Multiple regression and mediator variables can be used to avoid double counting when economic values are derived using stochastic herd simulation

S. Østergaard[1,2], J. F. Ettema[1,2], L. Hjortø[3], J. Pedersen[3] and M. Kargo[1,3]
[1] Aarhus University, [2] SimHerd Inc., Denmark, [3] Knowledge Centre for Agriculture, Denmark

ABSTRACT: Multiple regression and model building with mediator variables was addressed to avoid double counting when economic values are estimated from data simulated with herd simulation modeling (using the SimHerd model). The simulated incidence of metritis was analyzed statistically as the independent variable, while using the traits representing the direct effects of metritis on yield, fertility and occurrence of other diseases as mediator variables. The economic value of metritis was estimated to be €78 per 100 cow-years for each 1% increase of metritis in the period of 1-100 days in milk in multiparous cows. The merit of using this approach was demonstrated since the economic value of metritis was estimated to be 81% higher when no mediator variables were included in the multiple regression analysis.

Keywords: dairy cattle; economic values; herd simulation

Introduction

The breeding goal defines which traits are to be improved and how much weight is given to each trait (Groen et al. 1997). In dairy cattle breeding, the weighting factors are usually economic values that are based on a model of a dairy herd production system.

The economic values are marginal economic values, which means the economic value of one unit improvement of the trait – keeping the remaining traits constant. However, this is not straight forward as relationships among traits exist. To avoid double counting, the economic value of a given trait should therefore not include the indirect effects of other traits. For example, reduced milk production as a result of metritis should not be included in the economic value of metritis.

Relationships between longevity, functional traits and milk yield traits are especially difficult to handle because longevity is highly affected by improvements of these other traits. If improvement of other traits decreases the culling rate or enables another focus on selection, then this has an economic value. If these structural relationships are not taken into account in the derivation of economic values for functional traits and milk yield, then one could argue that the economic values of the functional traits and milk yield are underestimated, and that the economic value of longevity is overestimated when interpreting longevity as residual longevity. Wolfova and Wolf (2013) suggest that relationships among traits should be taken into account when these relationships are structural and caused by changes in e.g. the age structure of the herd but not when they are caused by genetic correlations between traits. The structural interactions can for a great part be taken into account by using mechanistic, dynamic and stochastic dairy herd models that simulate the production and state changes of dairy cows and young stock. However, phenotypic correlations and thereby genetic correlations are also inherent in this type of model. Therefore, it is problematic to keep all other traits constant when one trait (the trait of interest) is changed without violating the mechanistic and dynamic modeling of the structural effects as they exists in real herds. The phenotypic relationships between traits are mediating the effect of the trait of interest on herd economics in such a model. However, the only phenotypic relationships that should be included in the estimation of economic values are the ones causing structural changes in the herd. A method for correcting for these mediator effects in the simulated output, while keeping the structural relationships in the simulation model, could be to use multiple regression and the use of mediator variables (Fairchild and Mackinnon, 2009).

Based on this background information, we hypothesize that multiple regression and mediator variables can be used to avoid double counting when economic values are estimated using stochastic herd simulation. This will be illustrated using metritis as a model trait

Materials and Methods

Herd simulation model. In order to estimate EV that include structural herd effects, we addressed the method of stochastic, dynamic and mechanistic herd simulation modeling. The existing SimHerd model simulates the production and state changes of a dairy herd including young stock and has been used to study various herd management tasks (Østergaard et al., 2003; Kristensen et al., 2008), the implication of genetic trend for the effect of reproduction management (Ettema et al. 2011), and the derivation of economic values of production and functional traits (Nielsen, 2004). For the application in this study, the SimHerd model was further developed to generate more detailed output. For example, the sub model for metritis is a logistic regression model with risk factors, which in each time step for each cow generates a cow-specific risk to get metritis. In case metritis occurs, it affects the production, health and reproduction of this cow and thereby the herd structure due to relationships among traits and due to between-cow dependencies caused by herd level management decisions.

Simulated scenarios. Default SimHerd input parameters and the applied prices were intended to represent a typical Danish loose housing system and management strategy for a dairy herd with the Holstein breed. In order to simulate within a relevant area of the response surface and to be able to distinguish between direct and the structural effects on the response, we simulated separate scenarios for each trait. Each scenario was simulated over 20 years and in 1000 replicates. For each replicate, the SimHerd variable that corresponds to the trait of interest was increased by 1/1000 of the range of interest. For metritis, we increased the intercept parameter of the logistic sub-model for metritis risk from zero to 18% per cow-year.

Statistical analyses. Output data from each scenario was analyzed separately. Initially, we calculated the mean over simulation years 11 to 20 for each simulated replicate. These data was analyzed by first order multiple regression. In all models, we analyzed the herd level response variable ‘annual net return over operation costs’. With regard to relationships among traits, we on the one hand want to use herd simulation to include them and their structural effects, and on the other hand we want to exclude mediator effects from traits that have their own EV. To control for the latter ‘double counting’ phenomenon, our model building addressed the principles of mediator variables as illustrated in Fig. 1.

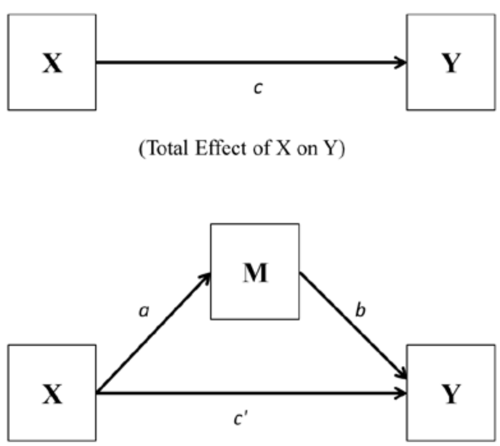

Fig 1. Path diagram for the single-mediator model. X = the independent variable, Y = the dependent variable, and M = the mediating variable. The mediation model decomposes the total effect of X on Y (c), into two parts: the indirect effect of X on Y, quantified by ab (the product of a and b), and the direct effect of X on Y with the effect of the mediator removed, quantified by c’. c=ab+c’ (Fairchild and MacKinnon, 2009).

