later than ten days after mixing. This relatively high threshold may not permit the detection of more “gentle” aggressions. According to Turner et al. (2006), for example, behaviours that do not lead to physical injuries, such as pushing, are not quantifiable using skin lesions score. Furthermore, it should be emphasized that the level of skin lesions may have been underestimated in Iberian pigs due to their dark skin. The breed is an important confounding factor that should be considered when comparing benckmark of wounds between production systems. No significant difference was observed between the conventional and straw bedded systems for the prevalence of wounds, in accordance with Scott et al. (2006).

When focusing on pigs in the conventional system, several possible causal factors could be identified. Severe wounds on the body decreased with the age of the pigs indicating the harmful effects of fighting among newly mixed pigs and the relative stability that followed hierarchical formation (Lyons et al., 1995). To make the assessment more reliable when achieving social stability, the evaluation was done ten days after arrivals. This interval appeared to be too short to avoid the sensitive period around mixing in the conventional system. Seventy two hours after mixing, dominance hierarchy is known to be defined (Meese and Ewbank, 1973); however, in some cases, exhaustion or starvation of the animals may have delayed the increase in fighting until their recuperation. Competition for food has shown to have negative effects on growing pigs, resulting in more skin injuries when fed restrictedly or when the feeding area is limited (Botermans and Svendsen, 2000). In the present study, liquid feeding in trough leaded to an increased risk of severely wounded animals compared with dry or wet hoppers. Liquid food was provided from two to four times per day (restricted in time) whereas pigs were fed *ad libitum* in dry or wet feed hoppers. This difference in food management may explain the increasing risk of wounds in liquid fed pigs in trough. Additionally, the significantly highest prevalence of wounds was recorded on pigs fed by means of Turbomat. In addition to be a restricted feeding system in time providing food once or twice a day, Turbomat may have a lower food accessibility enhancing competition between pigs. Mounting behaviour can cause lesions in the back area of the body. Several studies reported that entire males are more aggressive and sexually active than castrates (Cronin et al., 2003; Fredriksen et al., 2008). A higher prevalence of skin lesions would thus be expected in pens that housed entire males. In the present study, however, there was a trend for an increased risk of wounds in pens that housed castrated males. From that one cannot suggest that castration induces aggressions, but that castration may be used as a supposedly solution on farms with persistent problems of aggressions and consequent lesions on the body.

Tail biting

Tail biting is a serious form of harmful social behaviour (van de Weerd et al., 2005) resulting from a disharmony between the animal and its environment. This behaviour has been mainly described in commercial indoor environment (Schrøder-Petersen and Simonsen, 2001) even though it has also been reported in outdoor herds (Walker and Bilkei, 2006). In the present study, tail biting was associated with intensive production systems as none tail with blood or crust was recorded in extensive conditions. Some types of tail biting originate as misdirected foraging behaviour when there is a lack of proper environment stimulation (Fraser and Broom, 1990; Taylor et al., 2010). Many studies have shown that the provision of straw reduces behaviour directed at pen mates or tail biting levels (Fraser et al.,1991; Lyons et al., 1995; Beattie et al., 2000; Scott et al., 2006). However, no significant differences in the prevalence of tail biting were observed among the three intensive production systems studied. Intensively kept pigs appeared to chew tails even when other substrates were present indicating that pigs were housed in several inappropriate environments independently of the intensive production system studied. It should be considered, though, that outbreaks of tail biting are usually described on affected farms (Paul et al., 2007). This characteristic increases the between-farm as well as the between-pen variability, which together with the relatively low prevalence of tail biting can make difficult the differentiation between production systems. In addition of tail biting associated with rooting behaviours, other types of tail biting are seen in situations of competitiveness and frustration (Taylor et al., 2010). The access to a desired resource such as feeders can give rise to such situations (van de Weerd and Day, 2009). In the conventional production system, liquid feeding in trough was associated with a higher risk of tail biting compared to wet feed hopper. Chambers et al., (1995) reported similar observations; however other farm surveys (Hunter et al., 2001; Moinard et al., 2003) found lower levels of tail biting when pigs were liquid fed rather than fodders. Restriction of food in time in the liquid feeding system may confound the effect of the meal *per se*. Indeed, provision of food in too many meals per day is known to promote disturbance in the pen (Hessel et al., 2006) and possible episodes of tail biting. High stocking density may either contribute to enhance situations of frustration or represent a stressful factor by itself as it increases disturbance into groups. Even though high stocking density was not identified as a possible causal factor in the conventional system, it was associated to a high prevalence of tail biting in pigs in straw bedding. The significant effect of densities on the apparition of tail biting was supported by several studies (Arey, 1991; Beattie et al., 1996; Moinard et al., 2003); but other ones such as Chambers et al. (1995) did not find any association. Thermal stress, resulting from temperatures that are higher or lower than the pigs' thermal comfort zone, also increases tail biting (Geers et al., 1989; Schrøder-Petersen and Simonsen, 2001; Taylor et al., 2010). Estimations of the thermal comfort zone range from 15 to 24ºC and from 14 to 21ºC in growing and finishing pigs, respectively (Gonyou et al., 2006). The highest room temperatures recorded in pigs in straw bedding (24 - 40ºC) were out of the upper limit of the thermal comfort zone and contained all the cases of tail biting in that system. Lohse (1977) also reported higher prevalence of tail biting at high temperatures (35ºC). The straw bedded system was, in the present study, associated with hot environment temperatures which may have contributed to produce outbreaks of tail biting. On the contrary, on conventional farms, the range of temperatures (16 – 29ºC) was closer to the thermal comfort zone and pigs housed in the lowest room temperatures (16 – 21ºC) presented a higher risk of tail biting than pigs housed in the highest ones. An increased activity of the animals under cooler room temperatures may have increased the risk of tail biting in the conventional system. Finally, the age of the pigs was a relevant internal factor of tail biting. Pigs in the conventional and the straw bedded systems reached their predilection age for tail biting during the mid growing phase as also commented by Bracke et al.,(2004). Likewise, Sambraus (1985) observed that tail biting is not a common problem before pigs reach of 90-120 days old. The age of the pigs should therefore be taken into account when interpreting level of tail biting on farm assessments.

Lameness

Lameness was not significantly influenced by the production system. On one hand, there was no indication that extensively reared animals were associated with an increased prevalence of lameness. Similar results were reported by Kilbride et al. (2009) in outdoor housing. On the other hand, prevalences of lameness were similar between the three intensive systems studied. Likewise, Cagienard et al. (2005) did not find any significant difference in the risk of lameness between "animal friendly" and traditional systems in Switzerland. The relatively low prevalence of lameness partially due to the fact that only severely lame animals were recorded may make difficult the discrimination between systems. Taking into account moderate lameness, other studies (Mouttotou et al., 1998; Jørgensen, 2003) found lower prevalences of lameness in the straw bedded system.

Although lameness was not associated with a particular production system, a strong relationship was seen between the prevalence of lameness and fully slatted floors in the conventional system. Slatted floors have been highlighted as a risk for lameness in previous studies (Jørgensen, 2003; Kilbride et al., 2009) and partly slatted floors appeared to reduce considerably the prevalence of lameness in the present study. Slats placed too wide apart or in worn and damaged conditions can originate severe foot injuries. Management, quality and design associated with fully slatted floors are crucial to prevent those lesions. Higher stocking densities increased the risk of lameness in the conventional production system as also commented by Jørgensen (2003). The increase in the prevalence of lameness with the age of conventional pigs might be due to the longer period of time spent on slatted floors or to the increased pressure exerted on the feet as pigs get heavier. Some foot lesions such as wall separation, for example, are known to be associated with an increased body-weight. Finally, the presence of hospital pens should be carefully considered when interpreting prevalence of lameness on farm units. Conventional farms that provided hospital pens presented a significantly lower odd of lameness. As hospital pens were not included in the sample of pens assessed, this last result indicates that farmers isolate correctly severely lame animals. An incorrect management of hospital pens or a lack of availability of them increases the prevalence of lameness assessed by means of the Welfare Quality® protocol. As proposed by Scott et al. (2007), removed or treated animals should be taken into account to have a better understanding of high prevalence of lameness on conventional farms.

Skin condition (localized and widespread skin discolouration)

Skin condition is an unspecific measure that can be a symptom of a variety of different diseases. Local affection of the skin as ear necrosis, systemic infection as PRRS, and environmental diseases as sunburn are some examples of diseases of the skin (Cameron, 2006). Both types of skin condition were correlated moderately in the straw bedded system and weakly in the conventional system; these correlations were not strong enough to decide to analyze them as a sole measure. When comparing production systems, no significant differences could be detected, the between-farm variability was important and both type of skin condition presented over-dispersion. Skin condition is more likely to sporadically affect a farm, independently of the production system. The high variability between farms (and between paddocks) leads towards large confidence intervals that include unity, especially in extensive pigs for which few farms have been assessed. This high variability between farms may suggest that classifying farms by production system is less informative than individual management considerations. Focusing on the conventional system, no possible causal factor could be detected for localized skin condition. As for widespread skin condition, it was highly dependent of one of the four observers implicated in the assessments, whether because of a low inter-observer reliability or because of an association between the observer and the farms affected by that type of problem. Having controlled for the observer effect, discolouration of the skin on more than 10% of the body was associated with pigs that finalized their growing period. However, a specific cause for the problem was not studied in the present work.

Scouring

Scouring was a welfare concern particularly in intensive Iberian pigs that presented a much higher prevalence of pens affected (54%) compared with pigs in the conventional system (9%). Pigs in straw bedding, extensive Mallorcan Black pigs and extensive Iberian pigs presented similar prevalences. It should be considered that more than 30 % of the pens or paddocks assessed could not be evaluated in intensive Iberian pigs for feasibility reasons. Faeces can become liquid when mixing with urine in some conditions of humidity and can hide the evaluation of scouring, especially in dirty pens, affecting the feasibility and the validity of this measure in some production conditions. A wide range of factors are implicated in diarrhea disorders (Thomson et al., 1998; Pearce, 1999). Still, the sole significant possible causal factor identified in the conventional and the straw bedded systems was the age of the animals. Younger animals were associated with an increased risk of scouring. Changes in environments, whether in post-weaned piglet or pigs early in the growing stage, may cause a stress response which is known to be involved in the incidence and severity of enteric diseases (Pearce, 1999). Among intensive Iberian pigs, none possible causal factor could be identified. All the important enteric pathogens in pigs rely on faecal-oral transmissions for their propagation, and the identification of factors linked to the environmental contamination should provide a more effective prediction of scouring problems.

CONCLUSIONS AND ANIMAL WELFARE IMPLICATIONS

In general, the health measures from Welfare Quality® protocol assessed in the present study differed poorly between the five production systems. Like that, the straw bedded and the conventional systems did not differ in the prevalence of any animal-based measures from the "good health" principle. Low prevalences in addition to high between-farm or between-pen variabilities may explain, in part, this homogeneity across systems. For some animal-based parameters such as skin condition, the high variability between farms may suggest that classifying farms by production system may be less informative than individual management considerations. However, Mallorcan Black pigs and Iberian pigs kept extensively showed the lowest prevalence of tail biting and a lower prevalence of severely wounded animals than pigs in the conventional system. In addition, intensive Iberian pigs presented a higher prevalence of scouring compared to pigs in the conventional system. Several possible causal variables for severely wounded animals, tail biting and lameness were identified in the conventional system. The recording of simple environmental-based factors can be useful to detect farms that are more likely to show one of these health problems.

ACKNOWLEDGEMENTS

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. The authors are grateful to the farmers, veterinarians and pig associations for their cordial collaboration. They further thank Emma Fàbrega, Joel Gonzalez, the epidemiology unit of CReSA and two Referees for their helpful suggestions. Déborah Temple was supported by an FPU scholarship from the Ministerio de Educación of Spain.

REFERENCES

Altman DG, Lausen B, Sauerbrei W and Schumacher M 1994 Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute 86:* 829 - 835

Aparicio Tovar MA and Vargas Giraldo JD 2006 Considerations on ethics and animal welfare in extensive pig production: Breeding and fattening Iberian pigs. *Livestock Science 103:* 237–242.

Arey DS 1991 Tail biting in pigs. *Farm Building Progress 105:* 20 – 23

Beattie VE, Walkera N and Sneddon IA 1996 An investigation of the effect of environmental enrichment and space allowance on the behaviour and production of growing pigs. *Applied Animal Behaviour Science 48:* 151-158

Beattie VE, O' Connell NE and Moss BW 2000 Influence of environmental enrichment on the behaviour, performance and meat quality of domestic pigs. *Livestock Production Science 65:* 71–79

Blokhuis HJ 2008 International cooperation in animal welfare: the Welfare Quality® project. *Acta Veterinaria Scandinavica 50:* 1 – 5

Botermans JAM and Svendsen J 2000 Effect of feeding environment on performance, injuries and behaviour in growing– finishing pigs: group-based studies. *Acta Veterinaria Scandinavica 50:* 237– 249

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007 Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16:* 225 - 228

Bracke MBM, Hulsegge B, Keeling L and Blokhuis HJ 2004 Decision support system with semantic model to assess the risk of tail biting in pigs 1. Modelling. *Applied Animal Behaviour Science 87:* 31–44

Broom DM 2006 Behaviour and welfare in relation to pathology. *Applied Animal Behaviour Science* 9*7:* 73–83

Broom DM 2010 Welfare of animals: behavior as a basis for decisions. In: Breed MD and Moore J (eds) *Encyclopedia of animal behavior* pp 580 – 584. Oxford: Academic Press

Cagienard A, Regula G and Danuser L 2005 The impact of different housing systems on health and welfare of grower and finisher pigs in Switzerland. *Preventive Veterinary Medicine 68:* 49-61

Cameron R 2006 Diseases of the skin. In: Straw B, Zimmerman JJ, D'Allaire S and Taylor D (eds) *Diseases of swine* pp 179 – 198. Blackwell Publishing, Iowa

Capdeville J and Veissier I 2001 A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. *Acta Veterinaria Scandinavica 51:* 62-68

Chambers C, Powell L, Wilson E and Green LE 1995 A postal survey of tail biting in pigs in south-west England. *Veterinary Record 136:* 147–148

Courboulay V, Eugène A and Delarue E 2009 Welfare assessment in 82 pig farms: effect of animal age and floor type on behaviour and injuries in fattening pigs. *Animal Welfare 18:* 515-521

Cronin GM, Dunshea FR, Butler KL, McCauley I, Barnett JL and Hemsworth P 2003 The effects of immuno- and surgical-castration on the behaviour and consequently growth of group housed, male finisher pigs. *Applied Animal Behaviour Science 81:* 111–126

Dohoo I, Martin W and Stryhn H 2009 Model-buiding strategies. In: Dohoo I, Martin W and Stryhn H (eds) *Veterinary Epidemiological Reseach* pp 368 – 374. VER Inc: Charlottetown, Price Edward Island, Canada.

Farm Animal Welfare Council 1992 FAWC updates the five freedoms. *Veterinary Record 17*: 357

Fraser D and Broom DM 1990 Abnormal behaviour 3: Addressed to another animal. In: Fraser D and Broom DM (eds) *Farm Animal Behaviour and Welfare* pp 327–328. Baillière Tindall, London

Fraser D, Phillips P, Thompson BK and Tennessen T 1991. Effect of straw on the behaviour of growing pigs. *Applied Animal Behaviour Science 30*: 307–318

Fredriksen B, Lium BM, Marka CH, Mosveen B and Nafstad O 2008 Entire male pigs in farrow-to-finish pens: Effects on animal welfare. *Applied Animal Behaviour Science 110:* 258-268

Geers R, Dellaert B, Goedseels V, Hoogerbrugge A, Vranken E, Maes F and Berckmans D 1989 An assessment of optimal air temperatures in pig houses by the quantification of behavioural and health-related problems. *Animal Production 48:* 571-578

Gonyou HW, Lemay SP and Zhang Y 2006 Effects of the environment on productivity and disease. In: Straw B, Zimmerman JJ, D'Allaire S and Taylor D (eds) *Diseases of swine* pp 1027 – 1038. Blackwell Publishing, Iowa

Guy JH, Rowlinson P, Chadwick JP and Ellis M 2002 Health conditions of two genotypes of growing-finishing pig in three different housing systems: implications for welfare. *Livestock Production Science 75:* 233–243

Hessel EF, Wulbers-Mindermann M, Berg C, Van den Weghe HFA and Algers B 2006 Influence of increased feeding frequency on behavior and integument lesions in growing–finishing restricted-fed pigs. *Journal of Animal Science 84:* 1526–1534.

Hunter EJ, Jones TA, Guise HJ, Penny RHC and Hoste S 2001 The relationship between tail biting in pigs, docking procedure and other management practices. *The Veterinary Journal 161:* 72–79.

Jaume J and Alfonso L 2000 The Mallorcan Black pig. *Agri 2000 27:* 53-58

Jørgensen B 2003 Influence of floor type and stocking density on leg weakness, osteochondrosis and claw disorders in slaughter pigs. *Animal Science 77*: 439-449 (Abstract)

Kilbride AL, Gillman CE, Green LE 2009 A cross-sectional study of the prevalence of lameness in finishing pigs, gilts, and pregnant sows and associations with limg lesions and floor types on commercial farms in England. *Animal Welfare 18:* 215 - 224

Lee Y, Nelder JA and Pawitan Y 2006 Generalized linear models. In: Lee Y, Nelder JA and Pawitan Y (eds) *Generalized Linear Models with Random effects* pp 54 – 64. Chapman and Hall/ CRC, Boca Raton, FL, USA

Lohse E 1977 Einfluss unterschiedlicher umwelttemperaturen auf die motorische aktivität von ferklen. Institut für tierproduktion der technischen *Universität Berlin Journal 925:* 68–71

Lyons CAP, Bruce JM, Fowler VR and English PR 1995 A comparison of productivity and welfare of growing pigs in four intensive systems. *Livestock Production Science 43:* 265–274

MAPA 2007 Real Decreto 1469/2007, de 2 de noviembre, por el que se aprueba la norma de calidad para la carne, el jamón, la paleta y la caña de lomo ibéricos. BOE 264, 45087–45104

Meese GB and Ewbank R 1973 The establishment and nature of the dominance hierarchy in the domesticated pig. *Animal Behaviour 21:* 326 - 334

Molony V and Kent JE 1997 Assessment of acute pain in farm animals using behavioral and physiological measurements *Journal of Animal Science 75:* 266 - 272

Moinard C, Mendl M, Nicol CJ and Green LE 2003 A case control study of on farm risk factors for tail biting in pigs. *Applied Animal Behaviour Science 81:* 333 - 355

Mouttotou N, Hatchell FM and Green LE 1998 Adventitious bursitis of the hock in finishing pigs: prevalence, distribution and association with floor type and foot lesions. *Veterinary Record 142:* 109–114

Paul ES, Moinard C, Green LE and Mendl M 2007 Farmers' attitudes to methods for controlling tail biting in pigs. *Veterinary Record 160:* 803–805

Pearce GP 1999 Epidemiology of enteric disease in grower-finisher pigs: A postal survey of pig producers in england. *Veterinary Record 144:* 338-342

Rainville P 2002 Brain mechanisms of pain affect and pain modulation. *Current Opinion in Neurobiology 12:* 195-204

Rodríguez-Estévez V, García A, Peña F and Gómez AG 2009 Foraging of Iberian fattening pigs grazing natural pasture in the dehesa. *Livestock Science 120:* 135-143

Sambraus HH 1985 Mouth-based anomalous syndromes. In: Fraser AF (ed) *Ethology of Farm Animals* pp 391 – 422. Elsevier, Amsterdam, the Netherlands.

Schrøder-Petersen DL and Simonsen HB 2001 Tail biting in pigs. *The Veterinary Journal 162:* 196-210

Scott K, Chennells DJ, Campbell FM, Hunt B, Armstrong D, Taylor L, Gill BP and Edwards SA 2006 The welfare of finishing pigs in two contrasting housing systems: Fully-slatted versus straw-bedded accommodation. *Livestock science 103:* 104 - 105

Scott K, Chennells DJ, Armstrong D, Taylor L, Gill BP and Edwards SA 2007 The welfare of finishing pigs under different housing and feeding systems: liquid versus dry feeding in fully-slatted and straw-based housing. *Animal Welfare 16:* 53-62

Stolba A and Wood-Gush DGM 1989 The behaviour of pigs in a semi-natural environment. *Animal Production 48:* 419–425

Taylor NR, Main DCJ, Mendl M and Edwards SA 2010 Tail-biting: A new perspective. *The Veterinary Journal 186:* 137 – 147

Temple D, Dalmau A, Ruiz de la Torre JL, Manteca X and Velarde A 2011a Application of the Welfare Quality® protocol to assess growing pigs kept under intensive conditions in Spain. *Journal of Veterinary Behavior 6*: 138–149

Temple D, Manteca X, Velarde A and Dalmau A 2011b Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Applied Animal Behaviour Science 131:* 29 – 39

Temple D, Courboulay V, Velarde A, Manteca X and Dalmau A 2012 The welfare of growing pigs in five production systems: Assessment of Feeding and Housing. *Animal 6:* 656-667

Thomson JR, Smith WJ and Murray BP 1998 Investigations into field cases of porcine colitis with particular reference to infection with serpulina pilosicoli. *Veterinary Record 142:* 235-239

Turner SP, Farnworth MJ, White IMS, Brotherstone S, Mendl M and Knap P 2006 The accumulation of skin lesions and their use as a predictor of individual aggressiveness in pigs. *Applied Animal Behaviour Science 96*: 245 - 259

van de Weerd HA, Docking CM, Day JEL and Edwards SA 2005 The development of harmful social behaviour in pigs with intact tails and different enrichment backgrounds in two housing systems. *Animal Science 80:* 289-298

van de Weerd HA and Day JEL 2009 A review of environmental enrichment for pigs housed in intensive housing systems. *Applied Animal Behaviour Science 116:* 1–20

Walker PK and Bilkei G 2006 Tail-biting in outdoor pig production. *The Veterinary Journal 171*: 367–369

Welfare Quality® 2009 Welfare Quality® applied to growing and finishing pigs. In Welfare Quality® Assessment protocol for pigs (eds A Dalmau, A Velarde, K Scott, S Edwards, I Veissier, L Keeling, A Butterworth), pp 49-78. Welfare Quality® Consortium, The Netherlands

Whay HR, Main DCJ, Greent LE and Webster AJF 2003 Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. *Animal Welfare 12:* 205- 217

Whay HR 2007 The journey to animal welfare improvement. *Animal Welfare 16:* 117-122

Wolfinger R and O'Connell M 1993 Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *Journal of Statistical Computation and Simulation* 4: 233–243

CHAPTER 4

ASSESSMENT OF ANIMAL WELFARE THROUGH BEHAVIOURAL PARAMETERS IN IBERIAN PIGS IN INTENSIVE AND EXTENSIVE CONDITIONS

Based on Temple D, Manteca X, Velarde A, Dalmau A 2011. Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Applied Animal Behaviour Science 131:* 29-39

ABSTRACT

Behaviour is a significant component of well-being and should be properly considered in a pig welfare assessment. However, interpretation of variations in behaviour is usually not straightforward, especially when dealing with animals housed and managed in a variety of systems. The fourth principle of the Welfare Quality® project, labelled "Appropriate behaviour", was assessed on 21 farms (11 extensive and 10 intensive) with a total of 25 856 Iberian pigs (*Sus scrofa)*. The methodology included scan samplings of social and exploratory behaviour, human-animal relationship tests and qualitative behaviour assessments (QBA). Changes in occurrence of behaviour and qualitative measures were evaluated and discussed when comparing Iberian pigs either in intensive or extensive rearing conditions. Negative as well as positive social behaviours were significantly ($P < 0.001$) more frequent among intensive Iberian pigs compared to extensive animals. In intensive Iberian pigs, both types of social behaviours were highly correlated ($r_s = + 0.62$). No significant differences were observed in the occurrence of exploratory behaviour and in the proportion of animals exhibiting a panic response to the presence of humans between intensive and extensive Iberian pigs. It is suggested that the high occurrences of social negative behaviours recorded in intensive conditions are clearly an indicator of poor welfare; but that interpretations of variations in "positive" social and explorative behaviour are not straightforward and may lead to misinterpretations. Despite its subjectivity, the qualitative behaviour assessment appears useful to discriminate farms.

