

Ovelhey, A<sup>(1,2)</sup>, Ruddat, I<sup>(1,2)</sup>, Kösters, S<sup>(1,2)</sup>, Scholz, B<sup>(3)</sup>, Kreienbrock, L<sup>(1,2)</sup>

<sup>1</sup>Department of Biometry, Epidemiology and Information Processing, University of Veterinary Medicine, Hannover, Germany,

<sup>2</sup>WHO Collaborating Centre for Research and Training in Veterinary Public Health,

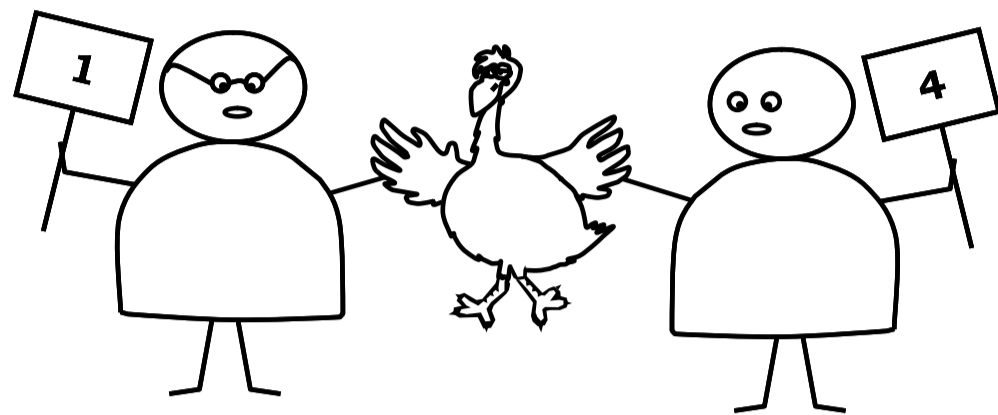
<sup>3</sup>Friedrich-Loeffler-Institute, Institute of Animal Welfare and Animal Husbandry, Celle, Germany

## Introduction

A network project was set up to improve small group housing systems for laying hens.

The effect of housing and management on behavioural traits and health of laying hens is investigated on 5 experimental stations.

Observers' agreement on scoring of different welfare-related variables was evaluated to minimize possible interviewer bias (1).

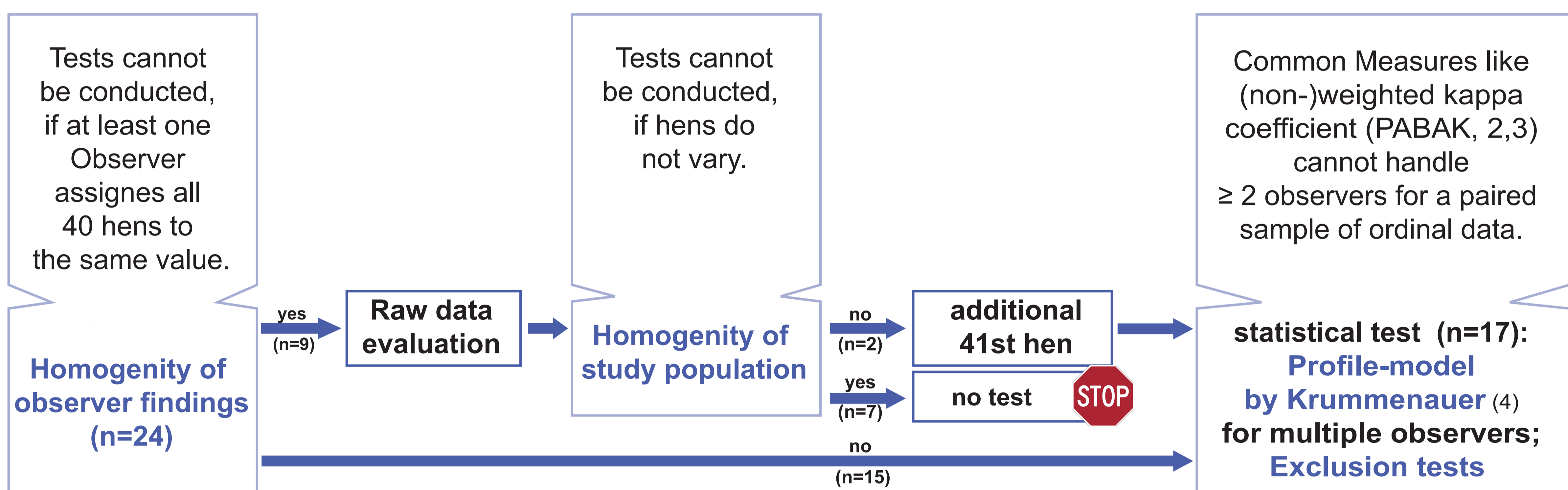


## Do observers agree on hen health? Evaluation of inter-observer reliability related to welfare parameters



**8 observers** rated a population of **40 hens** using a **scoring system** to quantify **feather quality**, **skin lesions** and **other health characteristics** developed in the LAYWEL EU-project. In total, 24 qualitative parameters were observed on ordinal, nominal and dichotomous scales.

## Drawbacks & Approaches



If more than 50% of the observers assigned at least 1 hen to the infrequent value, a 41st hen was added to the dataset in order to simulate obvious heterogeneity.

Sensitivity analysis did not reveal any effect of an additional, imaginary hen on estimates.

