**Supplementary material S1**

**General methodology**

*Utility function determination (MACBETH)*

MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) is a methodology which requires only qualitative judgements to quantify the relative attractiveness (utilities) of options (farms). To elicit a marginal utility function with MACBETH, the first step is to define whether the measure performs as a quantitative measure or as a qualitative one (Figure S1a) and which are the quantitative/qualitative performance levels of the measure (Figure S1b).

C:\Users\Paula\Documents\Trabajo\PhD\papers\Paper2\Animal\manuscript2\FigureS3a.tifFigure S1a MACBETH matrix of qualitative judgements for the measure castration. Definition of type of performance levels.



Figure S1b MACBETH matrix of qualitative judgements for the measure castration. Definition of the quantitative or qualitative performance levels of the measure.

The next step is to fill in a matrix, giving qualitative judgements regarding the difference of attractiveness between the different quantitative performance levels of the measure (Figure S1c). The qualitative judgements of difference can be rated as ‘very weak’, ‘weak’, ‘moderate’, ‘strong’, ‘very strong’ or ‘extreme’.



Figure S1c MACBETH matrix of qualitative judgements for the measure castration. Matrix filling giving qualitative judgements regarding the difference of attractiveness between the different quantitative performance levels of the measure.

As each judgement was given, the software automatically verified the matrix’s consistency (Figure S1d), and suggested judgement modifications which could be made to fix any detected inconsistency (Figure S1e).



Figure S1d MACBETH matrix of qualitative judgements for the measure castration. Automatic verification of the matrix’s consistency by the software as each judgement was given.



Figure S1e MACBETH matrix of qualitative judgements for the measure castration. Example of judgement modification suggestions proposed by the software in case of inconsistency.

From the complete and consistent matrix of judgements, MACBETH creates a numerical scale (Figure S1f).



Figure S1f MACBETH matrix of qualitative judgements for the measure castration. Numerical scale produced by MACBETH from the complete and consistent matrix of judgements.

With the numerical scale, MACBETH produces the marginal utility function (u) for each measure. This method also allows, in order to be able to aggregate the different measures into criteria, to normalise the raw data expressed in different scales into an absolute value scale, ranging, for example, between 0 and 100, corresponding 0 to the worst situation one can find on a farm and 100 the best situation.

*Aggregation with the help of the Choquet integral*

-Interaction index

According to the definition of Marichal, (2000), subtitutiveness between criteria can be understood as when a decision maker demands that the satisfaction of only one criterion produces almost the same effect than satisfaction of both. Of course, it is better that they be good on both directions, but it is less important. For instance, in this study and considering two criteria i and j, they would be regarded as substitutive when it is important that farms are good at criterion i or j, in other words, compensation is allowed between them, but they will be considered complementary when for the DM the satisfaction of only one criterion produces a very weak effect compared with the satisfaction of both.

-The concept of k-additivity

The number of variables involved in the CI increases exponentially with the coefficients which define a capacity. Let us consider a decision problem involving a set X of n elements, here (criteria). Defining a capacity on X requires the definition of coefficients. This could be too complex to handle if n goes beyond, say 8 (Grabisch, 1997). As a consequence it is frequent to consider that the capacity is additive, what identifies the Choquet integral with the weighted arithmetic mean (Marichal, 2000), and that can be defined with only n coefficients, at the price of a very poor modelling tool, avoiding in this way the complexity of using non-additive capacities but also losing their richness (Kojadinovic, 2007). The fundamental notion of *k-*additive proposed by Grabisch (1997) enables to find an intermediate solution between the complexity of representation and the richness of the model. *K*-additive measures for need less than coefficients to be defined. Only n coefficients are needed for (additive capacity), for , and in general for *k-*additive measures.

C:\Users\Paula\Documents\Trabajo\PhD\papers\Paper2\Animal\manuscript2\FigureS1.tif

Figure S2 Methodology followed in the Welfare Quality® project to aggregate the health measures into the health criterion (adapted from Welfare Quality, 2009).

C:\Users\Paula\Documents\Trabajo\PhD\papers\Paper2\Animal\manuscript2\FigureS2.tif

Figure S3 Methodologyproposed in this study to aggregate the Welfare Quality® health measures into the health criterion.

**References**

Kojadinovic I 2007. Minimum variance capacity identification. European Journal of Operational Research 177, 498-514.

Marichal JL and Roubens M 2000. Determination of weights of interacting criteria from a reference set. European Journal of Operational Research, vol. 124, no 3, 641-650.

Welfare Quality 2009. Welfare Quality® Assessment Protocol for Fattening Pigs*.* Lelystad: Wefare Quality® Consortium.