Key words: animal welfare; behaviour assessment; extensive; housing systems; Iberian pig

INTRODUCTION

It is generally accepted that animal welfare comprises physical and mental health (Dawkins, 2004; Webster et al., 2004) and that it includes several aspects such as absence of thirst, hunger, discomfort, disease, pain and injuries, stress and the expression of normal behaviour (Farm Animal Welfare Council, 1992). Based on this multidimensional concept of welfare, the Welfare Quality® project aimed to develop standardised on-farm assessment systems which provide reliable information on the way animals are produced. To that effect a multicriteria approach was built on four main principles of animal welfare: good feeding, good housing, good health and appropriate behaviour. Each of these four principles included several independent but complementary criteria, with a total of 12 criteria (Botreau et al., 2007). A set of predominantly animal-based measures was used to assess each one of these criteria after estimation of their validity and repeatability in pilot studies.

The fourth principle, labelled "Appropriate Behaviour", deals with the evaluation of behaviours and includes the assessment of the expression of social and other behaviours (such as exploratory behaviour in pigs), good human-animal relationship and the presence of a positive emotional state. Behaviour is a component of animal welfare and occurs as a consequence of the animal's motivational state and therefore the quantification of behaviour patterns is in fact a measure of motivation (Dellmeier, 1989). Expression of social and other behaviour as well as good human-animal relationship refer to variations in frequency of behaviours whereas the presence of a positive emotional state is based on a qualitative assessment. The qualitative assessment of animal behaviour relies on the ability of observers to integrate perceived details of behaviour, posture, and context into descriptions of animal's style of behaving using descriptors such as "relaxed", "tense", "frustrated" and "content" (Wemelsfelder, 2007). Even though a qualitative assessment may open the door to an anthropomorphic projection, recent studies have reported its scientific and biological validity (Rousing and Wemelsfelder, 2006; Wemelsfelder, 2008; Wemelsfelder et al., 2001).

Changes in behavioural patterns often represent the first level of response of an animal to an aversive or stressful environment. Then, behaviours which differ in pattern, frequency or context from those which are shown by most members of the species in conditions that allow a full range of behaviours are usually considered "abnormal" behaviours (Fraser and Broom, 1990). Many animal welfare studies seek to compare how animals behave in various housing environments or rearing conditions. More concretely, the comparison of behaviour of intensively and extensively kept animals has been used to determine how confined environment affects the development of behaviour. For instance, Krohn et al. (1994) compared grooming, exploration and abnormal behaviour of dairy cows kept in loose housing and tie stall; Kerr and Wood-Gush (1987) observed changes in frequencies and sequences of behaviour of intensively and extensively reared calves and proportions of beak-related behaviours were studied in intensive and free ranging hens (Shimmura et al., 2008). Nonetheless, interpretation of variations in behaviour is usually not straightforward and originate many discussions on what should be considered as normal or not or even prejudicial or not for the animal, especially when dealing with animals housed in a variety of systems (Rushen and de Passillé, 2009).

Derived from the previous concern, this study focuses on the evaluation of “Appropriate behaviour” in Iberian pigs. Iberian pig farming system was the one of election for its wide diversity (Aparicio Tovar and Vargas Giraldo, 2006). Some medium-sized farms (around 500 animals per farm) follow a traditional system in which Iberian pigs meet most of their food requirements grazing on pastureland during a long growing period (fattened animals are over 24 month-old). Depending on the availability of resources, especially of acorns, Iberian pigs may receive a food supplementation. This last rearing system is generally considered as extensive. More recently, farming methods tend to reduce the production cycle and use systematically fodder leading to an intensification of the rearing system. As a result, intensive Iberian pig rearing ranges from conventional farms with fully slatted floor to farms with outdoor access. Besides this diversity in production systems, the Iberian pig is a rustic breed highly adapted to an extensive rearing system. Its adaptation to more recent intensive rearing systems may, thus, rises new welfare implications inherent to the Iberian breed.

To sum up, this study does not pretend to have an extensive approach of the Iberian pig behaviour but to evaluate the changes in occurrence of behaviour and qualitative measures from the “Appropriate Behaviour” principle of the Welfare Quality® scheme when assessing Iberian pigs in intensive and extensive rearing systems. It will give causes for discussing the validity of these different measures across Iberian pig rearing systems. Finally, it will provide a first sight on the Iberian pig welfare applying the Welfare Quality® protocols in extensive and intensive production systems.

MATERIAL AND METHODS

Welfare assessment system

The Welfare Quality® assessment system consisted of four main principles of animal welfare: good feeding, good housing, good health and appropriate behaviour. Each of these four principles included 12 independent criteria (Botreau et al., 2007). This study focuses on the different animal-based measures from the "Appropriate Behaviour" principle evaluated on Iberian pigs in intensive and extensive conditions (**Table 1**).

Table 1 Recorded measures for the welfare assessment system of growing pigs on intensive and extensive Iberian pig farms (Welfare Quality®, 2009).

	Welfare Criteria	Measures
Appropriate behaviour	Expression of social behaviours	Social behaviour
	Expression of other behaviours	Exploratory behaviour
	Good human-animal relationship	Fear of humans
	Positive emotional state	Qualitative Behaviour Assessment

Animals and farming systems

Iberian pigs from 11 extensive and 10 intensive farm units were evaluated on the basis of the "Appropriate Behaviour" principle.

Extensive Iberian pigs were kept in outdoor paddocks whose size varied from 2000 m^2 to 400 ha. All paddocks provided covered facilities as well as wallows, puddles or small temporal streams to allow the animals to cool themselves. Pigs were grazed on pasture lands eating acorns and other resources (grass, bushes, roots, tubers, etc.), although during the assessment period (April-May) a supplement of concentrated fodder was provided manually once or twice a day. The average group size was 185 pigs per paddock and the average space allowance per pig was 387 m^2 ± SD = 419.6 m^2. Pigs entered the paddocks at an average weight of 48 kg and left to be slaughtered at 150-160 kg. Extensive pigs assessed were meanly 280 day-old ranging from 42 to 420 days old. Iberian pigs from intensive farms were raised within a great diversity of housing conditions showing different levels of intensification. Three out of the 10 farms evaluated housed pigs on concrete floor providing outdoor access only during the finishing phase (three last months of the growing period); two farms housed pigs indoor on straw or sand; and five farms housed pigs on concrete floor with resting areas providing an outdoor access during the whole period. All pigs were fed *ad libitum* with concentrated fodder. Group size ranged from seven pigs per pen to 320 pigs per paddock and the average space allowance per pig was 2.2 m^2, ranging from 0.66 m^2 in more conventional housings to 1.5 m^2 in indoor straw or sand bedding systems and 3.2 m^2 on farms with outdoor access during the whole period. Pigs entered at an average weight of 22 kg and left to be slaughtered at 140-160 kg. Intensive pigs assessed were meanly 160 days old ranging from 37 to 300 days old.

Behavioural measurements

Behavioural measures were assessed applying the methodology of assessment found in the Welfare Quality® Assessment protocol for pigs (2009). Two observers carried out the farm visits. In order to minimize difference between observers and to standardize the behavioural observations, observers followed an identical training (in intensive conditions as well as in extensive ones) prior beginning the assessment.

Social and exploratory behaviour

Social and exploratory behaviours were assessed on 3707 extensive Iberian pigs and 6524 intensive Iberian pigs by means of scan samplings. Firstly, pigs were scored as active or inactive. Then, the behaviours recorded from active pigs were: positive social behaviour, negative social behaviour, exploration and “other” behaviour (**Table 2**). The measure exploration included exploration of the pen/paddock and exploration of enrichment material. Before beginning the scan, the assessor ensures that all animals observed were standing up. If necessary, the observer would clap the hands. The observation of all the animals was recorded five minutes later from outside the pen in intensive conditions and from a distance higher than three metres from the group in the case of extensive conditions. Each group of pigs was observed five consecutive times with a two and a half minute interval between two scans. Three different groups of animals per farm were assessed using this method. Each group included an average of 170 animals. In extensive conditions, the use of binoculars was sometimes necessary.

Table 2 Summary of recorded behaviours during the scan sampling (Welfare Quality®, 2009).

Active behaviour	
Negative social behaviour	Aggressive behaviour, including biting or any social behaviour with a response from the disturbed animal.
Positive social behaviour	Sniffing, nosing, licking and moving gently away from the animal without an aggressive or flight reaction from this individual.
Exploration of the pen	Sniffing, nosing, licking all features of the pen or paddock.
Exploration of enrichment material	Exploration towards straw or other suitable enrichment material.
"Other" active behaviour	All other active behaviours (eating, drinking, air sniffing, etc.)
Non- Active behaviour	Resting.

Human Animal Relationship test

Human Animal Relationship (HAR) was evaluated according to the fear to human test validated by Courboulay and Foubert (2007). Up to ten selected pens or groups of extensive animals were tested throughout the farm unit. Each pen or group of animals was assessed as a whole. Only two possibilities were considered for this test: **0** – No panic response to human presence; **2** – More than 60% of the animals in the pen/group showing panic to the human presence. Panic was defined as animals fleeing, facing away from the observer or huddled in the corner of the pen.

Qualitative behaviour assessment

The qualitative behaviour assessment (QBA) uses descriptive terms with an expressive connotation to reflect animals' experience of a situation (Wemelsfelder, 2007). A rating scale was used to score pigs at group level based on 20 different terms: 1.Active, 2.Relaxed, 3.Fearful, 4.Agitated, 5.Calm, 6.Content, 7.Tense, 8.Enjoying, 9.Frustrated, 10.Sociable, 11.Bored, 12.Playful, 13.Positively occupied, 14.Listless, 15.Lively, 16.Indifferent, 17.Irritable, 18.Aimless, 19.Happy and 20.Distressed. The scale was of 125 millimetre length whose left side (or minimum) meant that the expressive quality indicated by the term was entirely absent in any of the pigs observed, and right side (maximum) indicated that the descriptor was dominant across all pigs. The scoring was made inside this scale according to the number of animals showing each one of the terms used and the percentage of time that animals were observed in each state. The assessment was carried out at farm level after animals in a number of pens or paddocks had been observed for 20 minutes (i.e. 8 points of observation and 2.5 minutes per point of observation). This methodology was validated by Wemesfelder and Millard (2009).

Statistical analysis

The SAS statistical package (SAS Inst.Inc., Cary, NC) was used to analyse data.

Social and exploratory behaviour analysis

Negative social behaviour, positive social behaviour, exploration and "other" behaviour were expressed in proportion to the total number of active pigs. The percentage of active pigs was expressed in proportion to the total number of observations (active + resting animals). A Mann-Whitney Wilcoxon test was used to investigate whether significant differences existed between frequencies of behaviours of Iberian pigs kept in intensive and extensive conditions. Spearman's correlation coefficients between behavioural measures of the scan sampling were calculated in extensive Iberian pigs and in intensive Iberian pigs. A correlation was considered to be statistically significant when its P value was less than 0.05.

HAR analysis

The proportion of pens/groups of animals showing a panic response among pigs in extensive conditions and pigs housed on intensive farms was compared using a Chi-square test.

Qualitative analysis

QBA data were expressed at farm level and were analysed using a Principal Component Analysis (PCA) by means of the SAS PRINCOMP procedure. PCA reduces the number of original variables (descriptive terms) into a fewer dimensionality (principal components) explaining as much of the variability between farms as possible. A correlation matrix was used for a more balanced representation of the descriptive terms and no rotation was applied. The first two principal components whose eigenvalues were greater than 1.0 produced a two-dimensional interpretative word chart. In each chart, the 20 descriptive terms were plotted in relation to the first two axes. The word chart can be useful to lead the interpretation of the first two principal components. The eigenvectors (or loadings) from the descriptor data quantify the weight each descriptor has on the two main axes (Rencher, 2002). Each farm unit receives a score on each of the main principal component. Next, a Mann-Whitney Wilcoxon test was used to test whether the scores obtained for pigs in extensive units and pigs on intensive farms were significantly different.

RESULTS

Social and Exploratory behaviour

Table 3 shows the mean frequencies of behaviour recorded during the scans for Iberian pigs in extensive and intensive conditions. The occurrences of negative social behaviour as well as positive social behaviour were significantly higher in Iberian pigs kept in intensive conditions ($P < 0.001$ and $P = 0.0003$, respectively). On the other hand, there were no significant differences between intensive and extensive Iberian pigs in the expression of explorative and "other" behaviours. Given that pigs displayed exploration of enrichment material only in one pen of a single intensive farm, this category was summed to exploration of the pen and was studied as an overall explorative activity.

Table 3 Mean occurrences and SE of behaviours recorded during the scan sampling applied in Iberian pigs in extensive and intensive conditions.

Behavioural measure	Extensive conditions	Intensive conditions	*P* value
Negative social	1.0 ± 0.27	5.1 ± 1.03	< 0.001
Positive social	2.3 ± 0.68	10.0 ± 1.11	0.0003
Exploratory	40.9 ± 7.03	39.0 ± 3.14	NS
"Other"	55.8 ± 7.31	47.3 ± 2.26	NS
Active	71.4 ± 6.17	68.7 ± 3.77	NS

NS: Non-Significant

In intensive Iberian pigs, there was a significant positive correlation between the occurrence of negative social behaviour and positive social behaviour (**Table 4**). In addition, negative social behaviour was negatively correlated with the frequency of exploration. In extensive Iberian pigs, there was no significant correlation between positive and negative social behaviours (**Table 5**).

Table 4 Spearman's correlation coefficients between behavioural measures of the scan sampling of Iberian pigs housed in intensive conditions.

	Negative Social	Positive Social	Exploratory	"Other"	Active
Negative Social	1				
Positive Social	0.62***	1			
Exploratory	-0.38*	-0.11	1		
"Other"	0.07	-0.21	-0.89***	1	
Active	0.07	-0.34	0.15	-0.11	1

*** P < 0.001
** P < 0.01
* P < 0.05

Table 5 Spearman's correlation coefficients between behavioural measures of the scan sampling of Iberian pigs housed in extensive conditions.

	Negative Social	Positive Social	Exploratory	"Other"	Active
Negative Social	1				
Positive Social	0.18	1			
Exploratory	-0.11	0.26	1		
"Other"	0.05	-0.34	-0.99***	1	
Active	-0.28	-0.41*	0.16	-0.09	1

*** P < 0.001
** P < 0.01
* P < 0.05

HAR test

The frequency of groups of animals exhibiting a panic response was not significantly different in extensive Iberian pigs (29.0%) compared with Iberian pigs raised in intensive conditions (15.8%) ($Z = 2.2$, $P = 0.14$). However, when a panic response was seen on extensive units, it concerned from 50% to 100% of the groups of animals assessed whereas on intensive farms these proportions ranged from 17% to 40% (**Figure 1**).

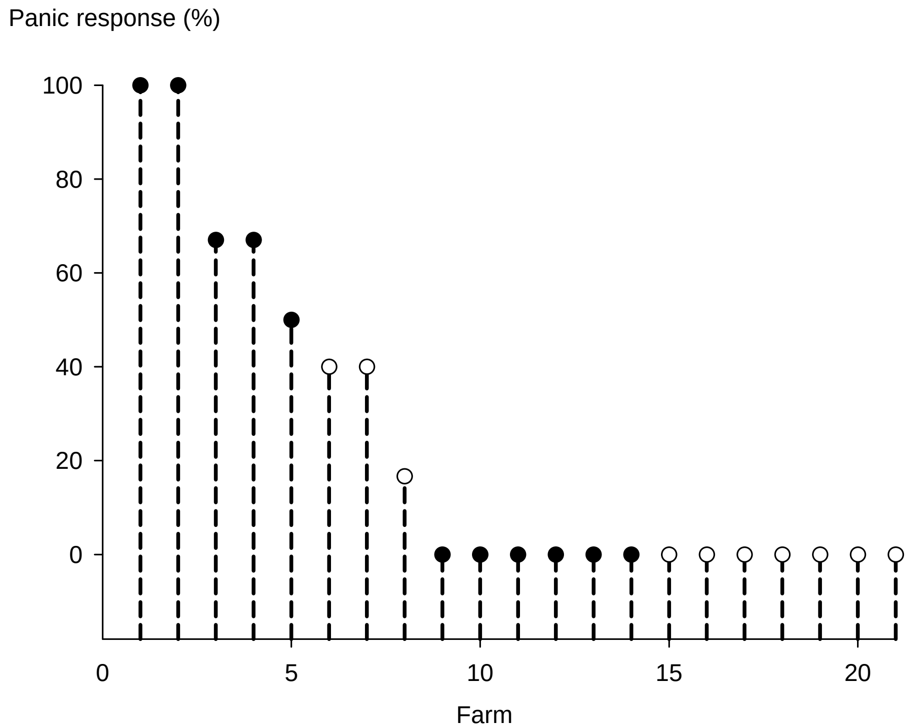


Figure 1 Proportion of pens or groups of animals showing a panic response to the observer on extensive units (black circles) and intensive farms (white circles).

Qualitative behaviour assessment

Two main axes were retained accounting for a cumulative percent of variability between farms of 58% (42% and 16%, respectively). Also, the eigenvalues for these first two axes were greater than 1. The PCA word chart is shown in **Figure 2**. Happy (0.34), enjoying (0.33), content (0.32), positively occupied (0.31), and lively (0.31) as well as listless (-0.26), bored (-0.24) and frustrated (-0.24) showed the highest loadings on the first axis. The second axis was characterised with terms such as sociable (0.50), agitated (0.43), indifferent (0.40) as well as distressed (-0.34) and listless (-0.28). The position of each farm on the basis of the QBA was plotted in **Figure 3**. The rearing system had a significant effect on the scores of the farms on the first axis (Z = -3.35; P = 0.0008). Pigs housed in extensive conditions obtained significantly higher scores, meaning that they were assessed as more happy, content, enjoying, positively occupied and lively than pigs housed in intensive conditions. There was no significant differences between the scores of intensive and extensive Iberian pigs on the second axis (Z = 1.23; P = 0.21).

The plot of Iberian pigs assessed only on intensive farms showed a homogeneous distribution of the farms throughout the two axes (**Figure 4**). These two axes explained 43% and 16% of the total variation between animals from intensive farms. As described previously, the first axis was characterised with terms ranging from listless (-0.30) and frustrated (-0.21) to positively occupied (0.32) or lively (0.31). The second axis presented a more ambiguous interpretation as it was characterized with terms ranging from fearful (-0.36) and active (-0.35) to agitated (0.38) and relaxed (0.40).

On the contrary, the plot of Iberian pigs assessed only in extensive conditions presented little dispersion (**Figure 5**). The first axis accounted for 37% of the total variance and was characterized with terms ranging from tense (-0.33) and irritable (-0.33) to enjoying (0.34) and happy (0.35). The second axis explained 22% of the total variance and was described by terms ranging from agitated (-0.41) and positively occupied (-0.31) to calm (0.32), bored (0.32) and aimless (0.31) resulting, thus, difficult to interpret. Pigs from six out of 11 extensive units obtained relatively high scores on the basis of the first axis while two outliers could be identified (**Figure 5**).

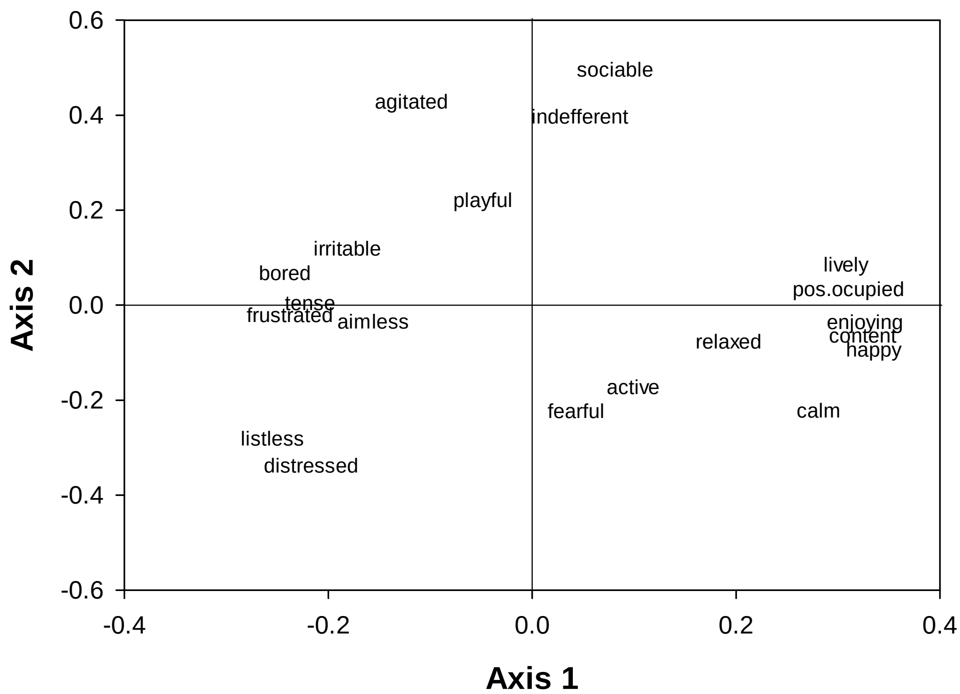


Figure 2 Word chart of the QBA assessed in extensive and intensive Iberian pigs. Axes reflect eigenvectors (loadings) of descriptive terms on the two principal PCA dimensions.

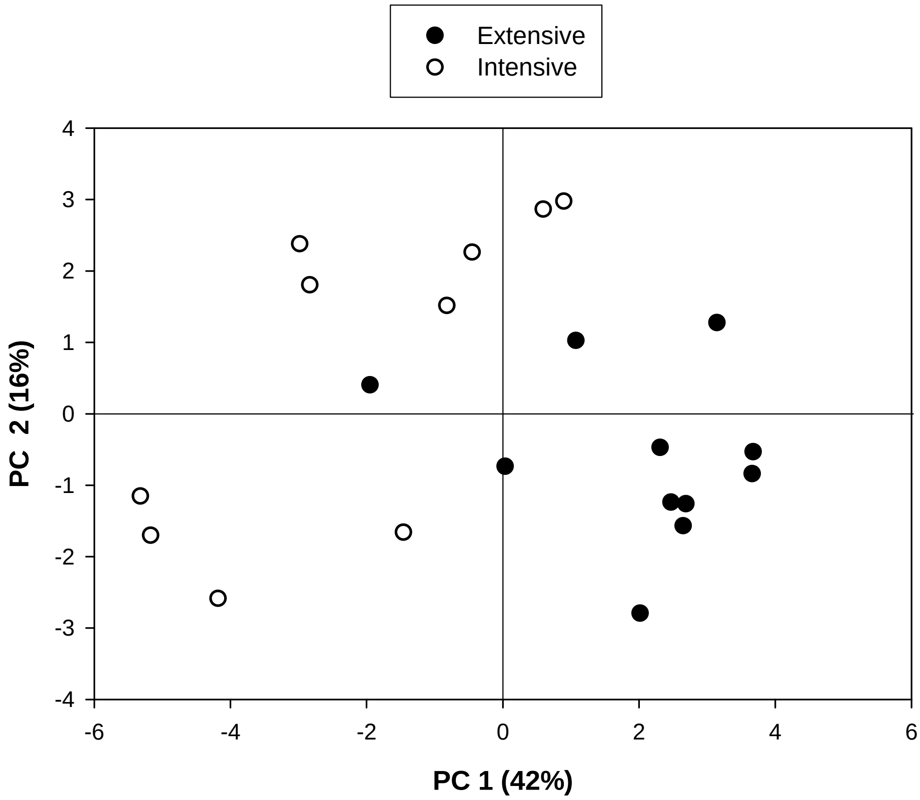


Figure 3 Plotted values of the first two main components for extensive and intensive farms.