Profile-model handles multiple observers and produces 1 global kappa coefficient for each variable to measure the agreement based on all observers simultaneously and 8 observer-specific kappa coefficients. Exclusion tests check whether the global measurement was significantly affected by one specific observer.

## Results & Discussion

### Overall, inter-observer reliability is greater than caused by accident (see fig. 1).

Disagreeing observers were informed about the requirement on adapting their scoring behaviour related to the specific parameter.

For some variables the number of missing values has to be reduced in the future.

Rating of hyperkeratosis / lesions at food pads and toes was influenced by the fact that examinations resulted in an aggravation of symptoms.

The overall impression of plumage condition does not seem to be an objective method to rate the health status of hens up to now.

### Measures to avoid homogeneity of observer findings

- Before the next reliability test a training of all observers should be conducted.
- Definitions of parameter values should be discussed and specified if necessary.

### Measures to avoid homogeneity of study population

For the next reliability test a heterogeneous group of hens is preferable though not realisable due to homogenous population within one housing system and laying period.

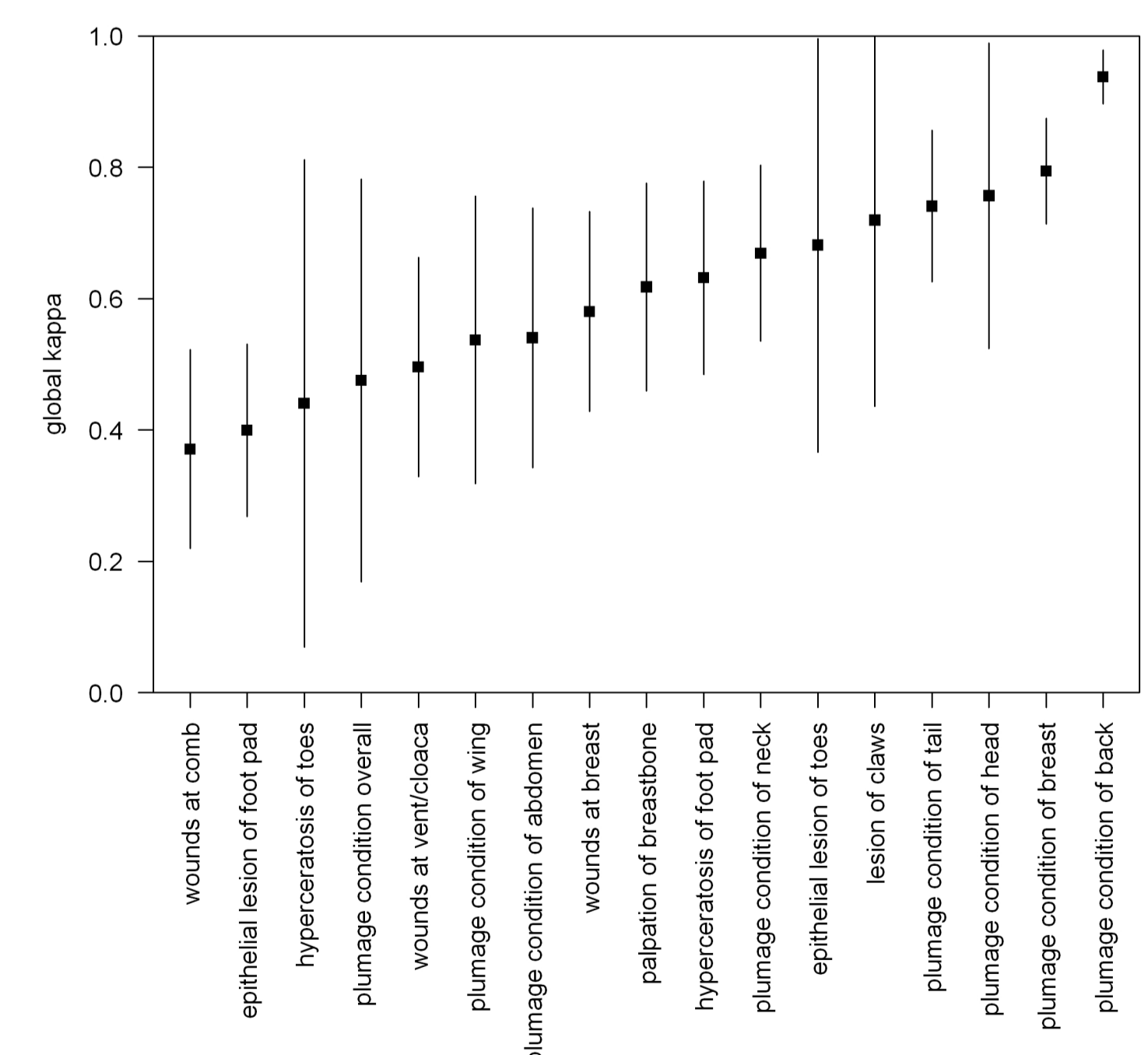


Figure 1: Variable specific Global kappa indices on the inter-observer reliability (possible values from -1 to 1)

- Thrusfield, M., 2007. Veterinary Epidemiology. Blackwell Publishing.
- Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. Journal of clinical epidemiology 46, 423-429.
- Fleiss, J., Levin, B., Myunghee, C.P., 2003. Statistical Methods for Rates and Proportions. John Wiley & Sons New York.
- Krummenauer, F., 2000. Statistische Methoden zum Vergleich diagnostischer Verfahren von begrenzter Reproduzierbarkeit. Fachbereich Medizin J.Gutenberg-Universität, Mainz.