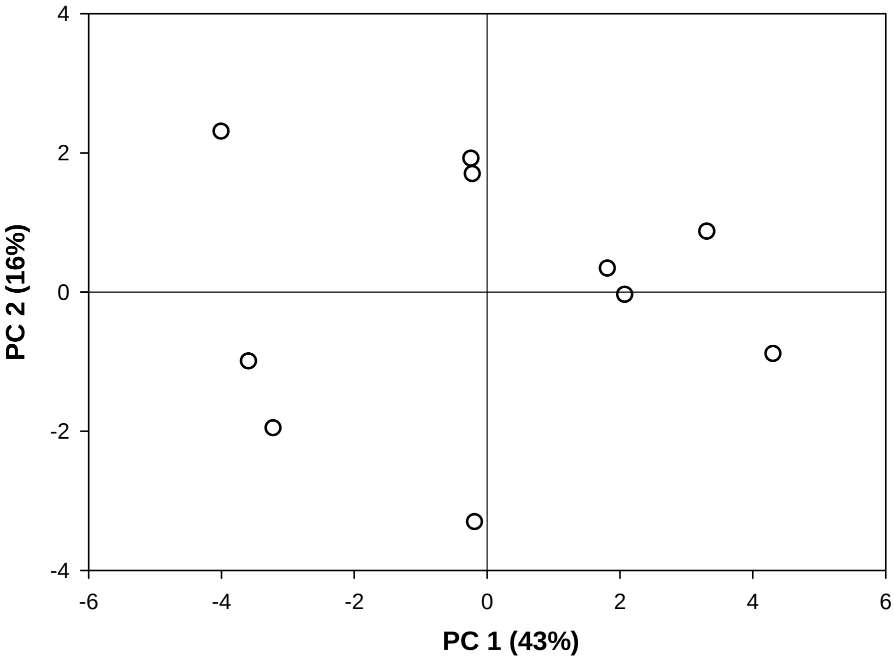


Figure 4 Plotted values of the first two main components of intensive farms.

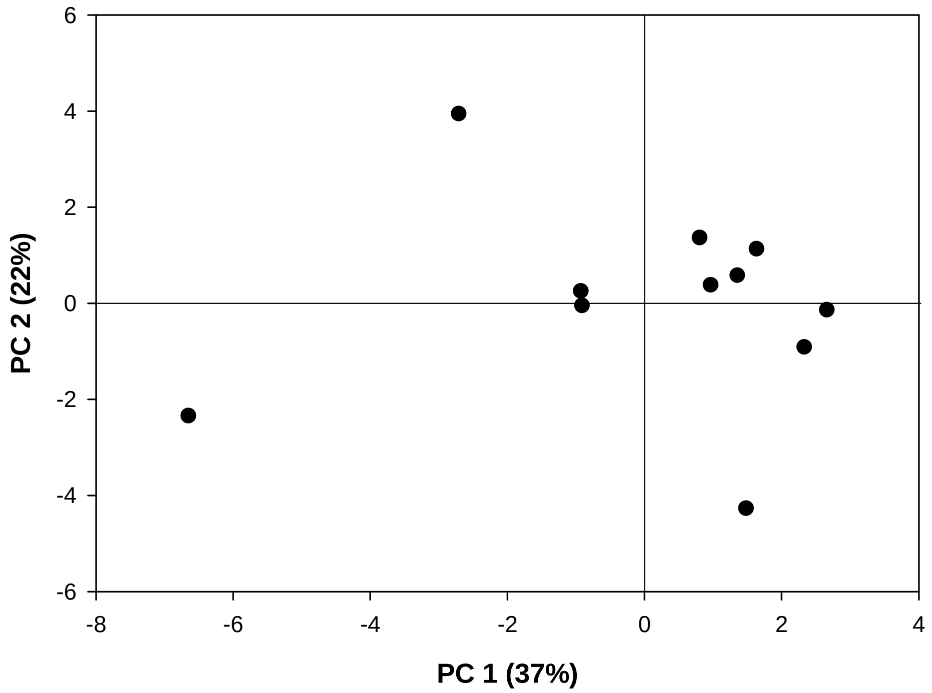


Figure 5 Plotted values of the first two main components of extensive units.

DISCUSSION

Expression of social behaviour

Positive social behaviour such as licking is related to good welfare. Firstly, it has shown to elicit physiological responses known to be pleasant (Panksepp, 2006). In extensive conditions, Iberian pigs performed 2.3% of positive social behaviours which was in accordance with the 3% of positive social behaviour recorded by Stolba and Wood-Gush (1989) when observing pigs in semi-natural environment. These authors reported social grooming in subgroups of pigs and emphasized the peculiarity of grunt or naso-nasal contact when members approached a nest site. In the present study, the frequency of positive social contacts increased as the Iberian pigs got less active in extensive conditions. Copado et al. (2004) also observed that many non-agonistic interactions in free ranging pigs were performed while pigs were more inactive or lying one next to the other. Positive social behaviour can therefore be considered as a pleasant emotion that promotes social cohesion and affiliation among conspecifics, under relaxed situations. Secondly, positive social behaviour reduces the negative effects of stressful events. This effect, known as “social buffering” (Kikusui et al., 2006), has been observed in several social species such as rats (Taylor, 1981), sheep

(Porter et al., 1995) and cattle (Mounier et al., 2006). Iberian pigs reared intensively presented much higher occurrence of positive social behaviour than pigs in extensive rearing systems (10.0% and 2.3%, respectively). Intensification can lead to an increase in the number of stressful situations to which the animals are submitted through their growing period. An increase in positive social contacts with conspecific could therefore be perceived as a strategy used by Iberian pigs to find relief from stressful experiences associated with a more intensive environment. De Jong et al. (1998) described an increase in allogrooming activity in pigs housed in barren environment compared with those in enriched housing systems. In intensive conditions, positive social behaviour may therefore have a positive influence on welfare even though it might be originated by a stressful situation.

Iberian pigs assessed in intensive conditions showed also a significantly higher occurrence of negative social behaviour compared with extensive Iberian pigs (5.1% and 1.0%, respectively). Intensive rearing conditions tend to foment competition for resources between pigs increasing the occurrence and duration of negative social interactions. A limited access to resources together with high stocking densities originates such competitive situations (Ewbank and Bryant, 1972). Besides the fact that high stocking densities cause dominance hierarchy to be less successful in controlling aggressions within a group, a limited space allowance impedes an animal to signal submission by fleeing when a fight occurs (Gonyou, 2001). The frequency of negative social behaviour out of the total of social behaviours among pigs in intensive and extensive conditions was similar (34% and 30%, respectively). An increased proximity between animals, as it happens in intensive conditions, may increase the probability of performing a social contact that can either be positive or negative, as well as the duration. In fact, a relatively high positive correlation was found between the occurrence of positive and negative social behaviours (r_s = 0.62) among Iberian pigs in intensive conditions. This strong relationship suggests the possibility of common underlying causes of both types of social behaviours. For example an interaction may begin as a positive one (i.e. nosing or licking) and end up in a negative one (i.e. biting). Therefore, high levels of positive social activity may not necessarily reflect a positive sates of the animal (Boissy et al., 2007). In cattle, allogrooming may be related to social dominance and whether an increased of allogrooming is performed by dominant or subordinate animals is contradictory (Mülleder et al., 2003; Sato et al., 1993). Non-agonistic social behaviour may even sometimes be considered as abnormal behaviour (Wiepkema et al., 1983). In consequence, while negative social behaviour is clearly an indicator of poor welfare, the interpretation of high frequencies of positive social behaviour is not straightforward and should be careful especially when registered among Iberian pigs in intensive conditions.

Expression of exploratory behaviour

Pigs are strongly motivated in performing exploratory behaviour such as rooting which in nature is necessary to search food. Preventing pigs from doing so by keeping them in a barren environment increases the risk of abnormal behaviours such as tail biting (Moinard et al., 2003; Petersen et al., 1995). In the present work, 41% of active behaviours were destined to exploratory activity in extensive conditions which was lower than the 54% daylight hours engaged in foraging proposed by Rodríguez-Estévez et al. (2009). In this last study, pigs were entirely reliant on natural resources while in the present assessment pigs received a food supplementation. Feeding necessity probably increases foraging activity and assessment of extensive Iberian pigs during the *montanera* (three months period of autumn-winter during which animals are fed essentially with acorns) will probably increase their motivation to explore. In intensive conditions, a decrease of exploratory behaviour was associated with an increase in negative social behaviour. Under unsuitable housing conditions, behaviour patterns which normally are directed towards the physical environment are released on inappropriate objects, such as the body of pen mates (Petersen et al., 1995; Van Putten and Dammers, 1976). Part of negative social interactions may be misdirected foraging behaviours when there is a lack of proper environment stimulation (Fraser and Broom, 1990). Thereafter, it seems that levels of exploratory behaviour in intensive conditions (39%) did not appear to fulfil the behavioural needs of the Iberian pigs. Even though more than 50% of the farms housed pigs in pens with straw or sand which have been suggested to decrease the manipulation of penmates (Appleby and Wood-Gush, 1988; Fraser et al., 1991) pigs from only one pen of a single farm displayed exploration of a suitable enrichment material. Definition of a suitable enrichment material gives too much room for interpretation (Bracke et al., 2006). Material use decreases over time (Docking et al., 2008) and as it is destroyed or soiled (Blackshaw et al., 1997). The presence of substrates spoiled with manure or not frequently renewed may explain the loss of interest for exploration and infrequent manipulation of enrichment material recorded among intensive Iberian pigs. When comparing Iberian pigs in extensive and intensive conditions, levels of exploratory behaviour were not significantly different. Moreover, those frequencies were very similar to the one found by Temple et al. (in press) applying the same methodology in “white” pigs on conventional farms with concrete floors (38%). Several reasons may explain this similarity in the expression of exploratory behaviour between pigs in extensive and intensive conditions. First of all, pigs are highly motivated to explore even if there are no obvious novel stimuli which may elicit the behaviour (Wood-Gush and Vestergaard, 1993). In barren environments, pigs usually spent more time examining surroundings (Fraser et al., 1991; Ruiterkamp, 1987). Furthermore, variation in the frequencies of exploration was inversely associated with variation in the frequencies of “other” behaviour using the scan sampling methodology. This last measure represented an important percentage of all active behaviours in intensive pigs (47%) and in extensive ones (56%). It included well defined behaviours that were seen in both farming conditions such as drinking, eating, and watching the surroundings; however, it also included behaviours that were more dependent of the rearing conditions. Motionless sitting, which is more likely interpreted as a form of apathy where pigs show no interest in the environment (Guy et al., 2002; Ruiterkamp et al., 1987), was usually seen in intensive conditions while wallowing into puddles and walking long distances were examples of “other” behaviours often observed in extensive conditions. Depending on the stimulation of the pigs in doing such behaviour, their motivation in exploring may change. In consequence, the occurrence of exploratory behaviour among extensive pigs is more likely to depend on their feeding motivation as well as on their motivation in performing other *a priori* “satisfying” behaviours. In contrast, a low level of exploration among intensive pigs has been mostly associated to the lack or deterioration of suitable manipulative substrate such as clean and renewed straw or other enrichment material (Beattie et al., 1995; Fraser et al., 1991). Therefore, a decrease in the frequency of exploration in extensive conditions is not always associated with an impaired welfare while it should be considered as so in intensive conditions.

Human-Animal Relationship

A poor human-animal relationship results in the animals being fearful of the stockpersons and other persons. Fear is considered as a major welfare problem as the sudden, intense or prolonged elicitation of fear can seriously damage the welfare, productivity, product quality and profitability of the farm animals (Waiblinger et al., 2006). The quality of stockmanship is the main determinant of a good or bad human-animal relationship. Stockmanship covers the way that animals are handled, the quality of their daily management and health care, and how well problems other than disease are recognised and solved (Waiblingler and Spoolder, 2007). The HAR test from the Welfare Quality® protocol is considered to be relevant to identify especially obvious poor human-animal relationships (Courboulay et al., 2009). In extensive management systems, animals often depend on the natural resources and are more likely to receive contacts from humans during infrequent but aversive procedures (Goddard et al., 2006; Turner and Dwyer, 2007), resulting in animals being more fearful to the human presence. Several studies have found a significant effect of the rearing system on the response of sheep (Hargreaves and Hutson, 1990) or heifers (Boivin et al., 1994) to humans. Nonetheless, in the present assessment, the panic response among groups of Iberian pigs was not significantly associated to the rearing system. Several environmental factors may cause Iberian pigs to be less fearful than expected in extensive conditions. First of all, extensive Iberian pigs received a daily food supplementation provided manually by the farmer. According to Hemsworth et al. (1996) pigs associate a rewarding experience of feeding with the handler and this conditioning result in pigs being less fearful of the handler and other humans.

Moreover, pigs were reared in conventional housings and management until they reach a sufficiently high body weight (30-40 kg) to get to outdoor paddocks. Long term effects of handling have been reported in young pigs (Hemsworth et al., 1987) and may influence their fear response to humans when getting older.

Although mean frequencies of the panic response did not depend significantly on the rearing system, there were marked differences between extensive and intensive units in the degree to which the animals were fearful of humans. Indeed, when a panic response was seen on an extensive unit, from 50% to 100% of the groups of animals assessed displayed a fear response to the observer. On the contrary, this response affected less than half of the pens assessed when seen on an intensive farm. Thereafter, the fear response tended to be consistent among groups of pigs from an extensive unit while it presented much higher intra-farm variability in intensive conditions. The HAR measure may be more sensitive in detecting animals' degree of fear of people among pigs in extensive conditions. In those conditions, pigs can easily flee from the observer whereas the escape response may be more ambiguous in intensive conditions. Animals that routinely receive frequent human contact are more likely to appear to ignore the stimulus person (Waiblinger et al., 2006) while animals that receive less direct contact, such as extensively reared pigs, are more likely to show a clear response to the observer presence. The degree of curiosity of the animals (de Passillé and Rushen, 2005) and passive avoidance behavioural pattern (Erhard et al., 1999) can also affect the animals' responses to people. These factors might be even more pronounced in more restraint conditions.

Qualitative behaviour assessment

On the basis of the first principal component (PC1), the qualitative behaviour assessment appear to be useful to distinguish between farms in which animals appeared to be in a more positive mood from farms in which animals showed a negative mood. The second principal component (PC2) by itself does not bear direct relevance to welfare, but it contributes to a meaningful transition between descriptive terms of positive and negative mood on the first PC. These observations corroborate the results found by Wemelsfelder and Millard (2009) when validating this methodology in growing pigs from other breeds. Extensive unit scores on PC1 were higher than those of intensive farms demonstrating that the QBA and subsequent analysis by means of a PCA appears to be a useful methodology to distinguish farms on the basis of expression of natural behaviour.

When applying the PCA only in Iberian pigs kept in intensive conditions, pigs' groups presented a homogeneous dispersion throughout the PC axes due to a high variability of the QBA among intensive farms. This variability makes the PCA sufficiently robust to distinguish between intensive farm scores on the first PC. In contrast, the PCA loses robustness when applied to the QBA from Iberian pigs in extensive conditions due to a high similarity between PC scores among extensive units. The variability between scores of intensive farms may be linked to the great diversity of housings in current Iberian intensive farming that reflect different levels of intensification. A first type of intensive farms raised Iberian pigs on concrete floor in facilities that resemble those of intensive “white” pigs in Spain until they get to the finishing phase (three last months of fattening). Iberian pigs then finalized their fattening period in outdoor paddocks with a greater space allowance (1.3 m^2 per pig). A second type of farms housed pigs indoor on straw or sand with an average space allowance of 1.5 m^2 per pig. Finally, a third type followed a semi-intensive rearing system providing indoor feeding and resting zones on straw or sand as well as an outdoor access during the whole growing period. On those farms the average space allowance was 3.2 m^2 per pig. On the one hand, housing conditions may have an effect on the observer’s perception of welfare that would systematically associate less intensively kept pigs to a more positive mood. Wemelsfelder et al. (2009) have shown that qualitative behaviour assessments in pigs are sensitive to contextual bias, but that this sensitivity does not appear to weaken the basic reliability of such assessment. On the other hand, housing and management can also have a real effect on the animal’s ongoing behaviour. It has indeed been suggested that pigs housed in enriched conditions show a different behavioural expression when observed in an experimental pen than pigs housed in non-enriched conditions (Wemelsfelder et al., 2000). Iberian pigs have been selected for over 2000 years for extensive production (Aparicio Tovar and Vargas Giraldo, 2006). They are thus confronted to aversive situations in intensive conditions that are notably different from those they find in their natural environment, resulting in the expression of a variety of behavioural responses in order to cope with aversive situations perceived in intensive environments (Wechsler, 1995). For example, animals may respond acting upon the aversive stimulus (Korte et al., 1992), increasing aggressive behaviour in situations of frustration (Duncan and Wood-Gush, 1971), showing an apathetic state (Fraser, 1975) or increasing their total activity and exploration directed at pen structures or pen mate (Scheepens et al., 1991). In consequence, pig’s behavioural expression may vary according to how it interacts and pays attention to a given situation. Subsequently, the QBA would be sufficiently sensitive to detect changes in the animals’ behavioural expression in a given situation and would be useful to give a perception of how well the animals are coping within the conditions of a particular farm.

CONCLUSIONS

Negative and "positive" social behaviours as well as the qualitative behaviour assessment are sensitive to changes in the housing and management of Iberian pigs. On the other side, no differences are seen between extensive and intensive reared Iberian pigs in the expression of other behaviours, such as exploratory behaviour, and in human-animal relationship. The high occurrences of negative social behaviour recorded in intensive conditions are clearly an indicator of poor welfare. However, interpretations of variations in the frequencies of "positive" social behaviour and exploration are not straightforward. Despite its subjectivity, the qualitative behaviour assessment appeared useful to discriminate farms on the basis of the expression of behaviour. Even though Iberian pigs in extensive conditions are more likely to cope with their environment and to perform normal patterns of behaviours, this is not clearly reflected by the results from the "Appropriate behaviour" assessment. Still, in overall, they appear to be in a more positive emotional state than their conspecific in more intensive rearing systems. Finally, since "Appropriate behaviour" is part of the integrative scoring system of the Welfare Quality® project, interpretation of variation in frequencies of behaviours in terms of animal welfare must be cautious, especially when comparing farm units from diverse rearing systems.

ACKNOWLEDGEMENTS

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. The authors are grateful to the farmers, veterinarians and associations of Iberian pigs for their cordial collaboration. Déborah Temple was supported by an FPU scholarship from the Ministerio de Ciencia e Innovación of Spain.

REFERENCES

Aparicio Tovar, M.A., Vargas Giraldo, J.D., 2006. Considerations on ethics and animal welfare in extensive pig production: Breeding and fattening Iberian pigs. Livest.. Sci. 103, 237–242.

Appleby, M.C., Wood-Gush, D.G.M., 1988. Effect of earth as an additional stimulus on the behaviour of confined piglets. Behav. Processes 17, 83–91.

Beattie, V.E.,Walker, N., Sneddon, I.A., 1995. Effects of environmental enrichment on behaviour and productivity of growing pigs. Anim. Welf. 4, 207–220.

Blackshaw, J.K., Thomas, F.J., Lee, J., 1997. The effect of a fixed or free toy on the growth rate and aggressive behaviour of weaned pigs and the influence of hierarchy on initial investigation of the toys. Appl. Anim. Behav. Sci. 53, 203–212.

Boissy, A., Manteuffel, G., Jensen, M.B., Oppermann Moe, R., Spruijt, B., Keeling,L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. Physiol. Behav. 92, 375-397.

Boivin, P., Le Neindre, J.P., Garel, J.M., Chupin, J.M., 1994. Influence of breed and rearing management on cattle reactions during human handling. Appl. Anim. Behav. Sci. 39, 115-122.

Botreau, R., Veissier, I., Butterworth, A., Bracke, M.B.M., Keeling, L.J., 2007. Definition of criteria for overall assessment of animal welfare. Anim. Welf. 16, 225-228.

Bracke, M.B.M., Zonderland, J.J., Lenskens, P., Schouten, W.G.P., Vermeer, H., Spoolder, H.A.M., Hendriks, H.J.M., Hopster, H., 2006. Formalised review of environmental enrichment for pigs in relation to political decision making. Appl. Anim. Behav. Sci. 98, 165–182.

Copado, F., de Aluja A.S., Mayagoitia L., Galindo, F., 2004. The behaviour of free ranging pigs in the Mexican tropics and its relationships with human faeces consumption. Appl. Anim. Behav. Sci. 88, 243–252.

Courboulay, V., Foubert, C., 2007. Testing different methods to evaluate pig welfare on farm. Anim. Welf. 16, 193-196.

Courboulay, V., Meunier-Salaün, M.C., Edwards, S.A.., Guy, J.H., Scott, K. 2009. Human-Animal relationships: validation of measures on fattening pigs, in: Forkman, B., Keeling, L. (Eds.), Assessment of animal welfare measures for sows, piglets and fattening pigs. Slu service, Uppsala, pp. 177-184.

Dawkins, M.S., 2004. Using behaviour to assess animal welfare. Anim. Welf. 13, 3-7.

De Jong, I.C., Ekkel, E.D., van de Burgwal, J.A., Lambooij, E., Korte, S.E., Ruis, M.A.W., Koolhaas, J.M., Blokhuis, H.J., 1998. Effects of strawbedding on physiological responses to stressors and behavior in growing pigs. Physiol. Behav. 64, 303-310.

Dellmeier, G.R., 1989. Motivation in relation to the welfare of enclosed livestock. Appl. Anim. Behav. Sci. 22, 129-138.

de Passillé, A.M., Rushen, J., 2005. Can we measure human–animal interactions in on-farm animal welfare assessment? Some unresolved issues. Appl. Anim. Behav. Sci. 92, 193–209.

Docking, C.M., Van de Weerd, H.A., Day, J.E.L., Edwards, S.A., 2008. The influence of age on the use of potential enrichment objects and synchronisation of behaviour of pigs. Appl. Anim. Behav. Sci. 110, 244–257.

Duncan, I.J.H., Wood-Gush, D.G.M., 1971. Frustration and aggression in the domestic fowl. Anim. Behav. 19, 500-504.

Erhard., H.W., Mendl, M., Christiansen, S.B., 1999. Individual differences in tonic immobility may reflect behavioural strategies. Appl. Anim. Behav. Sci. 64, 31–46.

Ewbank, R., Bryant, M.J., 1972. Aggressive behaviour amongst groups of domesticated pigs kept at various stocking rates. Anim. Behav. 20, 21-28.

Farm Animal Welfare Council, 1992. FAWC updates the five freedoms. Vet. Rec. 17, 357.

Fraser, D., 1975. The effect of straw on the behaviour of sows in tether stalls. Anim. Prod. 21, 59-68.

Fraser, D., Broom, D.M., 1990. Abnormal behaviour, in: Fraser, D., Broom, D.M. (Eds.), *Farm Animal Behaviour and Welfare.* Baillière Tindall, London, pp. 305-334.

Fraser, D., Phillips, P., Thompson, B.K.,Tennessen, T., 1991. Effect of straw on the behaviour of growing pigs. Appl. Anim. Behav. Sci. 30, 307–318.

Goddard, P., Waterhouse, T., Dwyer C., Stott A., 2006. The perception of the welfare of sheep in extensive systems. Small Rumin. Res. 62, 215–225.

Gonyou, H.W., 2001. The social behaviour of pigs, in: Keeling, L.J., Gonyou, H.W., *The social behaviour in farm animals.* Cabi Publishing, Wallingford, pp. 147-176.

Guy, J.H., Rowlinson, P., Chadwick, J.P., Ellis, M., 2002. Behaviour of two genotypes of growing-finishing pig in three different housing systems. Appl. Anim. Behav. Sci. 75, 193-206.

Hargreaves, A.L., Huston, G.D., 1990. Some effects of repeated handling on stress responses in sheep. Appl. Anim. Behav. Sci. 26, 253-256.

Hemsworth, P.H., Barnett, J.L., Hensen, C., 1987. The influence of inconsistent handling by humans on the behaviour, growth and corticosteroids of young pigs. Appl. Anim. Behav. Sci. 17, 245-252.

Hemsworth, P.H., Verge, J., Coleman, G.J., 1996. Conditioned approach-avoidance responses to humans: the ability of pigs to associate feeding and aversive social experiences in the presence of humans with humans. Appl. Anim. Behav. Sci. 50, 71–82.

Kerr, S.G.C., Wood-Gush, D.G.M., 1987. A comparison of the early behaviour of intensively and extensively reared calves. Anim. Prod. 45, 181-190.

Kikusui, T., Winslow, J.T., Mori, Y., 2006. Social buffering: relief from stress and anxiety. Phil. Trans. R. Soc. B. 361, 2215–2228.

Korte, S.M., Bouws, G.A.H., Koolhaas, J.M., Bohus, B., 1992. Neuroendocrine and behavioral responses during conditioned active and passive behavior in the defensive burying/probe avoidance paradigm: effects of Isapirone. Physiol. Behav. 52, 355-361.

Krohn, C.C., 1994. Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments. III. Grooming, exploration and abnormal behaviour. Appl. Anim. Behav. Sci. 42, 73-86.

Moinard, C., Mendl, M., Nicol, C.J., Green, L.E., 2003. A case control study of on farm risk factors for tail biting in pigs. Appl. Anim. Behav. Sci. 81, 333-355.

Mounier, L., Veissier, I., Andanson, S., Delval, E., Boissy, A., 2006. Mixing at the beginning of fattening moderates social buffering in beef bulls. Appl. Anim. Behav. Sci. 96, 185-200.

Mülleder, C., Palme, R., Menke, C., Waiblinger, S., 2003. Individual differences in behaviour and in adrenocortical activity in beef-suckler cows. Appl. Anim. Behav. Sci. 84, 167–183.

Panksepp, J., 2006. The neurobiology of positive emotions. Neurosci. Biobehav. Rev. 30, 173–87.

Petersen, V., Simonsen, H.B., Lawson, L.G., 1995. The effect of environmental stimulation on the development of behaviour in pigs. Appl. Anim. Behav. Sci. 45, 215–224.

Porter, R.H., Nowak, R., Orgeur, P., 1995. Influence of a conspecific agemate on distress bleating by lambs. Appl. Anim. Behav. Sci. 45, 239–244.

Rencher, A.C., 2002. Principal Component Analysis, in: Rencher, A.C. (Eds.), *Methods of Multivariate Analysis.* John Wiley and sons, New York, pp. 380-405.

Rodríguez-Estévez, V., García, A., Peña, F., Gómez, A.G., 2009. Foraging of Iberian fattening pigs grazing natural pasture in the dehesa. Livest. Sci. 120, 135-143.

Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Appl. Anim. Behav. Sci. 101, 40–53.

Ruiterkamp, W.A., 1987. The behaviour of growing pigs in relation to housing. Neth. J. Agri. Sci. 35, 67-70.

Rushen, J., de Passillé, A.M., 2009. The scientific basis of animal welfare indicators, in: Smulders, F.J.M., Algers, B. (Eds.), Food safety assurance and veterinary public health Volume 5. Welfare of production animals : assessment and management of risks. Wageningen Academic Publishers, The Netherlands, pp. 391-415.

Sato, S., Tarumizu, K., Hatae, K., 1993. The influence of social factors on allogrooming in cows. Appl. Anim. Behav. Sci. 38, 235-244.

Scheepens, C.J.M., Hessing, M.J.C., Laarakker, E., Schouten, W.G.P., Tielen, M.J.M., 1991. Influences of intermittent daily draught on the behaviour of weaned pigs. Appl. Anim. Behav. Sci. 31, 69-82.

Shimmura, T., Suzuki, T., Azuma, T., Hirahara, S., Eguchi, Y., Uetake, K., Tanaka, T., 2008. Form but not frequency of beak use by hens is changed by housing system. Appl. Anim. Behav. Sci. 115, 44–54.

Stolba, A., Wood-Gush, D.G.M., 1989. The behaviour of pigs in a semi-natural environment. Anim. Prod. 48, 419–425.

Taylor, G.T., 1981. Fear and affiliation in domesticated male rats. J. Comp. Physiol. Psychol. 95, 685–693.

Temple, D., Dalmau, A., Ruíz de la Torre, J.L., Manteca, X., Velarde, A., in press. Application of the Welfare Quality® protocol on intensively kept growing pigs in Spain. J. Vet. Beh.

Turner, S.P., Dwyer, C.M., 2007. Welfare assessment in extensive animal production systems: Challenges and opportunities. Anim. Welf. 16, 189-192.

Van Putten, G., Dammers, J., 1976. A comparative study of the well-being of piglets reared conventionally and in cages. Appl. Anim. Behav. Sci. 2, 339-356.

Waiblinger, S., Spoolder, H.A.M., 2007. Quality of stockmanship, in: Velarde, A., Geers, R. (Eds.), On farm monitoring of pig welfare. Wageningen Academic Publishers, The Netherlands, pp. 156-166.

Waiblinger, S., Boivin, X., Perdersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K., Jones, R.B., 2006. Assessing the human animal relationship in farmed species : a critical review. Appl. Anim. Behav. Sci. 101, 185–242.

Webster, A.J.F., Main, D.C.J., Whay, H.R., 2004. Welfare assessment: Indices from clinical observation. Anim. Welf. 13, 98.

Wechsler, B., 1995. Coping and coping strategies: a behavioural view. Appl. Anim. Behav. Sci. 43, 123–134.

Welfare Quality®, 2009. Welfare Quality® applied to growing and finishing pigs, in: Dalmau, A., Velarde, A., Scott, K., Edwards, S., Veissier, I., Keeling, L., Butterworth, A. (Eds.), Welfare Quality® Assessment protocol for pigs. Welfare Quality® Consortium, The Netherlands.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25-31.

Wemelsfelder, F., 2008. Curiosity, sentience, integrity: why recognising the "whole animal" matters, in: Boyle, L., O'Connell, N., Hanlon, A. (Eds), Proceedings of the 42nd Congress for Applied Ethology. Wageningen Academic Publishers, The Netherlands, pp. 1

Wemelsfelder, F., Millard, F., 2009. Qualitative Behaviour assessment, in: Forkman, B., Keeling, L. (Eds.), Assessment of animal welfare measures for sows, piglets and fattening, Slu service, The Netherlands, pp. 213-219.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. Appl. Anim. Behav. Sci. 67, 193–215.

Wemelsfelder, F., Hunter, E. A., Lawrence, A. B., Mendl, M. T., 2001. Assessing the 'whole-animal': a free-choice-profiling approach. Anim. Behav. 62, 209–220.

Wemelsfelder, F., Nevison, I., Lawrence, A.B., 2009. The effect of perceived environmental background on qualitative assessments of pig behaviour. Anim. Behav. 78, 477-484.

Wiepkema, P.R., Broom, D.M., Duncan, I.J.H., Van Putten, G., 1983. Abnormal Behaviours in Farm Animals, in: C.E.C. Report., Brussels, pp. 16.

Wood-Gush, D.G.M., Vestergaard, K., 1993. Inquisitive exploration in pigs. Anim. Behav. 45, 185-187.

CHAPTER 5

ASSESSMENT OF TEST-RETEST REPEATABILITY OF ANIMAL-BASED MEASURES ON GROWING PIG FARMS

Based on Temple D, Manteca X, Dalmau A, Velarde A *Submitted*. Assessment of test-retest repeatability of animal-based measures on growing pig farms. *Livestock Science.*

ABSTRACT

Measures taken on animals are subject to variations and their repeatability may represent one of their main limitations when used in a welfare assessment tool. Test-retest repeatability refers to the likelihood that the same results will be obtained if the assessment is repeated by the same observer. This study discusses the test-retest repeatability of quantitative and qualitative animal-based measures included into the Welfare Quality® protocol on a sample of 15 intensive conventional farms of growing pigs. An average of 12 months elapsed between the two visits and no changes in management routines or housing conditions were done by the farmers during this interval. A Wilcoxon signed rank test was used to test whether the mean results obtained during the two visits were significantly different. Different methods were used to assess repeatability of mean prevalences at farm level: Spearman's correlation coefficient (rs), intra class correlation coefficient (ICC) and limits of agreement (LoA). Moderate bursitis, widespread skin discolouration and positive social behaviour presented mean results significantly different between the two visits ($P < 0.05$). From a discriminative point of view, only moderately soiled body (rs = 0.7; ICC = 0.8) and the panic response (rs = 0.7; ICC = 0.5) were consistent over a long period of time. However, inconsistencies between Spearman's correlation and ICC were found when interpreting the repeatability of several quantitative measures such as severely wounded animal and severely soiled body. The Qualitative Behaviour Assessment (QBA) was subjected to a significant effect of the visit, however, the two sets of scores were moderately related on the first axis of the principal component analysis (rs = 0.5). A more evaluative approach such as the limits of agreement can be useful to guide the interpretation of test-retest repeatability of quantitative measures based on farm averages and to detect farms that present a persistent welfare problem. Unless the assessment relates only to the relative ranking of a farm within the same visit, the QBA is not consistent over a long period of time.

Keywords: test-retest repeatability; animal welfare assessment; pig; Welfare Quality®; animal-based measures

INTRODUCTION

Consistency of a welfare assessment system over time is especially important when it is to be used for certification purposes. In order to be cost effective, the Welfare Quality® assessments should take place in long intervals of supposedly more than six months (Knierim and Winckler, 2009). The frequency of visits will then be influenced by how often welfare can feasibly be assessed. Furthermore, if the main objective is to promote improvement, then tracking change over time becomes fundamental (Keeling, 2011). This should be done considering the repeatability of each farm individually. The assessment of repeatability becomes fundamental to detect the risk of poor welfare at some point in the future on the individual farm. Regardless of whether the protocol is used as a certification tool or to promote improvement, the outcomes should be representative of the long-term situation of a farm detecting significant changes and being not sensitive to minor changes included within the usual variation in farm conditions (Plesch et al., 2010). Repeatability will depend on the nature, distribution and sensitivity of each measure along time.

Repeatability is defined as the similarity in the results of repeated measurements on the same item (e.g. farm or group of animals). Concerning on-farm welfare assessment, repeatability generally has two components: inter-observer and test-retest repeatability (de Passillé and Rushen, 2005). High inter-observer repeatability is achieved when two (or more) persons carrying out an observation of the same animals get similar scores. Test-retest repeatability refers to the chance that the same results will be obtained if the test is repeated. In the special case of welfare assessment, the intra-observer repeatability - which describes the extent to which a single assessor obtains consistent results when evaluating the same animals or farms on different occasions – is part of test-retest repeatability. Measures taken on animals tend to be subject to variations and their repeatability may represent one of the main limitations of the use of animal-based measures in a welfare assessment tool (de Rosa et al., 2003; Lensink et al., 2003). For that reason, the majority of parameters included into the Welfare Quality® protocols were evaluated for their inter-observer repeatability at individual level and under experimental conditions (Forkman and Keeling, 2009).

Animal welfare is an inherent characteristic of the individual animal and has a dynamic outcome; therefore, it can not be seen as a steady state. The lack of reliability of a measure over time may be due to: 1) an unpredictable time effect; 2) the sensitivity of a parameter to routine (not less important) changes but not to major changes in the management/resources of the farm; 3) methodological constrictions. Those three major causes of variability between visits are more ore less likely to happen depending on the measure assessed.

A variety of different methodologies have been used to determine repeatability. In animal welfare science, repeatability of continuous variable is often expressed as a Spearman/Pearson's correlation coefficient (Winckler et al., 2007; Dalmau et al., 2010). In studies where three o more observers measure the same parameter (Wemelsfelder and Millard, 2009; Mullan et al., 2011) or where consistency over time is assessed more than twice (de Rosa et al., 2003; Plesch et al., 2010), Kendall's coefficient of concordance have been calculated to quantify the overall reliability. The consistency of a test measurement can also be quantified by means of intra class correlation coefficients (ICC). Despite its infrequent use in animal welfare studies, ICC is a common metric of repeatability of continuous measurements in medical sciences and clinical epidemiology (Weir, 2005; Dohoo et al., 2009). ICC assesses repeatability by comparing the variance of different measurements of the same subject (e.g. farm) with the total variance across all measurements and subjects (Bartko, 1966). Correlation coefficients (both Spearman/Pearson and ICC) are measures of reliability meaning that they evaluate the degree to which subjects (e.g. farms) can be distinguished from each other, despite the measurement error (Terwee et al., 2007). The main limitation of these methods resides, then, in their strong dependence on the variance of the assessed population of farms (Costa-Santos et al., 2011). More recently, the limits of agreement described by Bland and Altman (1986) have come into vogue to assess test-retest repeatability in the biomedical studies (Weir, 2005). Agreement assesses how close results of the repeated measurements are, by estimating the measurement error. Limits of agreement are, then, not influenced by the variance of the assessed population (de Vet et al., 2006). Even though interpretation of the significance of its results is highly subjective (Costa-Santos et al., 2011), limits of agreement may be useful to compare results of overall prevalences of welfare outcomes at farm level.

The interpretation of repeatability of continuous variables should be performed with caution as inconsistencies between methodologies have been observed in different studies (Costa-Santos et al., 2011). To overcome this problem the use of both measures of reliability and agreement is recommended (de Vet el al., 2006; Dohoo et al., 2009). This study evaluates the test-retest repeatability over a long period of time of the animal-based measures from the Welfare Quality® assessment protocol on a sample of 15 intensive farms. Different methodologies are used to assess test-retest repeatability of the welfare outcomes and their appropriateness is discussed especially when applied as a tool to track changes over time at the individual farm level.

MATERIAL AND METHODS

Farm sample and data collection

The Welfare Quality® protocol was applied twice by the same assessor on 15 intensive growing pig farms in Spain. An average of 12 months (SD = 40 days) elapsed between the two visits to minimize a seasonal effect. No major changes in management routines and housing conditions were undergone by the farmers during this interval. According to the farmers and managers, the animals in both visits came from the same origin however a similar lineage was not assured. All the farms assessed followed an "all in all out" management and their size ranged from 330 to 4800 pigs per farm. Pigs were fed *ad libitum* and kept indoor on fully or partly slatted floors in groups of an average of 14 (SD = 5) animals. During the first visit, the age of the pigs ranged from 63 to 182 days old and the average space allowance was 0.77 m^2 per pig (SD = 0.2 m^2 per pig). During the second visit, the age of the pigs ranged from 43 to 171 days old and the average space allowance was 0.77 m^2 per pig (SD = 0.2 m^2 per pig).

The quantitative animal-based parameters evaluated in this study are summarised in **Table 1**. Measures of feeding, housing and health were scored at pen or individual level according to a three-point scale ranging from 0 (good welfare) to 2 (poor welfare). The majority of measures – except coughs and sneezes – were assessed on animals in 10 randomly selected pens throughout the farm. Coughs and sneezes were evaluated at 6 different points of the farm. Quantitative behavioural measures were carried out by means of scan sampling of social and exploratory behaviours and a human-animal relationship (HAR) test. The scan sampling, validated by Courboulay and Foubert (2007), was performed at 3 different observation points of the farm. Approximately, 50-60 animals were observed from each observation point. For the HAR test, validated by Courboulay and Foubert (2007), 10 selected pens were tested throughout the farm. Each pen was assessed as a whole and a score of 0 (no panic response) or 2 (more than 60% of the animals in the pen showing a panic response) was given at pen level.

Table 1 Quantitative animal-based measures with their respective scoring scale used in the welfare assessment (Welfare Quality®, 2009).

Animal-based measures	Score	Description
Body Condition	0	Animal with a good body condition
	2	Poor body condition: Animal with visible spine, hip and pin bones
Bursitis	0	No evidence of bursae
	1	Moderate bursitis: One or several small bursae on the same leg or one large bursa
	2	Severe bursitis: Several large bursae on the same leg or one extremely large bursa or any eroded bursae
Manure on the body	0	Less than 20% of the body surface is soiled
	1	Moderately soiled body: More than 20% but less than 50% of the body surface is soiled with faeces
	2	Severely soiled body: Over 50% of the body surface is soiled with faeces.
Huddling	0	Pig lying with less than half of its body lying on top of another pig
	2	Pig lying with more than half of its body lying on top of another pig
Panting	0	Normal breathing.
	2	Rapid breath in short gasps.
Shivering	0	No vibration of any body part
	2	Slow and irregular vibration of any body part, or the body as a whole
Wounds on the body	0	If all regions of the animal's body have up to 9 lesions
	2	Severely wounded: when more than 10 lesions are observed on at least two zones of the body or if any zone has more than 15 lesions
Tail biting	0	No evidence of tail biting; superficial biting but no evidence of fresh blood or of any swelling
	2	Bleeding tail and/or swollen infected tail lesion, and/or part of tail tissue missing and presence of crust
Lameness	0	Normal gait or difficulty in walking, but still using all legs; swagger of caudal body while walking; shortened stride
	1	Severely lame, minimum weight-bearing on the affected limb
	2	No weight-bearing on the affected limb, or not able to walk
Pumping	0	No evidence of laboured breathing
	2	Evidence of laboured breathing
Scouring	0	No liquid manure visible in the pen
	1	Areas in the pen with some liquid manure visible
	2	All faeces visible inside the pen is liquid manure
Skin Condition	0	No evidence of skin inflammation or discolouration
	1	Localized skin condition: More than zero, but less than 10% of the skin is inflamed, discoloured or spotted
	2	Widespread skin condition: More than 10% of the skin has an abnormal colour or texture.
Hernias	0	No hernia/rupture
	2	Hernias/ruptures with bleeding lesion or touching the floor.

Twisted snout	0 2	No evidence of twisted snouts Evidence of twisted snouts
Rectal prolapse	0 2	No evidence of rectal prolapse Evidence of rectal prolapse
HAR	0 2	No panic response to human presence Panic response: More than 60% o the animals fleeing, facing away from the observer or huddled in the corner of the pen
Negative social behaviour	% of active behaviour	Aggressive behaviour, including biting or any social behaviour with a response from the disturbed animal.
Positive social behaviour	% of active behaviour	Sniffing, nosing, licking and moving gently away from the animal without an aggressive or flight reaction from this individual
Explorative behaviour	% of active behaviour	Sniffing, nosing, licking all features of the pen or paddock. Exploration towards straw or other suitable enrichment material

The qualitative assessment of behaviour (QBA), validated by Wemelsfelder and Millard (2009), uses descriptive terms with an expressive connotation to reflect animals' experience of a situation (Wemelsfelder, 2007). A rating scale was used to score pigs at group level based on 20 different terms: 1.Active, 2.Relaxed, 3.Fearful, 4.Agitated, 5.Calm, 6.Content, 7.Tense, 8.Enjoying, 9.Frustrated, 10.Sociable, 11.Bored, 12.Playful, 13.Positively occupied, 14.Listless, 15.Lively, 16.Indifferent, 17.Irritable, 18.Aimless, 19.Happy and 20.Distressed. An overall assessment was carried out at farm level after the animals in six observation points per farm were observed.

In order to minimize intra-farm variability between the two visits, the same pens or observation points were evaluated in both visits by the same assessor. Hospital pens were not included in the sampling. More detailed information about the overall methodology of assessment can be found in the Welfare Quality® Assessment protocol for pigs (2009).

Statistical analysis

The SAS statistical package (SAS Inst. Inc., Cary, NC) was used to analyse data.
Data were analysed at farm level.

Quantitative measures analysis

Prevalences of animal-based measures of feeding, housing and health as well as the HAR test were expressed as proportion of animals or pens affected on the total of animals or pens evaluated. Social and exploratory behaviours were expressed in proportion to the total number of active pigs. A Wilcoxon signed rank test was used to test whether the mean results obtained during the two visits were significantly different. Three statistical methods were applied to evaluate the test-retest repeatability of quantitative measures. The Spearman's correlation coefficient (rs) and the intraclass correlation coefficient (ICC) were used to determine reliability whereas limits of agreements were used as an evaluative methodology. For each quantitative measure, partial Spearman's correlations between the two visits were calculated removing the effect of the age of the animals. A rs value of 0.4 – 0.7 indicates a moderate correlation and more than 0.7 indicates a high correlation (Martin and Bateson, 2007). ICC were evaluated using Generalised Linear Models (PROC GENMOD, CORRW option), controlling for the age of the animals. A negative binomial or poisson distribution was used on the basis of the value of the deviance (Cameron and Trivedi, 1998). The closer the ICC is to 1.0, the higher the reliability; and if the variability within groups (visits) is larger than the variability between groups (visits) then ICC will be negative (Zar, 1999). The limits of agreement between the two visits for each quantitative measure were calculated using the Bland and Altman (1986) method. Limits of agreement were estimated on the mean difference of prevalences between the two separated visits on the same farms and the standard deviation of these differences. Most of the difference is expected to be less than two standard deviations. The plots of the differences between the pairs of prevalences against their mean value were used to interpret the results.

Qualitative analysis

QBA data were expressed at farm level and were analysed using a Principal Component Analysis (PCA) by means of the SAS PRINCOMP procedure. A correlation matrix was used for a more balanced representation of the descriptive terms and no rotation was applied. The first two principal components whose eigenvalues were greater than 1.0 produced a two-dimensional interpretative word chart. In the chart, the 20 descriptive terms were plotted in relation to the first two axes. The word chart can be useful to lead the interpretation of the first two principal components. The eigenvectors (or loadings) from the descriptor data quantify the weight each descriptor has on the two main axes (Rencher, 2002). Each farm unit receives a score on each of the main principal component. Next, a Wilcoxon signed rank test was used to test whether the scores obtained during the first and the second farm visits were significantly different. Finally, partial Spearman's correlations between scores of farms from the two visits were calculated for each main principal component and removing the effect of the age of the animals.

RESULTS

Prevalences and occurences of quantitative measures

The mean prevalences and occurrences of the quantitative animal-based measures recorded during the two visits are shown in **Table 2**. Moderate bursitis, widespread skin discolouration/inflammation and positive social behaviour presented mean results significantly different between the two visits. The remaining quantitative animal-based measures did not present mean results significantly different between the two visits. Figure 1 described the mean prevalences of manure on more than 50% of the body (severely soiled body) and severely wounded animals recorded in each visit. The farms that presented the highest prevalences of severely soiled body and severely wounded animals in the first visit showed also the highest prevalences of those measures in the second visit.

Table 2 Mean prevalences/occurrences ± SD of quantitative animal-based measures recorded during the two visits using the Welfare Quality® protocol for growing pigs. Partial Spearman correlations coefficient, Intraclass correlation coefficient (ICC) and limits of agreement (mean differences between visits ± 2 SD) between two farm visits are shown for each quantitative animal-based measure.

Animal-based measure	First visit	Second visit	Wilcoxon signed rank test p value	Spearman Correlation between visits	ICC	Limits of agreement
Poor body condition	0.1 ± 0.32	0.1 ± 0.37	NS	- 0.19	- 0.19	0.0 ± 1.03
Panting	0.1 ± 0.27	0.0 ± 0.00	NS	-	-	0.0 ± 0.00
Shivering	0.0 ± 0.00	0.0 ± 0.00	NS	-	-	0.0 ± 0.00
Huddling	0.9 ± 1.30	1.4 ± 2.02	NS	0.49	0.45	0.5 ± 3.56
Moderate bursitis	61.2 ± 11.23	50.4 ± 13.9	0.015	0.22	- 0.20	-
Severe bursitis	6.9 ± 7.01	4.1 ± 3.66	NS	- 0.15	- 0.48	-2.9 ± 17.76
Moderately soiled body	10.9 ± 13.30	7.2 ± 7.30	NS	0.72*	0.83	3.7 ± 10.10
Severely soiled body	1.8 ± 3.10	1.2 ± 3.50	NS	0.09	0.72	0.7 ± 3.72
Severely wounded	0.9 ± 1.09	1.1 ± 2.62	NS	0.13	0.89	1.0 ± 3.96
Tail biting	0.7 ± 1.12	0.4 ± 1.16	NS	0.06	- 0.01	-0.2 ± 3.22
Lameness (total)	0.4 ± 0.77	0.2 ± 0.42	NS	0.28	- 0.04	-0.1 ± 1.69
Pumping	0.0 ± 0.00	0.0 ± 0.00	NS	-	-	0.0 ± 0.00
Snout	0.0 ± 0.00	0.0 ± 0.00	NS	-	-	0.0 ± 0.00
Prolapse	0.0 ± 0.00	0.04 ± 0.17	NS	-	-	0.0 ± 0.34
Hernia	0.1 ± 0.34	0.2 ± 0.35	NS	- 0.20	- 0.26	0.0 ± 1.02
Localized skin condition	3.0 ± 5.17	1.3 ± 1.12	NS	- 0.30	- 0.42	-1.6 ± 11.67
Widespread skin condition	5.8 ± 6.03	0.8 ± 1.75	0.002	0.15	0.44	-
Coughs	15.5 ± 24.41	20.4 ± 22.83	NS	- 0.10	- 0.10	4.9 ± 71.52
Sneezes	17.1 ± 21.76	13.3 ± 10.90	NS	0.17	0.25	-3.8 ± 41.81
Scouring scored as 1	5.3 ± 9.90	6.7 ± 9.00	NS	0.40	- 0.15	1.3 ± 24.22
Scouring scored as 2	0.0 ± 0.00	0.0 ± 0.00	NS	-	-	0.0 ± 0.00
Negative social	5.1 ± 2.73	5.6 ± 2.04	NS	0.06	- 0.30	-0.5 ± 7.50
Positive social	9.9 ± 2.81	13.6 ± 1.96	0.0006	0.19	0.29	-
Exploratory	27.3 ± 6.41	24.5 ± 5.97	NS	- 0.10	0.36	2.7 ± 16.08
"Other"	57.6 ± 5.17	56.3 ± 8.73	NS	- 0.29	- 0.10	1.3 ± 20.10
Panic response	9.3 ± 17.51	8.0 ± 12.07	NS	0.69*	0.51	1.3 ± 25.30

* P < 0.01
NS: non-significant

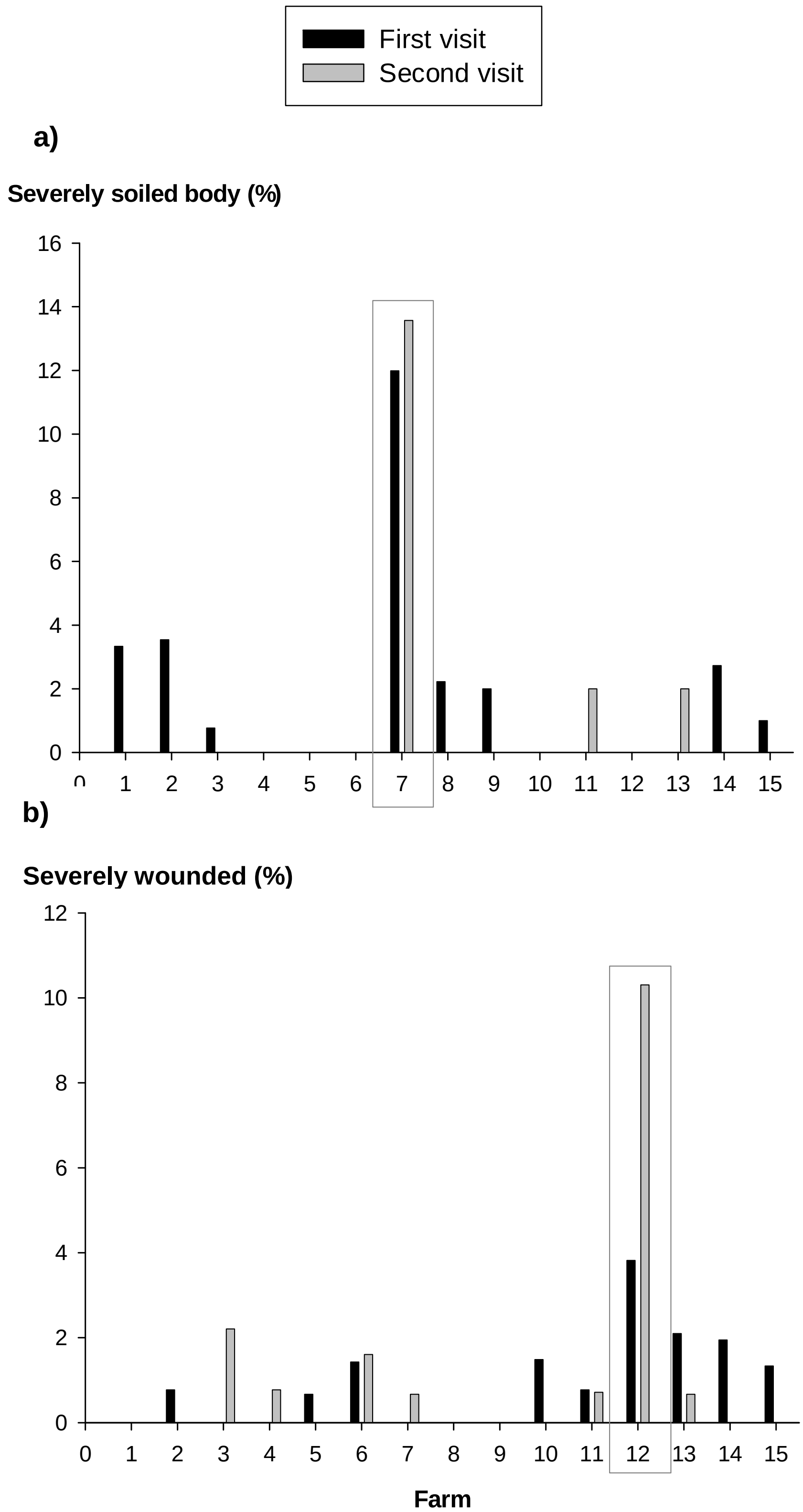


Figure 1 Distribution of the mean prevalences of manure on more than 50% of the body (severely soiled body, a) and wounds on the body (severely wounded animals) (b) in percentage of pigs affected per farm during the first visit (black) and the second visit (grey).

Test-retest repeatability of quantitative measures

Moderately soiled body (rs = 0.72) and the HAR test (rs = 0.69) presented the highest partial Spearman's correlation coefficient between the two visits. These two animal-based measures also presented acceptable ICC (ICC > 0.5). Based on Spearman's correlation coefficients, the other quantitative measures were not repeatable from one visit to the other (**Table 2**). According to ICC, severely soiled body and severely wounded animals presented a good reliability been in discrepancy with the results of Spearman's correlations.

The limits of agreement for the mean differences between visits of each quantitative animal based measure are shown in **Table 2**. The limits of agreement of moderate bursitis, widespread skin condition and positive social behaviour were not calculated as the mean difference was not significantly different from zero. Limits of agreement were plotted for four quantitative measures: severe bursitis, negative social behaviour, tail biting and scouring scored as 1. The scatter of the difference for severe bursitis increases as the combined mean increases (**Figure 2**). **Figure 3** shows the mean difference plots of negative social behaviour between the two visits. The range of difference of negative social behaviour varied from -3.5% to 2% between the two visits for 13 of the 15 farms assessed and a clear outlier could be identified. Tail biting (**Figure 4**) did not present any agreement as, except one farm, affected units were not affected twice. Finally, according to **Figure 5**, two of the six farms that presented diarrhoea on the first visit presented the same prevalences on the second visit, showing entire agreement. Also, nine farms that didn't present diarrhoea on the first visit did not present it on the second visit.

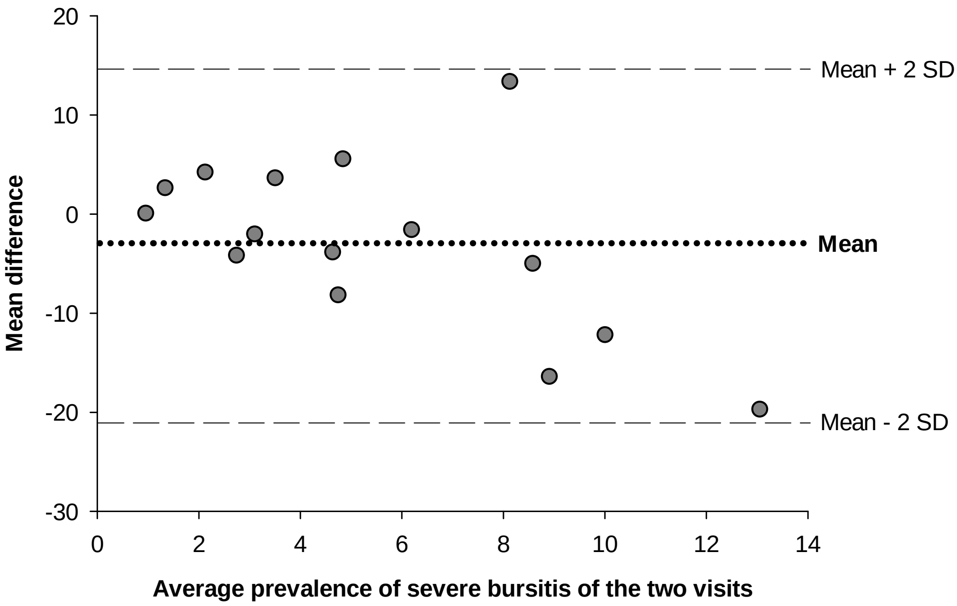


Figure 2 Plot of the mean difference (second visit minus first visit) against the mean prevalence of severe bursitis of the two visits for 15 intensive farms.

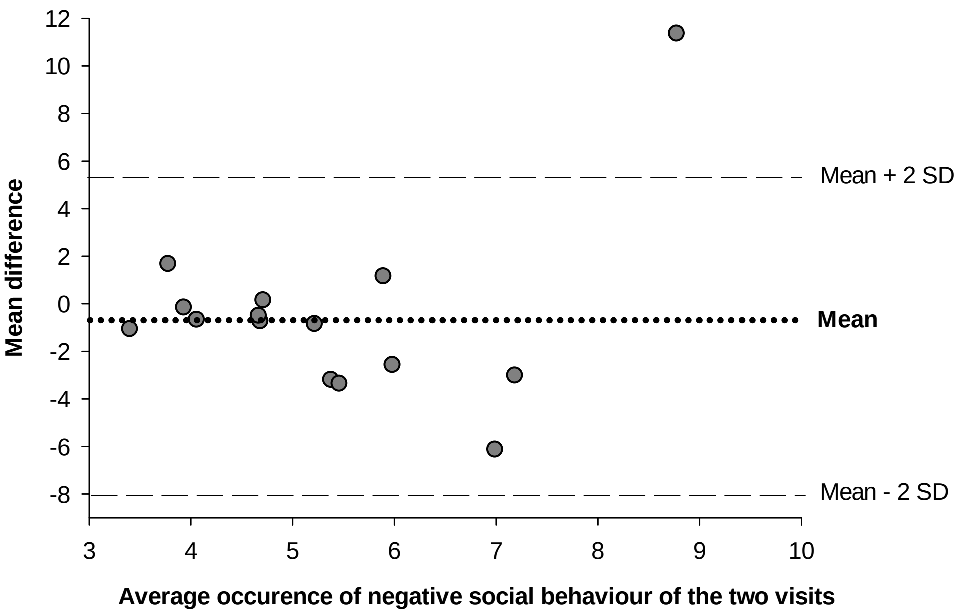
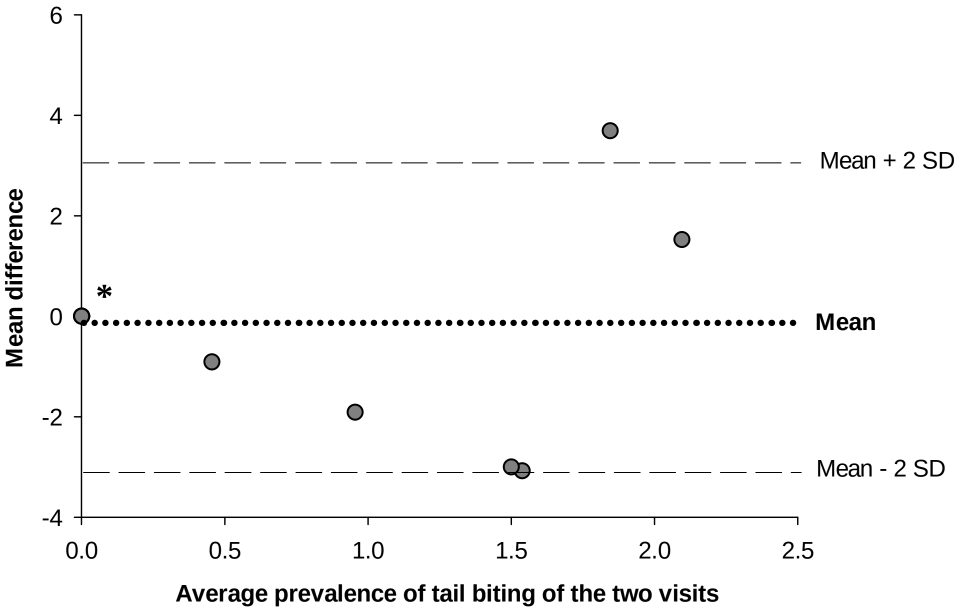
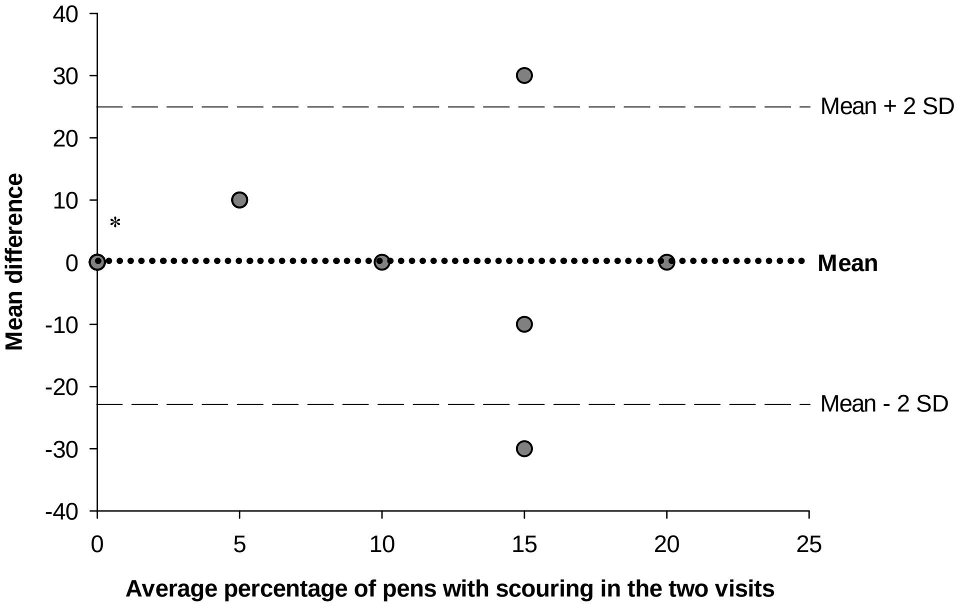


Figure 3 Plot of the mean difference (second visit minus first visit) against the mean occurrence of negative social behaviour of the two visits for 15 intensive farms.



* Nine farms with a 0% prevalence in the two visits are superposed.

Figure 4 Plot of the mean difference (second visit minus first visit) against the mean prevalence of tail biting of the two visits for 15 intensive farms.



* Nine farms with a 0% prevalence in the two visits are superposed.

Figure 5 Plot of the mean difference (second visit minus first visit) against the mean prevalence of pen with scouring scored as 1 on the two visits for 15 intensive farms.

Qualitative behaviour assessment

Two main axes were retained accounting for a cumulative percent of variability between farms of 52% (35% and 17%, respectively). The eigenvalues for these first two axes were greater than 1. The PCA word chart is presented in **Figure 6**. Relaxed (0.32), enjoying (0.31), positively occupied (0.30), and happy (0.29) as well as frustrated (-0.31), indifferent (-0.27) and irritable (-0.26) showed the highest loadings on the first axis. The second axis was characterised with terms such as playful (0.43), indifferent (0.28) as well as bored (-0.36), fearful (-0.27) and agitated (-0.26). The position of each farm on the basis of the QBA was plotted in **Figure 7**. The visit had a significant effect on the scores of the farms on the first axis (S = -59; $P < 0.0001$) and the second axis (S = -53; P = 0.0012). The animals observed during the second visit obtained significantly higher scores on the first axis, meaning that they were scored as more relaxed and enjoying. However, on the second axis, the scores from the second visit were lower, meaning that the animals were perceived more bored and fearful.

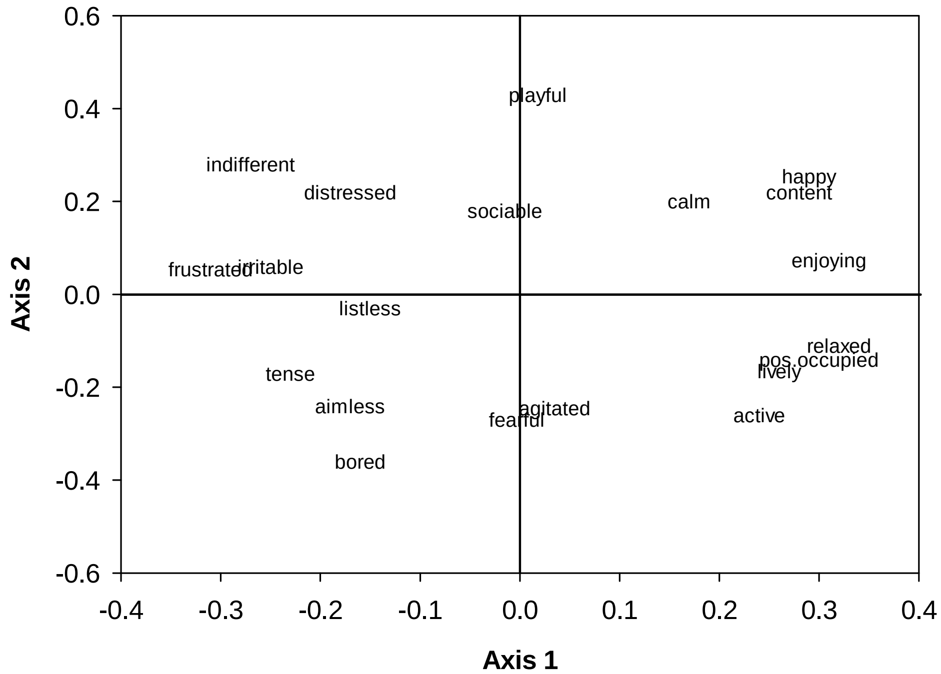


Figure 6 Word chart of the QBA assessed in intensively kept pigs during the first and the second farm visits. Axes reflect eigenvectors (loadings) of descriptive terms on the two principal PCA dimensions.

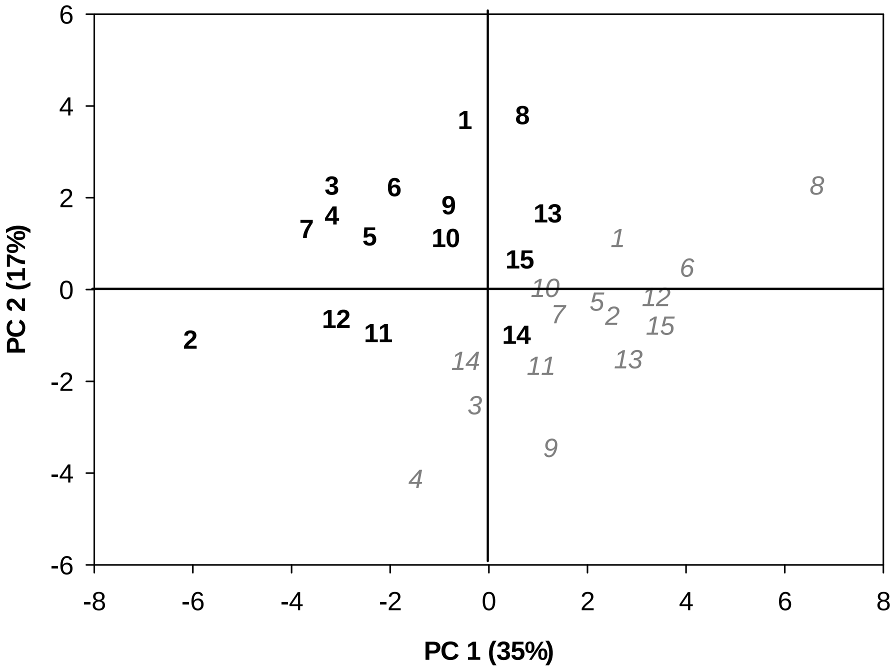


Figure 7 Plotted values of the first two main components for the farms assessed during the first (**bold numbers**) and the second visit (*cursive numbers*).

Test-retest repeatability of the QBA

The first principal component of the PCA from the QBA presented a moderate partial Spearman's correlation between the two visits (rs = 0.52) whereas the second principal component was less consistent (rs = 0.32). Limits of agreement were not calculated for the qualitative behaviour assessment due to the significant visit effect.

DISCUSSION

Quantitative measures

Several methodologies were used to assess the repeatability of each quantitative measure over time. Spearman's correlation coefficient measures the extent to which one set of test results varies with a second set; however, it does not directly compare the values obtained (Dohoo et al., 2009). The intraclass correlation coefficient (ICC) relates the amount of variability among units to the total variability which consists of variability among units plus variability among measurements within a same unit (between visits). It can be considered as one minus the proportion of variance due to measurement error (Dohoo et al., 2009). Both measures of reliability are more likely to be used for discriminative purposes and the measurement error is related to the variability between units (farms) in both cases (Bland and Altman, 1996; de Vet et al., 2006). Taking into account that the age of the animals was an influencing factor for many measures (Temple et al., 2012), Spearman's correlation coefficient and ICC were calculated taking into account the age of the animals.

From a discriminative point of view, only moderately soiled body and the human animal relationship test showed a high consistency over a long period of time. Fear of human and moderately soiled body presented the highest repeatability according to both reliability methodologies used. Concerning fear of human, Rousing and Waiblinger (2004) and Winckler et al. (2007) reported a good repeatability of the avoidance distance measure in cattle in intervals of four to five days and of two months, respectively. Fear of human beings is largely affected by the way stockpeople interact with their animals (Hemsworth et al., 1989). When no major changes in husbandry occurred, the HAR appeared to be a consistent measure over long period of time showing little dependence of minor environmental and management variations within the farms. Pig dirtiness depends on seasonal effects, cleanliness routines and farm's design such as the type of floor (Huynh et al., 2005) or the feeding system (Scott et al., 2007). Moderately soiled body appeared to be sensitive to differences between intensive farms in previous studies (Temple et al., 2011a). The high consistency of the measure moderately soiled body between visits shows that this sensitivity depends greatly on the housing and management rather than minor variations in the farm conditions. Severely soiled body showed low Spearman's correlations between visits but high ICC. Low Spearman's correlations are expected for a parameter such as severely soiled body which does not allow discrimination between farms (Temple et al., 2011a). One farm presented extreme prevalences of severely soiled body in both visits increasing the variance between units and thereafter producing a large ICC. The same contradictory results between both types of correlation coefficients were found for severely wounded animals and the opposite was observed for scouring.

Indeed, scouring presented a moderate Spearman's correlation but a low ICC, indicating that the pairs of prevalences were related but that the variability between the two sets of measurements was high. The former contradictions between Spearman's correlation and ICC highlight the dependence of Spearman's correlation coefficients and ICC on the variability between farms as commented by de Vet et al., (2006) and Bland and Altman (1996).

The remaining quantitative measures were not repeatable between both visits either using Spearman's correlations or ICC. From a discriminative point of view, to have a good reliability, the measurement error of the protocol should be small in comparison to the variability between farms (de Vet, 1998). Health data recorded by means of the Welfare Quality® assessment protocol showed low prevalences and little sensitivity to detect differences between intensive farms in a previous study (Temple et al., 2011a). Plesch et al. (2010) commented that the lower the incidence of a welfare outcome, the lower was the repeatability of results over time. In addition, the measurement error of a welfare assessment tool between two farm visits is expected to be relatively high. Any regular variations (but not major changes) in farm conditions as well as any unpredictable time effects can increase the measurement error. The inclusion of farms from different production systems in the sample will probably increase the correlation coefficient as it would increase the variability between farms; still, it does not mean that the consistency of prevalence on a given farm is higher. Consequently, for those peculiar data that show little sensitivity, the evaluation of the repeatability by means of Spearman's correlations or ICC does not seem to be robust enough and can lead to misinterpretations.

Agreement parameters are used for evaluative purposes (de Vet et al., 2006). If the welfare assessment system is considered from an evaluative point of view, the variability between farms in the population does not matter; only the measurement error of the assessment system is important. The limits of agreement's methodology proposed by Bland and Altman (1986) indicate the range of differences between the two sets of measures. The plot of the difference between the means of two visits against the average prevalence of the two visits helps to determine the range of errors, indicated by the spread of the points (de Vet, 1998). The mean difference between the two sets of prevalences is closed to zero for many parameters. The adoption of an evaluative approach can be complementary to guide the interpretation of the results from discriminative methods. For example, limits of agreement of severe bursitis were wider due to large values of the difference for high average prevalences. The plot is useful to determine if the level of disagreement between the two visits varies with the mean prevalences obtained at farm level. Severe bursitis is therefore not consistent over time for high prevalences. Older animals housed on concrete floors present a higher risk of bursitis (Gillman et al., 2008). The age of the animals was not taken into account when calculating and plotting the limits of agreement, and might explain part of this difference for high average prevalences.

Some parameters such as localized and widespread skin condition, respiratory and enteric disorders are more likely to be subject to unpredictable variations along time and sporadically affect a farm. Those indicators were actually difficult to predict by means of simple measurements on the environment (Temple et al., 2012). By nature, those parameters are expected to show a low repeatability. The outbreak of a particular disease may induce variations in the incidence of a given welfare measure on a particular month or year. Widespread skin discolouration can be a symptom of systemic infections such as PRRS and environmental diseases (such as mange) that may explain the significant differences in the mean prevalences between both visits. For those measures whose mean values did not differ significantly between both visits, attention should be given to farms that present one (or several) of those problems persistently. For example, many farms that didn't present diarrhoea on the first visit did not present it on the second visit. However, two of the six farms presented the same prevalence of diarrhoea during both visits, showing entire agreement. Those farms presented a persistent problem of diarrhoea and should therefore be considered as problematic units for that particular measure. Two other farms presented an outbreak of diarrhoea only in one of the two visits. Several environmental and management factors may increase the risk and duration of enteric disorders during the productive cycle, producing outbreaks of diarrhoea (Pearce, 1999). Routine management (but not "major" changes) can prevent a welfare problem from becoming persistent. The efficiency of the farmer/veterinarian in applying a correct treatment (vaccines), controlling the ventilation and keeping a good hygiene may prevent many respiratory and enteric disorders from becoming persistent. On intensive farms, tail biting is a multifactorial problem whose incidence is highly variable and often referred as "outbreak" (Hemsworth and Cronin, 2006; Taylor et al., 2010). The repeatability of such unpredictable welfare problem at farm level is expected to be very low and its persistency should be prevented by good management (e.g. hospital pens, environmental enrichment, etc.). In our study, except one farm, affected units were not affected twice either because of a complete absence of tail biting at the moment of the visit or due to proper management of sick animals. Conversely to tail biting, lameness and poor body condition do not usually appear as "outbreaks"; however, their occurrence may vary, as well, depending on management factors such as the correct use of hospital pens. Taking into account that no "major" housing (e.g. type of floor; mean densities) or nutritional changes occurred between the two visits, the lack of repeatability of lameness and poor body condition may show that the majority of farmers do, indeed, isolate sick animals as recommended. Routine changes on a farm affect repeatability of several measures even if "major" changes do not occurred. Therefore, adopting an evaluative approach such as the limits of agreement may be better to identify farms that have persistent welfare problems and thus higher risks of poor welfare.

Finally, methodological aspects should be considered when interpreting a lack of repeatability between visits. Some parameters may be too sensitive and their variation may be due to slight fluctuations in the environment or in the animals themselves. The range of difference of negative social behaviour varied from -3.5% to 2% between the two visits for 13 of the 15 farms assessed. Negative social behaviour appeared to be very sensitive to farm variations and its occurrence on intensive conventional farms ranged from 1% to 14% (Temple et al., 2011a). The majority of farms assessed in the present study showed, then, a homogeneous occurrence of negative social behaviour over the two visits and the range of variation can represent the usual variation of this behavioural measure when no major change occur on the farm. On the other side, two farms were not consistent over time due to an increase of aggressions during one of the two visits. Thereafter, from an evaluative point of view, negative social behaviour presented some agreement for the majority of farms. Other parameters that present high intra-farm variability might not be detected by the sampling in both visits (Temple et al., 2011a). The sample of animals assessed for tail biting or lameness might not be representative of the current welfare state of the overall group of animals present on the farm for those two measures. Intra-observer variability may represent a third methodological aspect to consider when interpreting test-retest repeatability. For instance, a low intra-observer repeatability linked to the higher training experience of the observer in the second visit may explain variations in the mean prevalence of positive social behaviour between both visits.

Qualitative assessment

The qualitative behaviour assessment (QBA) was analysed as an integrative measure of 20 descriptive terms. On the basis of the first principal component (PC1), the qualitative behaviour assessment distinguishes between farms with animals that appeared to be in a more positive mood from farms with animals that showed a more negative mood. The second principal component (PC2) differentiates these moods in low and high levels of arousal. These observations were in agreement with the clustering of descriptors at the positive and negative ends of the two principal components found by Wemelsfelder and Millard (2009) when studying the inter-observer reliability of the QBA. Similar PCA results were also reported by Temple et al. (2011b) when comparing Iberian pigs kept in intensive and extensive units. The PCA pointed out a clear visit effect in which the animals observed during the second visit appeared to be, in overall, in a more positive mood but also more bored and fearful than during the first visit. Besides this systematic difference, the two sets of scores were moderately related on the first axis using partial Spearman's correlations (r_s = 0.5). Therefore, unless the assessment relates only to the relative ranking of a farm within the same visit, it is not consistent over time. Wemelsfelder et al. (2001) reported good inter and intra-observer reliability using the "free choice profiling methodology" (FCP) on pigs observed individually. As well, Wemelsfelder and Millard (2009) found a good reliability of the QBA between observers using a fixed scale rating to assess groups of growing pigs. In both studies the reliability was tested on the basis of video clips of pigs (whether individual or groups) that were shown to the observers. In contrast, in the present study, the QBA was applied on pigs kept in commercial conditions and an interval of 12 months elapsed between the two visits. Variations in the physical or mental states of the animals might explain, in part, variations between visits and lower correlations than in experimental conditions. As no major changes (management, housing, and breed) occurred between the two farm visits, the QBA may be sensitive to minor variations in the pigs' general mood. Moreover, an effect of the mood of the observer (intra-observer effect) should not be dismissed as it may originate a systematic error between visits. The interpretation of what the terms mean may change depending on the experience of the observer over time. A previous training with videos displaying the full range of the species behaviour under diverse breeding conditions and a detailed description of the terms on animals kept in a wide range of housing conditions should control this possible intra-observer effect. In a recent study, Phythian et al. (2011) has found a relative stability of sheep farm scores throughout a year. Sheep farms are more likely to be different one from the other in husbandry and housing conditions. In contrast, the pig farms assessed in the present study were all intensive conventional farms. The QBA may gain stability over time when applied to sample of farms from a wide range of production systems. In this case, the measurement error due to possible seasonal effects or to an intra-observer effect for example would be lower than the variability between farms increasing the consistency of the QBA results. Finally, the QBA may provide a more stable picture of the animals on a farm over a long period of time and not only during a single visit. As commented by Meagher (2009), observer ratings of behaviour are, generally, intended to integrate information over a period of time and a variety of situations rather than reflect only a particular moment in time, and thus are expected to display stability. Thereafter, integrating information across visits, the QBA is more likely to be more reliable.

CONCLUSIONS

Moderate bursitis, widespread skin discolouration/inflammation and positive social behaviour presented mean results significantly different between the two visits. The remaining quantitative animal-based measures did not present mean results significantly different between the two visits. From a discriminative point of view only moderately soiled body and the human animal relationship test showed a high reliability over a long period of time. Inconsistencies between Spearman's correlation and ICC were found when interpreting the repeatability of several measures. Adopting a more evaluative approach such as the limits of agreement can be useful to guide the interpretation of test-retest repeatability of quantitative measures based on farm averages. It also allows the detection of farms that present a persistent welfare problem and can be practical to be used for improvement purposes. Unless the assessment relates only to the relative ranking of a farm within the same visit, the QBA scores from intensive conventional farms of pigs were not stable over time.

ACKNOWLEDGEMENTS

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. The authors are grateful to the farmers, veterinarians and pig associations for their cordial collaboration. Déborah Temple was supported by an FPU scholarship from the Ministerio de Educación.

REFERENCES

Bartko, J.J., 1966. The intraclass correlation coefficient as a measure of reliability. Psychol. Rep. 19, 3-11.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 327, 307 - 310.

Bland J.M., Altman, D.G., 1996. Statistical notes: measurement error and correlation coefficients. Br. Med. J. 313, 41-42.

Costa-Santos, C., Bernardes, J., Ayres-de-Campos D., Costa, A., Costa, C., 2011. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. J. Clin. Epidemiol. 64, 264-269.

Courboulay, V., Foubert, C., 2007. Testing different methods to evaluate pig welfare on farm. Anim. Welf. 16, 193-196.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Dalmau, A., Geverink, N.A.,Van Nuffel, A., van Steenbergen, L.,Van Reenen, K., Hautekiet V., Vermeulen, K.,Velarde, A., Tuyttens, F.A.M., 2010. Repeatability of lameness, fear and slipping scores to assess animal welfare upon arrival in pig slaughterhouses. Anim. 4, 804–809.

de Passillé, A.M., Rushen, J., 2005. Can we measure human–animal interactions in on-farm animal welfare assessment? Some unresolved issues. Appl. Anim. Behav. Sci. 92, 193–209.

de Rosa, G., Tripaldi, C., Napolitano, F., Grasso, F., Bisegna, V., Bordi, V., 2003. Repeatability of some animal related variables in dairy cows and buffaloes. Anim. Welf. 12, 625-629.

De Vet, H.C.W., 1998. Observer reliability and agreement. In: Armitage, P., Colton, T. (Eds.), Encyclopedia of Biostatistics. John Wiley and Sons, New York.

De Vet, H.C.W, Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59, 1033- 1039.

Dohoo, I., Martin, W., Stryhn, H., 2009. Screening and diagnostic tests. In: Dohoo, I., Martin, W., Stryhn, H. (Eds), Veterinary Epidemiological Reseach. Inc, Charlottetown, Price Edward Island, Canada, pp. 91 – 134.

Forkman, B., Keeling, L., 2009. Assessment of animal welfare measures for sows, piglets and fattening pigs Welfare Quality Reports, Nº. 10. Cardiff University, Cardiff, UK.

Gillman, C.E., KilBride, A.L, Ossent, P., Green, L.E., 2008. A cross-sectional study of the prevalence and associated risk factors for bursitis in weaner, grower and finisher pigs from 93 commercial farms in England. Prev. Vet. Med. 83, 308-322.

Hemsworth, P.H., Barnett, J.L., Coleman, G.J., Hansen, C., 1989. A study of the relationships between the attitudinal and behavioural profiles of stockpersons and the level of fear of humans and reproductive performance of commercial pigs. Appl. Anim. Behav. Sci. 23, 301-314.

Hemsworth, P.H., Cronin, G.M., 2006. Behavioral problems. In: Straw, B., Zimmerman, J.J., D'Allaire, S., Taylor, D., (Eds), *Diseases of swine*. Blackwell Publishing, Iowa, pp 847 – 859.

Huynh, T.T.T., Aarnink, A.J.A., Gerrits, W.J.J., Heetkamp, M.J.H., Canh, T.T. Spoolder, H.A.M., 2005. Thermal behaviour of growing pigs in response to high temperature and humidity. Appl. Anim. Behav. Sci. 91, 1-16.

Keeling, L.J., 2011. Measuring up to expectations: the use of animal outcome indicators in welfare assessment. *Proceedings of the 5th international conference of the assessment of animal welfare at farm and group level* (WAFL), Guelph, Canada. Wageningen Academic publishers.

Knierim, U., Winckler, C., 2009. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. Anim. Welf. 18, 451-458.

Lensink, B.J., Van Reenen, C.G., Engel, B., Rodenburg, T.B. Veissier, I., 2003. Repeatability and reliability of an approach test to determine calves' responsiveness to humans: 'a brief report'. Appl. Anim. Behav. Sci. 83, 325-330.

Martin, P., Bateson, P., 2007. Measuring Behaviour. Cambridge University Press, Cambridge.

Meagher, R.K, 2009. Observer ratings: validity and value as a tool for animal welfare research. Appl. Anim. Behav. Sci. 119, 1-14.

Mullan, S., Edwards, S.A., Butterworth, A., Whay, H.R., Main, D.C.J., 2011. Inter-observer reliability testing of pig welfare outcome measures proposed for inclusion within farm assurance schemes. Vet. J. *In press*.

Pearce, G.P., 1999. Epidemiology of enteric disease in grower-finisher pigs: A postal survey of pig producers in england. Vet. Rec. 144, 338-342.

Phythian, C.J., Wemelsfelder, F., Michalopoulou, E., Duncan, J.S., 2011. Qualitative behaviour assessment in sheep: consistency across time and association with health indicators. *Proceedings of the 5th* *international conference of the assessment of animal welfare at farm and group level* (WAFL), Guelph, Canada. Wageningen Academic publishers.

Plesch, G., Broerkens, N., Laister, S., Winckler, C., Knierim, U., 2010. Reliability and feasibility of selected measures concerning resting behaviour for the on-farm werfare assessment in dairy cows. Appl. Anim. Behav. Sci. 126, 19-26.

Rencher, A.C., 2002. Principal component analysis. In: Rencher, A.C. (Ed.), Methods of Multivariate Analysis. John Wiley and sons, New York, pp. 380-405.

Rousing, T., Waiblinger, S., 2004. Evaluation of on-farm methods for testing the human–animal relationship in dairy herds with cubicle loose housing systems test retest and inter-observer reliability and consistency to familiarity of test person. Appl. Anim. Behav. Sci. 85, 215–231.

Scott, K., Chennells, D.J., Armstrong, D., Taylor, L., Gill, B.P., Edwards, S.A., 2007. The welfare of finishing pigs under different housing and feeding systems: liquid versus dry feeding in fully-slatted and straw-based housing. Anim. Welf. 16, 53-62.

Taylor, N.R., Main, D.C.J., Mendl, M., Edwards, S.A., 2010. Tail-biting: A new perspective. Vet. J. 186, 137 – 147.

Temple, D., Dalmau, A., Ruiz de la Torre J.L., Manteca, X., Velarde, A., 2011a. Application of the Welfare Quality® protocol to assess growing pigs kept under intensive conditions in Spain. J. Vet. Behav. 6, 138–149.

Temple, D., Manteca, X., Velarde, A., Dalmau, A., 2011b. Assesssment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. Appl. Anim. Behav. Sci. 131, 29 – 39.

Temple, D., Courboulay, V., Velarde, A., Dalmau, A., Manteca, X., 2012. The welfare of growing pigs in five different production systems in France and Spain: assessment of health. Anim. Welf. 21, 257-271.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., de Vet H.C.W., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60, 34-42.

Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. 19, 231-240.

Welfare Quality®, 2009. Welfare Quality® applied to growing and finishing pigs, In: Dalmau, A., Velarde, A., Scott, K., Edwards, S., Veissier, I., Keeling, L., Butterworth, A. (Eds.), Welfare Quality® Assessment protocol for pigs. Welfare Quality® Consortium, The Netherlands.

Wemelsfelder, F., Hunter, E. A., Lawrence, A. B., Mendl, M. T., 2001. Assessing the 'whole-animal': a free-choice-profiling approach. Anim. Behav. 62, 209–220.

Wemelsfelder, F., 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Anim. Welf. 16, 25-31.

Wemelsfelder, F., Millard, F., 2009. Qualitative Behaviour assessment, in: Forkman, B., Keeling, L. (Eds.), Assessment of animal welfare measures for sows, piglets and fattening, Slu service, The Netherlands, pp. 213-219.

Winckler, C., Brinkmann, J.,Glatz, J., 2007. Long-term consistency of selected animal-related welfare parameters in dairy farms. Anim. Welf. 16, 197-199

Zar, J.H., 1999. Simple linear correlation, In: Zar (Ed.), Biostatistical analysis. Prentice Hall, New Jersey, USA, pp: 377-412.

GENERAL DISCUSSION

OVERALL EVALUATION OF THE MONITORING PROTOCOL

1. FEASIBILITY AND SENSITIVITY

1.1. Feasibility

The average time required for conducting the entire protocol on intensive conventional farms of growing pigs was 6 hours and 20 minutes. The interview and general information recording are expected to be much shorter on a second visit, which may then last one hour less. Current certification schemes last about 2 hours, which is less than needed for the Welfare Quality® (WQ®) assessment in its present form (Botreau et al., 2009b). Several stakeholders expressed their worries that the visit of farms to run the WelfareQuality® assessment would be too long and thus too costly (Veissier et al., 2007). Several strategies should then be considered to reduce the duration of the second visit. A first plausible strategy is to focus on a subset of measures on a second visit based on the results from a first visit. For improvement purposes, those measures should be the ones with higher prevalence / incidence. Secondly, based on the presence or absence of "predictors" and on risk evaluation, the visit may focus on welfare problems that are more likely to happen. For example, the assessment of moderate bursitis may not make sense in young pigs on straw bedded floors. Finally, having completed at least two visits, the assessment may focus only on persistent welfare problems. Farms that did not present any persistent problem may be evaluated less frequently than farms with constantly high prevalences across visits.

Although some WQ® methodologies to assess animal welfare in pigs have been developed under intensive conditions, methodological problems may arise when applying the protocol to extensive conditions. Difficulties in observing the animals due paddock dimensions and to the lack of close contact with humans may result in the assessment of some parameters such as bursitis or wounds on the body being difficult. However, it was possible to assess the vast majority of parameters – except behavioural ones - when animals were gathered for inspection or for feeding. The methodologies of behavioural recording were flexible enough to allow the recordings of measures in extensive conditions outside feeding time.

1.2. Sensitivity

The assessment system has to cover all aspects of welfare and its sensitivity depends on what is included in it. In the context of a welfare assessment system on farm, the sensitivity of a measure refers to its capacity to detect variations between commercial farms. Variations between farms include the effects of variations in the way the farming system is managed as well as specific housing-animal interactions. The differences in sensitivity of a welfare measure between farms within the same production system or between production systems can then be assessed. In both cases, it is necessary to describe the intra-farm variability in order to find out if the measure is representative of the actual welfare state of the overall group of animals present on the farm.

The majority of measures from the good housing and appropriate behaviour principles of the WQ® protocol were sensitive enough to detect differences between conventional farms. On the other hand, body condition, indicators of thermal comfort and the majority of health indicators (except coughs and sneezes) did not allow us to discriminate between conventional farms. A possible explanation is that the threshold established when defining the categorical scoring scale of several measures is too restrictive. To increase repeatability between observers but to the detriment of the sensitivity of these measures, clear categories of scores were defined. As an example, in order to increase the proportion of observers allocating the same score, Brenninkmeyer et al. (2007) and March et al. (2007) merged several more detailed classes (5-point lameness scale) into fewer classes (2-point scale: lame vs. not lame). Therefore, allowing greater differences between observers improves reliability but decreases discrimination ability (Engel et al., 2003). The main problem is finding a balance between repeatability and sensitivity. Tuyttens et al. (2009) have shown that continuous scales are potentially more sensitive than simple categorical scales without affecting the repeatability of the measure. The choice of a continuous scale should be taken into consideration for measures such as wounds on the body or lameness to increase their sensitivity at commercial level. A second possible explanation for a low sensitivity between farms may be the high within farm (or between pens) variability that is found in some measures such as tail biting, lameness, etc. A lack of sensitivity would then be linked to a lack of accuracy in the sampling strategy, so that the estimated prevalences of those measures would not be reliable. According to Endres et al. (2011) estimating the proportion of thin and fat cows required that 70% to 80% of the pens be measured. Therefore, a higher percentage of pen/group of animals must be sampled to generate accurate estimates for relatively rare parameters such as very thin animals in cattle (in European cattle rearing systems) or tail biting in pigs than for more frequent parameters such as moderate bursitis in pigs on conventional farms. This percentage is even more important on bigger farms where the ability to pick up such infrequent welfare problems will be lower than on small farms. Conventional farms of growing pigs can often housed more than 2 000 animals (representing approximately 50% of the farms assessed in the present study) which makes the evaluation of more than 50% of the animals unfeasible. Changes in the sampling strategy in order to increase the sensitivity of rare parameters are less likely to happen for feasibility reasons. For those rare parameters, the main issue to be decided is whether we shall consider the average state of the animals on the farm or put more attention on the worse animals. The assessment of hospital pens as well as cull animals for example may be more representative than a random sampling of rare problems. Finally, this low sensitivity can simply reflect homogeneity in the actual welfare state of the pigs housed on the farms assessed.

To increase its sensitivity, the assessment system was then applied on a large scale sampling of farms from several production systems. However, health measures kept presenting a low sensitivity between farms even though the sample of farms was intended to be heterogeneous from a sanitary point of view (based on veterinary records). As for measures of thermal comfort, they truly appeared to be not sensitive enough to detect stress induced by high environmental temperatures. Estimations of the thermal comfort zone range from 15 to 24ºC and from 14 to 21ºC in “white” growing and finishing pigs, respectively (Gonyou et al., 2006). On the days of assessment, the fourth quartile of environment temperatures ranged from 25 to 29ºC and from 24 to 40ºC on conventional and straw bedded farms, respectively. However, the prevalence of pigs panting was 0.2% on conventional farms and 0.7% in pigs on-straw. It is very likely that heat stress had not been detected in many of the animals assessed. Other feasible indicators of heat stress, more sensitive than panting, should be validated and included into the protocol for growing pigs on farm.

To some extent, it was also surprising not to detect any welfare concern linked to the rearing of pigs in extensive conditions. The prevalence of poor body condition was significantly higher in extensive Mallorcan Black pigs than in pigs kept on conventional farms. However, all the other parameters failed to reveal a welfare concern in extensive conditions and many prevalences were below 1%. The presence of diseases such as digestive disorders caused by parasites or bacteria is a particular concern in extensive Iberian pigs (Astorga et al., 2010). The animals are exposed to multiple environmental factors including the contact with wild boars, foxes and birds which contribute to the dissemination of pathogens. Even when livestock can be observed, the question remains of how representative health measures, such as presence of diarrhoea, really are at herd level, given the variability of the environment. Unlike intensively-housed livestock that generally have controlled and relatively constant environments, the environment of an extensive animal and the associated welfare status may change dramatically in a relatively short period of time, such as in response to severe weather, accident or disease outbreak. Beyond the difficulties in detecting health problems, there definitively remain welfare challenges in extensive production systems, including the large seasonal variations in food and water supply and quality, climatic conditions and in some cases the restricted opportunities for direct monitoring and supervision (Matthews, 1996; Petherick, 2005). All aspects that have a clear impact on welfare in intensive conditions have been included into the WQ® protocol. The sensitivity of the protocol in extensive conditions relies, then, on the capacity of those animal-based parameters to detect welfare problems. Considering that extensive systems present specific challenges to welfare, several items are likely to be missing in the protocol and animal based observations alone are unlikely to be sensitive enough to detect welfare problems in extensive conditions. Details on the environment such as the presence of shades, wallowing facilities as well as the evaluation of the quality of the stockmanship and the use of some management practices such as nose ringing should also be taken into consideration when assessing the welfare of extensively kept animals.

2. INTERPRETATION OF THE OUTCOMES: SELECTIVITY AND CONTEXTUAL VALIDITY

Welfare is a mental state, resulting from how the animal experiences its environment (Dawkins, 1980; Duncan, 1996; Bracke et al., 1999). Parameters used to assess animal welfare are not direct measures of this mental state but only indices that need to be interpreted. Complex interactions between environmental aspects and with the animal itself lead generally to unpredictable outcomes (Waiblinger et al., 2001). For that reason animal-based parameters are generally considered to be more closely linked to the welfare of animals. However, their relationship with the actual mental state of the animals is not always certain. Despite being practical and focused, interpretation of the significance of the results to the animals themselves presents an ongoing challenge (Whay et al., 2007). Although the validity of each independent measure integrated into the WQ® protocol for growing pigs has been discussed in pilot studies (Forkman and Keeling, 2009), many indicators may not be specific enough and some of them have been accepted only for consensus validity.

2.1. Selectivity

We should consider if we are measuring what we want and not something else, which refers to the selectivity of a measure once applied in commercial conditions as well as what are the factors that can influence the outputs and if that measure can be interpreted as regards the criteria for which it is used. The prevalence of bursitis, for example, was significantly different between production systems (e.g. conventional vs. straw bedded) and possible environmental causes of bursitis such as the density and the type of floor were detected. Bursitis seems, therefore, a selective indicator of comfort around resting. The prevalence of poor body condition was associated with pigs feeding on pasture in extensive Mallorcan Black pigs and it seems an appropriate indicator of prolonged hunger. Nevertheless, poor body condition was also associated with pigs on-straw in their early and mid growing period. As commented by Scott et al. (2006), poor body condition may also be indicator of Postweaning Multisystemic Wasting Syndrome (PMWS) and could therefore relate to the "absence of diseases" criteria. Coughs and sneezes which are included in the protocol as indicators of respiratory diseases were sensitive to differences between farms; however, this variation was not easy to interpret in terms of welfare, and neither possible causes of coughs and sneezes nor differences between production systems were detected (not published). The incidence of coughs and sneezes may vary due to many causes, such as viruses, bacteria, parasites or environmental contaminants (dust, ammonia, etc.) Similarly, the link between diarrhoea and diseases was confusing, as the consistency of faeces may vary to some extent according to the diet and the age of the animals. Faeces can become more liquid when the pigs are liquid fed or when extensive pigs are fed on pasture or during the *montanera* period. Detection of diarrhoea may also be impaired in dirty floors, slatted floors or in extensive conditions. Therefore, each one of these signs of disorders, individually assessed can be nonspecific (Scott et al., 2007; Wemelsfelder, 2007) and combining those signs with other types of measure may provide more specific information on the presence of diseases. For instance, the detection of inactive animals with a low body condition and scruffy hair coat can be a selective indicator of the fever reaction caused by an active infection (Hart, 2010). Prevalence of "sick" animals can be high in some farms with acute infection. In fact, detection of "runt" pigs is usually used by vets to detect sick animals (Consensus validity). A recent study on the validity of sickness behaviour in dairy calves has proposed to integrate behavioural tools such as lethargy score, standing motionless and willingness to approach the observer as predictors of disease (Stanton et al., 2011). Similar indicators of "sickness" may increase the selectivity (and probably the sensitivity) of health measures in growing pigs and should be of interest in further studies. In addition to the validation of new animal-based indicators of diseases, evidence of prevention and active management of diseases can be of special interest to increase our interpretation of health outcomes (Main et al., 2001). Some points such as the correctness of a treatment or the management of hospital pens can be selective measures of health problems. Studies are needed to assess attitudes of farmers towards the prevention of diseases and to select valid and feasible management-based indicators of health.

2.2. Contextual validity

Another source of lack of validity can arise when the meaning of a result differs depending on the conditions (Contextual validity). For example, variation in occurrences of behaviour from one production system to another may not have the same significance. The difficulties in interpretation of an outcome measure have been outlighted by de Passillé and Rushen (2005) when validating a test to measure human animal interactions on farm. The validity of "avoidance distance towards an observer" as an indicator of human-animal relationship in cattle was critically discussed by the authors who ended up asking whether the measure really discriminates among farms with and without a true welfare deficiency in terms of human-animal relationship. The assessment of welfare through behavioural parameters from the WQ® protocol revealed clear difficulties of interpretation when comparing farm units of Iberian pigs in intensive and extensive conditions. For examples, interpretations of variations in the frequencies of "positive" social behaviour and exploration were not straightforward. In addition, even though the mean frequencies of the panic response did not depend on the rearing system, there were marked differences between extensive and intensive units in the degree to which the animals were fearful to humans. Several factors such as the restricted space to flee and the frequency of human contact (Waiblinger et al., 2006) may explain variations in the fear response to the observer presence. Those influencing factors may be even more pronounced in intensive conditions. As a result, the outcomes of the human animal relationship test would not vary only depending on the handling of the animals but also on confounding factors that may differ between rearing systems.

An erroneous interpretation of a behavioural outcome can wrongly favour or prejudice a farmer. As part of the WQ® quality protocol, the assessment of "appropriate behaviour" has to be valid across systems and it should be possible to extrapolate from intensive to extensive conditions (Appleby, 1996). A parameter and its associated methodology of assessment should be validated under both rearing conditions. Different thresholds may need to be established depending on the production system (e.g. different flight distances); however, the final score obtained (e.g. panic response) will have to reflect a welfare problem (obvious poor human-animal relationships) in both conditions. Finally, it is necessary to find indicators of emotions (or affect) that should be valid across systems and applicable on farm.

3. BENCHMARKING AND SETTING THRESHOLDS

3.1. Benchmarking system

A benchmarking system informs farmers about their own performance and encourages them to make improvement in problem areas. Although welfare assessment systems can be used for legislation or certification purposes, they may also be used as an advisory/management tool. The results from the welfare assessment system need to be sufficiently simple and transparent to be fully understood by the farmer.

Generalized Linear Mixed Models (GLMM) were used for modelling non-parametric data with very low prevalences and count data. The methodology allowed for a practical detection of farms with welfare problems in relation to a given population. The model was applied on 30 conventional farms; however its assessment should be feasible on a larger sample of farms from different production systems. This benchmarking system provides individual farms with data from their own performance in relation to averages from other farms. Moreover, mean prevalences and distributions of several welfare outcomes were obtained on a wide range of commercial farms.

Graded production systems can be useful to raise awareness on the apparition of welfare problems. Benchmarking can also help to set thresholds (Keeling, 2011). For example, it may guide the aggregation system and the experts' decisions when establishing thresholds of awareness for each measure. Information on the range of variation of a measure within a large population of farms (minimum, maximum, mean and standard error) is needed to adjust the aggregation model to the characteristics of that measure (Botreau et al., 2007b). As mentioned before, several measures from the WQ® protocol presented low prevalences and small variability once applied on a large sample of commercial farms. Discrimination thresholds of those measures are thus expected to be low in order to detect significant values of poor welfare.

3.2. Detecting causal factors

Welfare assessment alone is unlikely to deliver welfare improvement without some additional strategies to improve welfare and driver to motivate farmers to respond positively to the assessment (Whay et al., 2007). The identification of risk factors is fundamental to quantify the level of impact of an intervention. Reducing risk is one aim of recommendation on housing and management. Predictive factors can give an estimation of which type of welfare problem is more likely to appear on a given farm. The visit could then be focused on the target animal-based measures increasing the feasibility of the WQ® protocol as a commercial tool. A descriptive epidemiological approach was undertaken to consider prevalences with a view to determining risk factors. It provides information on the links between possible causal factors and their consequences for several welfare outcomes. Important differences between production systems can be established based on the majority of animal-based parameters of the good feeding and housing principles. Pigs in the conventional system presented the lowest prevalence of poor body condition while extensive Mallorcan Black pigs and extensive Iberian pigs had a decreased prevalence of bursitis and pig dirtiness. The straw bedded system presented a lower prevalence of bursitis but poorer hygiene and poorer body condition than the conventional system. The odds ratios reported for several parameters of feeding and housing were very high when comparing production systems suggesting that the production system by itself has an important effect on the apparition of such welfare problems (e.g.

bursitis or pig dirtiness). It should be emphasised though that no indicator of thermal comfort (huddling, shivering and panting) differed between production systems.

In contrast to measures of feeding and housing, health measures differed, in general, very little between the five production systems studied. Odds ratios included unity or were closed to 1. A lack of sensitivity of health measures recorded by the WQ® protocol may explain this homogeneity between production systems. Another possible explanation is that production systems may be truly similar in the prevalence of health problems. To answer this question health measures recorded by means of the Welfare Quality® protocol should be compared to a "gold standard" or to epidemiological studies applied on the same population of farms. The impact of causal factors often varied depending on the production system and according to the severity (score) of a parameter. To make the interpretation of the results easier, models were developed for each score and by production system. Nevertheless, for some parameters that presented over-dispersion (e.g. skin condition), the high variability between farms may suggest that classifying farms by production system is less informative than individual management considerations.

4. TEST-RETEST REPEATABILITY

Consistency of a welfare assessment system over time is especially important when it is to be used for certification purposes. In order to be cost effective, the Welfare Quality® assessments should take place in long intervals of supposedly more than six months (Knierim and Winckler, 2009). The frequency of visits will then be influenced by how often welfare can be feasibly be assessed. Furthermore, if the main objective is to promote improvement, then tracking change / stability over time becomes fundamental (Keeling, 2011). This should be done considering the repeatability of each farm individually. The assessment of repeatability becomes fundamental to detect the risk of poor welfare at some point in the future on the individual farm. Regardless of whether the protocol is used as a certification tool or to promote improvement, the outcomes should be representative of the long-term situation of a farm detecting significant changes and not being sensitive to minor changes included within the usual variation in farm conditions (Plesch et al., 2010). Repeatability will depend on the nature, distribution and sensitivity of each parameter along time.

Several methodologies were combined to evaluate the repeatability of the WQ® assessment protocol on a sample of 15 intensive conventional farms. The results were discussed from a discriminative (reliability) and an evaluative (agreement) point of view. The main limitation of discriminative methods is their strong dependence on the variance of the assessed farms (Bland and Altman, 1996; Costa-Santos et al., 2011). The contradictions between Spearman's correlation and

ICC observed for wounds on the body, moderate soiled body and scouring highlight the dependence of those discriminative methods on the variability between farms. In addition to low hazards and to the lack of sensitivity of some parameters to discriminate farms, a high measurement error is expected between visits. This makes the evaluation of repeatability using discriminative methods difficult. In contrast, the limits of agreement used as an evaluative methodology are not influenced by the variance of the assessed population (de Vet et al., 2006); still, interpretation of the plots and of the limits values is highly subjective (Costa Santos et al., 2011).

Six measures (panting, shivering, pumping, twisted snout, prolapse, hernia and scouring scored as 2) had prevalence close to 0% and a low variability in the two visits thus hampering the estimation of repeatability. Body condition, tail biting and lameness had low prevalence, but sufficient variability to assess repeatability over time. From a discriminative point of view, only moderately soiled body and the human animal relationship test showed a high reliability over a long period of time. Huddling had a moderate repeatability. Severely soiled body and wounds on the body presented a high reliability using ICC. The two sets of scores from the PCA of the QBA were moderately related on the first axis using partial Spearman's correlations; however, the scores from the QBA suffered a clear visit effect. The remaining measures were not repeatable over time using Spearman's correlation coefficient and/or ICC. The lack of reliability of a parameter over time may be due to: 1) an unpredictable time effect; 2) the sensitivity of a parameter to routine (not less important) changes but not to major changes in the management/resources of the farm; 3) methodological problems. Those three major causes of variability between visits are more ore less likely to happen depending on the parameter studied. Adopting an evaluative approach appears complementary to guide the interpretation of the results from discriminative methods. Some parameters such as localized and widespread skin condition, tail biting, respiratory and enteric disorders are more likely to be subject to unpredictable variations along time and therefore sporadically affect a farm. Those indicators were actually difficult to predict by means of simple measurements on the environment, as commented before. Between-farm and intra-farm variability were high and few causal factors were identified. By nature, those parameters are expected to show a low repeatability. Attention should then be given to farms that present one (or several) of those problems persistently across visits.

Depending on the welfare problem and on the situation in which it occurs, the farmer will have more or less capacity to control its occurrence. Routine management (but not "major" changes) can prevent a welfare problem from becoming persistent. Following a proper management, farmers will be able to control a respiratory or an enteric disorder. The efficiency of the farmer/veterinarian in applying a correct treatment (vaccines), controlling the ventilation and keeping a good hygiene will prevent most of the coughs/sneezes and diarrhoeas from becoming persistent. On intensive farms, tail biting may be difficult to prevent; however, a pig with a bitten tail should be placed immediately in a hospital pen. Tail biting usually appears as an outbreak (Paul et al., 2007) and repeated occurrence of tail biting on a given farm should be prevented by good management. In our study, no farm assessed had an outbreak of tail biting on the two visits either because of a complete absence of tail biting at the moment of the visit or due to proper management of sick animals. Lameness and poor body condition do not appear as outbreaks; however, their occurrence may depend, as well, on the incorrect management of hospital pens. Their lack of repeatability may show that the majority of farmers do, indeed, isolate sick animals as recommended. Therefore, it can be concluded that routine changes on a farm will obviously affect repeatability of several measures even if “major” changes did not occurred. Therefore, adopting an evaluative approach such as the limits of agreement may be better to identify farms that have persistent welfare problems and thus higher risks of poor welfare.

Finally, methodological aspects should be considered when interpreting a lack of repeatability between visits. Some parameters may be too sensitive and their variation may be due to slight fluctuations in the environment or in the animals themselves (e.g. social behaviour). Other parameters that present high intra-farm variability (between pen variability) might not be detected by the sampling in both visits. As mentioned before, parameters such as tail biting or lameness might not be representative of the current welfare state of the overall group of animals present on the farm. Intra-observer variability represents a third methodological aspect that should be taken into account when interpreting repeatability.

CONCLUSIONS

Feasibility of the welfare assessment system

- The overall Welfare Quality® protocol for growing pigs on farm is easy to perform on intensive conventional farms and requires little input from the farmers; however, it takes an average time of 6 hours and 20 minutes per visit.

- Under extensive conditions it is feasible to assess the great majority of measures – except behavioural ones - when the animals are gathered for inspection or feeding and the methodologies of behavioural recording are flexible enough to allow recording outside feeding time.

Sensitivity of the welfare assessment system to detect variations between farms and between production systems

- The majority of measures from the good housing and appropriate behaviour principles of the Welfare Quality® protocol are sensitive enough to detect differences between intensive conventional farms of growing pigs; however, poor body condition, indicators of thermal comfort and the majority of health indicators do not allow discrimination among intensive conventional farms.

- Important differences between production systems can be established based on animal-based measures of the good feeding and housing principles of the Welfare Quality® protocol; in contrast, health measures do not differ significantly between the five production systems studied.

- Negative and "positive" social behaviours as well as the qualitative behaviour assessment are sensitive to changes in the housing and management of Iberian pigs; in contrast, no differences are found between extensive and intensive reared Iberian pigs in the expression of exploratory behaviour and in human-animal relationship.

Selectivity and contextual validity of the outcome results

- The variation in the prevalence of some animal-based measures - such as poor body condition, coughs and sneezes, diarrhoea, skin condition – is not easy to interpret in terms of welfare reflecting a possible lack of selectivity of those measures.
- The high occurrence of negative social behaviour recorded in intensive conditions is clearly an indicator of poor welfare; however, interpretations of variations in the frequencies of "positive" social behaviour and exploration are not straightforward.
- The qualitative behaviour assessment appears to be valid to discriminate farms on the basis of the expression of natural behaviour.

Test-retest repeatability over a long period of time

- Most of the quantitative animal-based measures, except moderate bursitis, widespread skin discolouration and positive social behaviour, do not present mean results significantly different between two visits on conventional farms with a 12 months interval between visits.
- From a discriminative point of view, only moderately soiled body and the human animal relationship test show a high reliability over a long period of time; however inconsistencies between Spearman's correlation coefficient and intraclass correlation coefficient are found when interpreting the repeatability of several health measures.
- The scores of the qualitative behaviour assessment from intensive conventional farms of pigs are not stable over time.

Testing the Welfare Quality® protocol as a tool to identify and prevent welfare problems

- A benchmarking system based on the use of Generalized Linear Mixed Models for non-parametric data with low prevalence allows the identification of farms with welfare problems in a given population.

- Some possible causal factors relating to severely wounded animals, tail biting and lameness were identified in the intensive conventional system; then, the recording of simple environmental-based factors can be useful to detect farms that are more likely to show one of these health problems.

- Studying stability over time adopting an evaluative approach such as the Limits of Agreements allows a practical detection of farms that present a persistent welfare problem.

REFERENCES

From the general introduction and discussion

Almond GW, Flowers WL, Batista L and D'Allaire S 2006. Diseases of the reproductive system. In: Straw BE, Zimmerman JJ, D'Allaire S and Taylor DJ (eds.), *Diseases of Swine* pp 113-147. Blackwell Publishing, Iowa, USA

Amon T, Amon B, Ofner E and Boxberger J 2001. Precision of assessment of animal welfare by the TGI 35L Austrian Animal Needs Index. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 114–117

Amory JR, Mackenzie AM, Eckersall PD, Stear MJ and Pearce GP 2007. Influence of rearing conditions and respiratory disease on haptoglobin levels in the pig at slaughter. *Research in Veterinary Science 83:* 428–435

Appleby MC 1996. Can we extrapolate from intensive to extensive conditions? *Applied Animal Behaviour Science 49:* 23-27

Archer J 1973. Tests for emotionality in rats and mice: a review. *Animal Behavior 21:* 205-235

Astorga RJ, Gómez-Laguna J, Moreno P, Hernández M, Carrasco L 2010. Sanidad animal en el Cerdo Ibérico. *Sólo Ibérico 24*: 43 - 50

Baldock NM and Sibly RM 1990. Effects of handling and transportation on the heart rate and behaviour of sheep. *Applied Animal Behaviour Science 28*:15-39

Barnett JL and Hemsworth PH 1990. The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science 25*: 177-187

Barnett JL, Hemsworth PH, Cronin GM, Jongman EC and Hutson GD 2001. A review of the welfare issues for sows and piglets in relation to housing. *Australian Journal of Agricultural Research 52:* 1-28

Bartussek H 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science 61*: 179-192

Bartussek H 2001. An Historical Account of the Development of the Animal Needs Index ANI-35L as Part of the Attempt to Promote and Regulate Farm Animal Welfare in Austria: An Example of the Interaction Between Animal Welfare Science and Society. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 34-41

Bell DE, Raiffa H and Tversky A 1988. *Decision making: descriptive, normative and prescriptive interactions*. Cambridge University Press, Cambridge, UK

Bennett RM 1996. People's willingness to pay for farm animal welfare. *Animal Welfare 5*: 3-11

Bergeron R, Badnell-Waters AJ, Lambton S, Mason G 2006. Stereotypic oral behaviour in captive ungulates: foraging, diet and gastrointestinal function. In: Mason G and Rushen J (eds), *Stereotypic Animal Behaviour, Fundamentals ans Applications to Welfare,* pp 19 – 57. CABI, Wallingford, UK

Blache D, Terlouw C, Maloney SK 2011. Physiology. In: Appleby MC, Mench JA, Olsson IAS and Hughes BO (eds.) *Animal Welfare*, pp.155-182. Second edition, CAB International, Wallingford, UK

Bland JM and Altman DG 1996. Statistical notes: measurement error and correlation coefficients. *British Medical Journal 313:* 41-42

Blokhuis HJ 2008. International cooperation in animal welfare: the Welfare Quality® project. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 1-5

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. *Animal Welfare 12*: 445–455

Boissy A, Bouix J, Orgeur P, Poindron P, Bibe B and Le Neindre P 2005. Genetic analysis of emotional reactivity in sheep: Effects of the genotypes of the lambs and of their dams. *Genetics Selection Evolution 37:* 381-401.

Boissy A, Arnould C, Chaillou E, Désiré L, Duvaux-Ponter C, Greiveldinger L, Leterrier C, Richard S, Roussel S, Saint-Dizier H, Meunier-Salaün MC, Valance D and Veissier I 2007a. Emotions and Cognition: a new approach to animal welfare. *Animal Welfare 16:* 37-43

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J, Bakken M, Veissier I and Aubert A 2007b. Assessment of positive emotions in animals to improve their welfare. *Physiology and Behavior 92:* 375–397

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

Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, Van Reenen CG and Veissier I 2007b. Aggregation of measures to produce an overall assessment of animal welfare. Part 2: analysis of constraints. *Animal 1:* 1188-1197

Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, Van Reenen CG and Veissier I 2007c. Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16:* 225-228

Botreau R, Veissier I and Perny P 2009a. Overall assessment of animal welfare: strategy adopted in Welfare Quality®. *Animal Welfare 18:* 363-370

Botreau R, Bonde M, Buller H, Butterworth A, Evans A, Forkman B, Keeling L, Van Reenen K, Veissier I, Velarde A, Winckler C 2009b. *Scenarios for farm visits.* Deliverable D2.36 and D2.37, WP2.5.

Bouyssou D 1990. Building criteria: A prerequisite for MCDA. In: Bana e Costa CA (ed.) Readings in Multiple Criteria Decision-Aid. Springer-Verlag: Heidelberg, Germany

Bracke MBM, Spruijt BM and Metz JHM 1999. Overall animal welfare reviewed. Part 3: welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science 47:* 307–322

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a. Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. *Journal of Animal Science 80:* 1819-1834

Bracke MBM, Metz JHM, Spruijt BM and Schouten WGP 2002b. Decision support system for overall welfare assessment in pregnant sows B: Validation by expert opinion. *Journal of Animal Science 80:* 1835–1845

Brenninkmeyer C, Dippel S, March S, Brinkmann J, Winckler C and Knierim U 2007. Reliability of a subjective lameness scoring system for dairy cows. *Animal Welfare 16*: 127-129

Broom DM 1986. Indicators of poor welfare. *British Veterinary Journal 142:* 524–526

Broom DM 1987. The veterinary relevance of farm animal ethology. *Veterinary Record 121:* 400–402

Broom DM 1991a. Assessing welfare and suffering. *Behavioural Processes 25:* 117-123

Broom DM 1991b. Animal Welfare. Concepts and measurements. *Journal of Animal Science 69*: 4167-4175

Broom DM 2006. Behaviour and welfare in relation to pathology. *Applied Animal Behaviour Science 97:* 73–83.

Broom DM 2010a. Cognitive ability and awareness in domestic animals and decisions about obligations to animals. *Applied Animal Behaviour Science 126:* 1–11

Broom DM 2010b. Welfare of animals: behaviour as a basis for decisions. In: Breed MD and Moore J (eds.) *Encyclopedia of Animal Behavior,* pp 580-584. Academic Press: Oxford, UK

Broom DM and Johnson KG 1993. *Stress and Animal Welfare*. Chapman and Hall, London, UK

Candiani D, Salamano G, Mellia E, Doglione L, Renato B, Toussaint M and Gruys E 2008. A combination of behavioral and physiological indicators for assessing pig welfare on the farm. *Journal of Applied Animal Welfare Science 11:* 11 – 13

Capdeville J and Veissier I 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 62-68

Cook CJ, Mellor DJ, Harris PJ, Ingram JR, Mathews LR 2000. Hands-on and hands-off measurement of stress. In: Moberg GP and Mench JA (eds.), *The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare,* pp. 123–146. CAB International, Wallingford, UK

Costa-Santos C, Bernardes J, Ayres-de-Campos D, Costa A and Costa C 2011. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. *Journal of Clinical Epidemiology 64:* 264-269

Dalin AM, Magnusson U, Haggendal J and Nyberg L 1993. The effect of transport stress on plasma levels of catecholamines, cortisol, corticosteroid-binding globulin, blood cell count and lymphocyte proliferation in pigs. *Acta Agriculturae Scandinavica, Section A – Animal Science 34:* 59-68

Dalmau A, Velarde A, Gispert M 2009. Standardisation of the measure “meat quality” to assess the welfare of pigs at slaughter. In: Forkman B and Keeling L (eds.), *Assessment of animal welfare measures for sows, piglets and fattening pigs*, pp. 117-124, Welfare Quality Reports, Nº. 10. Cardiff University, Cardiff, UK

Dantzer R 2002. Can farm animal welfare be understood without taking into account the issues of emotion and cognition? *Journal of Animal Science 80*: 1 - 9

Dawkins MS 1980. *Animal Suffering: The Science of Animal Welfare*. Chapman & Hall, London, UK

Dawkins MS 1983. Battery hens name their price: Consumer demand theory and the measurement of ethological ‘needs’. *Animal Behaviour 31:* 1195-1205

Dawkins MS 2000. Animal minds and animal emotions. *American Zoologist 40:* 883–888

Dawkins MS 2004. Using behaviour to assess animal welfare. *Animal Welfare 13:* 3-7 de Passillé AM and Rushen J 2005. Can we measure human–animal interactions in on-farm animal welfare assessment? Some unresolved issues. *Applied Animal Behaviour Science 92:* 193–209

de Rosa G, Tripaldi C, Napolitano F, Grasso F, Bisegna V and Bordi V 2003 Repeatability of some animal related variables in dairy cows and buffaloes. *Animal Welfare 12*: 625-629

De Vet HCW, Terwee CB, Knol DL and Bouter LM 2006. When to use agreement versus reliability measures. *Journal of Clinical Epidemiology 59:* 1033- 1039

Dobromylskyj P, Flecknell PA, Lascelles BD, Livingston A, Taylor P, Waterman-Pearson A 2000. Pain assessment. In: Flecknell P.A. and Waterman-Pearson A *(eds.), Pain management in animals, pp. 53-79.* Harcourt Publishers, London, UK

Doyle RE, Vidal S, Fisher AD, Hinch GN, Boissy A and Lee C 2010. The effect of repeated testing on judgement biases of sheep. *Behavioural Processes 83:* 249–352

Duncan IJH 1996. Animal welfare defined in terms of feelings. *Acta Agriculturae Scandinavica, Section A – Animal Science 27:* 29–35

Dwyer CM 2008. Environment and the sheep: breed adaptations and welfare implications. In: Dwyer CM (ed.), *The Welfare of Sheep*, pp. 41–79. Springer-Verlag, Heidelberg, Germany

Dwyer CM 2009. Welfare of sheep: Providing for welfare in an extensive environment. *Small Ruminant Research 86:* 14–21

Dwyer CM, Lawrence AB, Brown HE and Simm G 1996. Effect of ewe and lamb genotype on gestation length, lambing ease and neonatal behaviour in lambs. *Reproduction Fertility and Development 8:* 1123–1129

Eckersall PD 1995. Acute Phase Proteins as Markers of Inflammatory Lesions. *Comparative Haematology International 5:* 93-97

Eckersall PD 2000. Recent advances and future prospects for the use of acute phase proteins as markers of disease in animals. *Revue de Médecine Vétérinaire 151:* 577-584.

Endres MI, Espejo LA and Tucker CB 2011. Evaluation of the sample needed to accurately estimate outcome-based measurements of dairy welfare on farm. *Proceedings of the 5th international conference of the assessment of animal welfare at farm and group level (WAFL)*, Wageningen Academic publishers, Guelph, Canada

Engel BG, Bruin G, Andre G and Buist W 2003. Assessment of observer performance in a subjective scoring system: visual classification of the gait of cows. *Journal of Agricultural Science 140*: 317-333

FAWC 1992. Farm Animal Welfare Council updates the five freedoms. *Veterinary Record 17:* 357

Forkman B and Keeling L 2009. *Assessment of animal welfare measures for sows, piglets and fattening pigs*, Welfare Quality Reports, Nº. 10. Cardiff University, Cardiff, UK

Fraser D 1995. Science, values and animal welfare: exploring the 'inextricable connection'. *Animal Welfare 4:* 103-117

Fraser D 2006. Animal welfare assurance programs in food production: a framework for assessing the options. *Animal Welfare 15:* 93-104

Fraser AF and Broom DM 1990. *Farm Animal Behaviour and Welfare.* Baillière Tindall, London, UK

Fraser D and Nicol CJ 2011. Preference and motivation research. In: Appleby MC, Mench JA, Olsson IAS and Hughes BO (eds.). *Animal Welfare*, pp. 183-199. Second edition, CAB International, Wallingford, UK

Freire R, Walker A and Nicol CJ 1999. The relationship between trough height, feather cover and behaviour oflaying hens in modified cages. *Applied Animal Behaviour Science* 63: 55-64

Geers R, Petersen B, Huysmans K, Knura-Deszczka S, De Becker M, Gymnich S, Henot D, Hiss S and Sauerwein H 2003. On-farm monitoring of pig welfare by assessment of housing, management, health records and plasma haptoglobin. *Animal Welfare 12:* 643–647

Gonyou HW 1994. Why the Study of Animal Behavior Is Associated with the Animal Welfare Issue. *Journal of Animal Science 72*: 2171-2177

Gonyou HW, Lemay SP and Zhang Y 2006. Effects of the environment on productivity and disease. In: Straw B, Zimmerman JJ, D'Allaire S and Taylor D (eds.), *Diseases of Swine*, pp 1027-1038. Blackwell Publishing, Ames, IA, USA

Goossens X, Sobry L, Ödberg F, Tuyttens FAM, Maes D, De Smet S, Nevens F, Opsomer G, Lommelen F, Geers R 2008. A population-based on-farm evaluation protocol for comparing the welfare of pigs between farms. *Animal welfare 17:*35-41

Grandinson K 2005. Genetic background of maternal behaviour and its relation to offspring survival. *Livestock Production Science 93:* 43–50

Graves HB 1984. Behavior and ecology of wild and feral swine (*Sus Scrofa*). *Journal of Animal Science 58*: 482-492

Harding EJ, Paul ES and Mendl M 2004. Animal behaviour, cognitive bias and affective state. *Nature 427:* 312

Hargreaves AL and Hutson GD 1990. Changes in heart rate, plasma cortisol and haematocrit of sheep during a shearing procedure. *Applied Animal Behaviour Science 26:* 91-101

Hart BL 2010. Beyond Fever: comparative perspectives on sickness behavior. In: Breed MD and Moore J (eds) *Encyclopedia of Animal Behavior*, pp 205-210. Academic Press, Oxford, UK

Hemsworth PH, Bamett JL, Beveridge L, Matthews LR 1995. The welfare of extensively managed dairy cattle: a review. *Applied Animal Behaviour Science 42:* 161 – 182.

Hessing MJC, Hagelso AM, Schouten WGP, Wiepkema PR and Van Beek JAM 1994. Individual behavioral and physiological strategies in pigs. *Physiology and Behavior* 55:39–46

Hörning B 1998. Scoring systems for an on-farm assessment of housing conditions. In: *Proceedings of the 32nd Congress of the International Society for Applied Ethology*, 21–25 July Clermont- Ferrand, France

Hörning B 2001. The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 42– 47

Houpt KA and Wolski TR 1982. *Domestic animal behaviour for veterinarians and animal science.* The Iowa State university Press, Iowa, USA

Hughes BO and Duncan IJH 1988. The notion of ethological 'need', models of motivation and animal welfare. *Animal Behaviour 36:* 1696-1707

Hurnik JF 1988. Welfare of farm animals. *Applied Animal Behaviour Science 20:* 105-117

Hurnik JF and Lehman H 1988. Ethics and farm animal welfare. *Journal of Agricultural and Environmental Ethics 1:* 305-318.

Jensen P 1986. Observations on the maternal behaviour of freeranging domestic pigs. *Applied Animal Behaviour Science 16*: 131-142

Jensen P and Toates FM 1993. Who needs 'behavioural needs'? Motivational aspects of the needs of animals. *Applied Animal Behaviour Science* 37: 161-181

Johnsen PF, Johannesson T and Sandøe P 2001. Assessment of farm animal welfare at herd level: Many goals, many methods. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 26– 33

Jones RB and Boissy A 2011. Fear and other negative emotions. In: Appleby MC, Mench JA, Olsson IAS and Hughes BO (eds.) *Animal Welfare*, pp.78-97. Second edition, CAB International, Wallingford, UK

Keeling LJ 2011. Measuring up to expectations: the use of animal outcome indicators in welfare assessment. *Proceedings of the 5th international conference of the assessment of animal welfare at farm and group level (WAFL),* Wageningen Academic publishers, Guelph, Canada.

Kerr SGC and Wood-Gush DGM 1987. A comparison of the early behaviour of intensively and extensively reared calves. *Animal Production 45:* 181-190

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

Krohn CC 1994. Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments. III. Grooming, exploration and abnormal behaviour. *Applied Animal Behaviour Science 42:* 73-86

Lazarus RS 1991. Progress on a cognitive–motivational–relational theory of emotion. *American Psychology 46*: 819-834

Leeb B, Leeb Ch, Troxler J and Schuh1 M 2001. Skin Lesions and Callosities in Group-Housed Pregnant Sows: Animal-Related Welfare Indicators. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 82–87

Lensink BJ, Van Reenen CG, Engel B, Rodenburg TB and Veissier I 2003. Repeatability and reliability of an approach test to determine calves' responsiveness to humans: 'a brief report'. *Applied Animal Behaviour Science 83*: 325-330

Losinger WC and Heinrichs AJ 1997. Management practices associated with high mortality among preweaned dairy heifers. *Journal of Dairy Research 64*: 1–11

Lynch JJ, Flinch GN and Adams DB 1992. *The Behaviour of Sheep: Biological Principles and Implications for Production*. CAB International, Wallingford

Main DCJ 2009. Welfare into practice. In: Marchant-Forde JN (ed.), *The Welfare of Pigs*, pp. 287 – 299. Springer-Verlag, Heidelberg, Germany

Main DCJ and Green LE 2000. A descriptive analysis of the operation of the Farm Assured British Pigs. *Veterinary Record 147:* 162–163

Main DCJ, Webster AJF and Green LE 2001. Animal welfare assessment in farm assurance schemes. *Acta Agriculturae Scandinavica, Section A - Animal Science 51*: 108-113

Main DCJ, Kent JP, Wemelsfelder F, Ofner E and Tuyttens FAM 2003a. Applications for methods of on-farm welfare assessment. *Animal Welfare 12*: 523-528

Main DCJ, Whay HR, Green LE, Webster AJF 2003b. Effect of the RSPCA Freedom Food scheme on the welfare of dairy cattle. *Veterinary Record 153:* 227-231

Main DCJ, Whay HR, Leeb C and Webster AJF 2007. Formal animal-based welfare assessment in UK certification schemes. *Animal Welfare 16:* 233-236

Mainau E, Dalmau A, Ruiz de la Torre JL and Manteca X 2010. A behavioural scale to measure ease of farrowing in sows. *Theriogenology 74:* 1279–1287

Manteca X and Ruiz de la Torre JL 1996. Transport of extensively farmed animals. *Applied Animal Behaviour Science 49:* 89-94

March S, Brinkmann J and Winckler C 2007. Effect of training on the inter-observer reliability of lameness scoring in dairy cattle. *Animal Welfare 16*: 131-133

Martin P and Bateson P 2007. Measures of behaviour. In: *Measuring behaviour, an introductory guide*. Cambridge University Press, Newcastle, UK

Mason GJ and Latham N 2004. Can't stop, won't stop: Is stereotypy a reliable animal welfare indicator? *Animal Welfare 13:* 57–69

Mason GJ, Cooper J and Clarebrough C 2001. Frustrations of fur-farmed mink. *Nature 410:* 35

Matthews LR 1996. Animal welfare and sustainability of production under extensive conditions: a non-EU perspective. *Applied Animal Behaviour Science 49:* 41-46

Meagher RK 2009. Observer ratings: validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science 119:* 1-14

Mellor DJ and Murray L 1989. Cortisol responses in young lambs. *Research in Veterinary Science 46:* 387-391

Mendl M and Paul ES 2004. Consciousness, emotion and animal welfare: insights from cognitive science. *Animal Welfare 13:* 17 - 25

Mendl M, Burman OHP, Parker RMA and Paul ES 2009. Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science 118:* 161–181

Merlot E, Meunier-Salaün M and Prunier A 2004. Behavioral, endocrine and immune consequences of mixing in weaned piglets. *Applied Animal Behaviour Science 85:* 247–257

Meunier-Salaün MC, Vantrimponte MN, Raab A and Dantzer R 1987. Effect of floor area restriction upon performance, behavior and physiology of growing-finishing pigs. *Journal of Animal Science 64:* 1371-1377

Miele M and Parisi V 2001. 'L'Etica del Mangiare, i valori e le preoccupazioni dei consumatori per il benessere animale negli allevamenti: un'applicazione dell'analisi Means-end Chain'. *Rivista di Economia Agraria, Anno LVI, No 1*: 81-103 [Title translation: The ethics of eating. Consumers' values and concerns for farm animals' welfare: a means-end chain analysis]

Moberg GP 2000. Biological responses to stress: implications for animal welfare. In: Mench J and Moberg G (eds.), *The Biology of Stress*, pp. 1–21. CAB International, Wallingford, UK

Mollenhorst H, Rodenburg TB, Bokkersb EAM, Koeneb P, de Boera IJM 2005. On-farm assessment of laying hen welfare: a comparison of one environment-based and two animal-based methods. *Applied Animal Behaviour Science 90:* 277–291

Morisse JP, Cotte JP, Huonnic D 1995. Effect of dehorning on behaviour and plasma cortisol responses in young calves. *Applied Animal Behaviour Science 43*: 239-247

Mouttotou N, Hatchell FM and Green LE 1999. Prevalence and risk factors associated with adventitious bursitis in live growing and finishing pigs in south-west England. *Preventive Veterinary Medicine 39:* 39-52

Murata H, Shimada N and Yoshioka M 2004. Current research on acute phase proteins in veterinary diagnosis: an overview. *The Veterinary Journal 168:* 28–40

Nowak R, Poindron P, Le Neindre P and Putu IG 1987. Ability of 12-hour old Merino and crossbred lambs to recognise their mothers. *Applied Animal Behavioural Science 17:* 263–271

Pallarés FJ, Martínez-Subiela S, Seva J, Ramis G, Fuentes P, Bernabe A, Muñoz A and Cerón JJ 2008. Relationship between serum acute phase protein concentrations and lesions in finishing pigs. *The Veterinary Journal 177:* 369–373

Paul ES, Harding EJ and Mendl M 2005. Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews 29:* 469–491

Paul ES, Moinard C, Green LE and Mendl M 2007. Farmers' attitudes to methods for controlling tail biting in pigs. *Veterinary Record 160:* 803–805

Petherick JC 2005. Animal welfare issues associated with extensive livestock production: the northern Australian beef cattle industry. *Applied Animal Behaviour Science 92:* 211-234

Plesch G, Broerkens N, Laister S, Winckler C and Knierim U 2010. Reliability and feasibility of selected measures concerning resting behaviour for the on-farm werfare assessment in dairy cows. *Applied Animal Behaviour Science 126:* 19-26

Porges SW 2002. Is there a major stress system at the periphery other than the adrenals? In: Broom DM, Coe CL, Dallman MF, Dantzer R, Fraser D, Gartner K, Hellhammer DH and Sachser N (eds.), *Coping with Challenge: Welfare in Animals Including Humans*, pp. 133–149. Dahlem University Press, Berlin

Price EO 1984. Behavioral aspects of animal domestication. *Quarterly Review of Biology 59:* 1–32

Rodenburg TB, Komen H, Ellen ED, Uitdehaag KA and van Arendonk JAM 2008. Selection method and early-life history affect behavioural development, feather pecking and cannibalism in laying hens: A review. *Applied Animal Behaviour Science 110:* 217–228

Romeyer A and Bouissou MF 1992. Assessment of fear reactions in domestic sheep, and influence of breed and rearing condition. *Applied Animal Behaviour Science 34:* 93-119

Rousing T and Waiblinger S 2004. Evaluation of on-farm methods for testing the human-animal relationship in dairy herds with cubicle loose housing systems: test-retest and inter-observer reliability and consistency to familiarity of test person. *Applied Animal Behaviour Science 85*: 215-231

Rousing T and Wemelsfelder F 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science 101*: 40–53

Rousing T, Bonde M and Sorensen JT 2001. Aggregating welfare indicators into an operational welfare assessment system: A bottom-up approach. *Acta Agriculturae Scandinavica, Section A - Animal Science 51*: 53-57

Rushen J 1991. Problems associated with the interpretation of physiological data in the assessment of animal-welfare. *Applied Animal Behaviour Science 28:* 381–386

Rushen J 2003. Changing concepts of farm animal welfare: bridging the gap between applied and basic research. *Applied Animal Behaviour Science 81:* 199–214

Rushen J and de Passillé AM 1992. The scientific assessment of the impact of housing on animal welfare: a critical review. *Canadian Joumal of Animal Science 72:* 721-743

Rushen J and de Passillé AM 2009. The scientific basis of animal welfare indicators. In: Smulders FJM and Algers B (eds.) *Food safety assurance and veterinary public health Volume 5. Welfare of production animals : assessment and management of risks*, pp. 391-415. Wageningen Academic Publishers, The Netherlands

Rushen J, Pombourcq E and de Passillé AM 2007. Validation of two measures of lameness in dairy cows. *Applied Animal Behaviour Science 106*: 173-177

Rushen J, de Passillé AM, von Keyserlingk MAG and Weary DM 2008. Housing for adult cattle. In: *The Welfare of Cattle*, pp 142 – 180. Springer-Verlag, Heidelberg, Germany

Russel AJF, Doney JM and Gunn RG 1969. Subjective assessment of body fat in live sheep. *The Journal of Agricultural Science 72 :* 451-454

Sandøe P and Christiansen SP 2008. The use of animals in food production. In: Sandøe P and Christiansen SP (eds.) *Ethics of animal use*, Blackwell Publishing, UK

Sandøe P, Munksgaard L, Badsgard NP and Jensen KH 1996. How to manage the management factor – assessing animal welfare at the farm level. In: *Proceedings of the Fourth International Symposium on Livestock Farming Systems, Foulum, Denmark, August 22–23, EAAP Publication No. 89*. Wageningen Pers, Wageningen, The Netherlands

Scientific Veterinary Committee 1997. The Welfare of Intensively Kept Pigs. Health and Consumer Protection Directorate-General, European Commission. Doc XXIV/B3/ScVC/005/1997, http://europa.eu.int/comm/food/fs/sc/oldcomm4/out17_en.pdf.

Scott EM, Nolan AM and Fitzpatrick JL 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 5-10

Scott K, Chennells DJ, Campbell FM, Hunt B, Armstrong D, Taylor L, Gill BP and Edwards SA 2006. The welfare of finishing pigs in two contrasting housing systems: Fully-slatted versus straw-bedded accommodation. *Livestock Science 103:* 104-115

Scott K, Chennells DJ, Armstrong D, Taylor L, Gill BP and Edwards SA 2007. The welfare of finishing pigs under different housing and feeding systems: liquid versus dry feeding in fully-slatted and straw-based housing. *Animal Welfare 16:* 53-62

Shettleworth SJ 2001. Animal cognition and animal behaviour. *Animal Behaviour 61:* 277-286

Shimmura T, Suzuki T, Azuma T, Hirahara S, Eguchi Y, Uetake K and Tanaka T 2008. Form but not frequency of beak use by hens is changed by housing system. *Applied Animal Behaviour Science 115:* 44–54

Smulders FJM 2009. A practicable approach to assessing risks for animal welfare – methodological considerations. In: Smulders FJM and Algers B (eds.) *Food safety assurance and veterinary public health. Volum 5. Welfare of production animals: assessment and management of risks*, pp 239-274. Wageningen Academic Publishers, The Netherlands

Smulders D, Verbeke G, Mormède P and Geers R 2006. Validation of a behavioral observation tool to assess pig welfare. *Physiology and Behavior 89:* 438–447

Sorensen JT, Sandøe P and Halberg N 2001. Animal welfare as one among several values to be considered at farm level: The idea of an ethical account for livestock farming. *Acta Agriculturae Scandinavica, Section A - Animal Science 51: 11-16*

Spinka M 2006. How important is natural behaviour in animal farming systems? *Applied Animal Behaviour Science 100:* 117–128

Spoolder H, De Rosa G, Hörning B, Waiblinger S and Wemelsfelder F 2003. Integrating parameters to assess on-farm welfare. *Animal Welfare 12*: 529-534

Spoolder H, Geudeke MJ, Van der Peet-Schwering CMC and Soede NM 2009. Group housing of sows in early pregnancy: a review of success and risk factors. *Livestock science 125:* 1-14

Stalder T and Kirschbaum C *In press.* Analysis of cortisol in hair – State of the art and future directions. *Brain, Behavior, and Immunity*

Stanton AL, Widowski TM, Kelton DF, Leslie KE, Leblanc SJ and Millman ST 2011. Prevalence of sickness behavior in neonatal dairy calves at a commercial heifer facility and associations with ADG and illness. *Proceedings of the 5th international conference of the assessment of animal welfare at farm and group level (WAFL)*, Guelph, Canada. Wageningen Academic publishers.

Stolba A and Wood-Gush DGM 1989. The behaviour of pigs in a semi-natural environment. *Animal Production 48:* 419

Straw BE, Zimmerman JJ, D'Allaire S and Taylor DJ, 2006. *Diseases of Swine*. Blackwell Publishing, Iowa.

Sundrum A and Rubelowski I 2001. The meaningfulness of design criteria in relation to the mortality of fattening bulls. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 48–52

Tosi MV, Canali E, Gregoretti L, Ferrante V, Rusconi C, Verga M and Carenzi C 2001. A Descriptive Analysis of Welfare Indicators Measured on Italian Dairy Farms: Preliminary Results. *Acta Agriculturae Scandinavica, Section A - Animal Science 51:* 69-72

Turner SP and Dwyer CM 2007. Welfare assessment in extensive animal production systems: Challenges and opportunities. *Animal Welfare 16:* 189-192.

Tuyttens FAM, Sprenger M, Van Nuffel A, Maertens W and Van Dongen S 2009. Reliability of categorical versus continuous scoring of welfare indicators: lameness in cows as a case study. *Animal Welfare 18:* 399-405

Veasey JS, Warran NK and Young RJ 1996. On comparing the behaviour of zoo housed animals with wild conspecifics as a welfare indicator. *Animal welfare 5:* 13-24

Veissier I, Forkman B and Jones RB 2007. Assuring animal welfare: from societal concerns to implementation. *Second Welfare Quality® stakeholder conference*. Berlin, Germany.

Viñuela-Fernández I, Jones E, Welsh EM and Fleetwood-Walker SM 2007. Pain mechanisms and their implication for the management of pain in farm and companion animals. *The Veterinary Journal 174:* 227-39

Von Borell E, Bockisch FJ, Buscher W, Hoy S, Krieter J, Muller C, Parvizi N, Richter T, Rudovsky A, Sundrum A and Van den Weghe H 2001. Critical control points for on-farm assessment of pig housing. *Livestock Production Science 72:* 177-184

Waiblinger S, Knierim U and Winckler C 2001. The development of an epidemiologically based on-farm welfare assessment system for use with dairy cows. *Acta Agriculturae Scandinavica, Section A - Animal Science 51*: 73– 77

Waiblinger S, Boivin X, Perdersen V, Tosi MV, Janczak AM, Visser EK and Jones RB 2006. Assessing the human animal relationship in farmed species : a critical review. *Applied Animal Behaviour Science 101:* 185–242.

Wasser SK, Hunt KE, Brown JL, Cooper K, Crockett CM, Bechert U, Millspaugh JJ, Larson S and Monfort SL 2000. A generalized fecal glucocorticoid assay for use in a diverse array of nondomestic mammalian and avian species. *General and Comparative Endocrinology 120:* 260–275

Webster AJF 2005a. *Animal Welfare. Limping towards Eden*. UFAW, Blackswell Publishing, UK

Webster AJF 2005b. The assessment and implementation of animal welfare: theory into practice. *Scientific and Technical Review of the Office International des Epizooties 24:* 723-734

Webster AJF, Main DCJ and Whay HR 2004. Welfare assessment: Indices from clinical observation. *Animal Welfare 13:* 93-98

Wemelsfelder F 1997. The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science 53:* 75–88

Wemelsfelder F 2007. How animals communicate quality of life: the qualitative assessment of behaviour. *Animal Welfare 16:* 25–31

Wemelsfelder F 2008. Curiosity, sentience, integrity: why recognising the “whole animal” matters. In: Boyle L, O’Connell N and Hanlon A (eds.), *Proceedings of the 42nd Congress for Applied Ethology*, pp. 1. Wageningen Academic Publishers, The Netherlands

Wewers ME and Lowe NK 1990. A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing and Health 13*: 227-236

Whay HR and Main DCJ 1999. "The way cows walk" steps towards lameness management. *Cattle Practice 7*: 357-364

Whay HR, Main DCJ, Greent LE and Webster AJF 2003a. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. *Animal Welfare 12:* 205-217

Whay HR, Main DCJ, Green LE and Webster AJF 2003b. Assessment of the welfare of dairy caftle using animal-based measurements: direct observations and investigation of farm records. *Veterinary Record 153:* 197-202

Whay HR, Main DCJ, Green LE and Webster AJF 2003c. An animal-based welfare assessment of group-housed calves on UK dairy farms. *Animal Welfare 12: 611–617*

Whay HR, Leeb C, Main DCJ, Green LE and Webster AJF 2007. Preliminary assessment of finishing pig welfare using animal-based measurements. *Animal Welfare 16:* 209-211

Williamson A and Hoggart B 2005. Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing 14*: 798-804

Winckler C, Capdeville J, Gebresenbet G, Hörning B, Roiha U, Tosi M, Waiblinger S 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. *Animal Welfare 12:* 619-624

Winckler C, Brinkmann J and Glatz J 2007. Long-term consistency of selected animal-related welfare parameters in dairy farms. *Animal Welfare 16:* 197-199

Wood MJ 1999. Pig assurance scheme – Comment. *Veterinary Record 145*: 531