The independent variable should represent the trait itself or a proper indicator for the trait under study in the specific scenario. The mediator variables were those variables that met two criteria: they represent the direct effects of the study trait, as they are modeled in the SimHerd model, and they have their own EV. The mediator variables should not remove the structural effects caused by the trait under study. To avoid that, we addressed only one cow-parity-DIM-specific independent and mediator variable per trait. Thereby, structural effects could not be represented by linear combinations of dependencies between parity and lactation group. In another words: the reason that cow level estimates need to be estimated by parity is that 1) structural effects will affect the distribution of primiparous and multiparous cows, and 2) many observed cow events and performances are significantly different for primiparous and multiparous cows, e.g. milk yield level and disease incidence. The same goes for different lactation stages. The genetic correlations between traits, between and within parity, are however most often high, and therefore it is acceptable to estimate the economic values for the specific parity group that is expected to be the most important.

Traits characterizing repeated animal performance, traits expressed in different lactations, and traits from a group of traits characterizing a specific biological potential were also to a large extend controlled for by the abovementioned modeling and the one cow-parity-DIM variable per trait. For metritis, this implies that we use ‘metritis cases per 0 to 100 DIM periods for multiparous cows’ as the independent indicator variable. As mediators we included one multiparous cow-parity-DIM-specific variable for milk yield, conception rate, ketosis, and displaced abomesum. Milk withdrawal was not included, although this is a direct effect of metritis. Milk withdrawal however, does not have its own EV.

To analyze the implication of the abovementioned concerns, we analyzed a number of statistical models that differed in which independent and mediator variables were included.

Results and Discussion

Results for the default herd. Key figures for the simulated default herd included 200.1 cow-years, a 35.7% replacement rate, 0.06 dead cows per cow-year, 1.11 calvings per cow-year, a 0.41 conception rate, a 0.40 estrus detection rate, 0.18 metritis cases per cow-year, 9600 kg ECM per cow-year, € 0.39 per kg ECM in milk payment, and € 1951 per cow-year in net return.

EV for metritis. Average key figures from the replicates simulated with increasing (from zero) risk of metritis included 0.093 metritis cases per cow-year, 9615 kg ECM per cow-year, €0.39 per kg ECM in milk payment, and €1963 per cow-year in net return.

Table 1 shows the economic values for metritis in the different models. Our default model (model 1) included the independent variable ‘metritis cases per period of 1-100 DIM in parity 2+’ (Me2) and the mediator variables ‘Avg. daily MY 1-168 DIM in parity 2+’ (MY2), ‘Conception rate in parity 2+’ (CR2), ‘ketosis cases per period of 1-100 DIM in parity 2+’ (Ke2) and ‘displaced abomasum cases per period of 1-100 DIM in parity 2+’ (DA2). In this model, the estimate for Me2 was 78 ct. This is the estimated reduction in annual net return per cow-year for each 1% increase in metritis cases per period of 1-100 DIM in parity 2+. In a 100-cow herd, this implies that the one extra case of metritis in the period of 1-100 DIM in parity 2+ reduces the annual net return by €78. This estimate also includes losses from cases in other parities and lactation periods.

can hardly be defined independently of herd dynamic effects.

Table 1. Models and results for the economic value for the trait 'metritis cases 1-100 DIM in parity 2+' in € cents, depending on the mediator variables included in the analysis.

Model No.	Mediator variables [¥]	Economic value (SE)	Model R^2
1	MY2, CR2, Ke2, DA2	78 (11)	0.51
2	No mediator variables	141 (9)	0.17
3 [§]	MY1, CR1, Ke1, DA1, MY2, CR2, Ke2, DA2	17 (26), 79 (30)	0.56
4	MY2, CR2, Ke2, DA2, CUL	64 (8)	0.72
5	MY2, CR2, Ke2, DA2, WD	23 (13)	0.53
6	MY1, CR1, Ke1, DA1, MY2, CR2, Ke2, DA2	59 (11)	0.56
7	MY, CR, Ke, DA	77 (11)	0.64
8	Remove mediator effects	86 (9)	0.07

[§] This model has independent variables for both 'metritis cases 1-100 DIM in parity 1' (Me1) and Me2.
[¥] MY1/MY2 = Avg. daily MY 1-168 DIM in parity 1 / 2+
CR1/CR2 = Conception rate in parity 1 / in parity 2+
Ke1/Ke2 = Ketosis cases 1-100 DIM in parity 1 / parity 2+
DA1/DA2 = Displaced abomasum cases 1-100 DIM in parity 1 / parity 2+
KE/DA = As above but per cow-year (herd level incidence)
CR = Conception rate calculated at herd level
CUL = Replacement rate and (in)voluntary culling
WD = Milk withdrawal

The estimate of the EV for Me2 from model 1 does deliberately not include the direct effect of metritis on milk yield, conception rate, ketosis, and displaced abomasum (as they have an EV of their own). The implication of not correcting for these by mediator variables is seen from model 2 with no mediator variables, where the estimate is 81% higher (141 ct).

The implication of only one cow-parity-DIM-specific independent and mediator variable per trait is seen from model 6 (trait mediator variables for each parity), where the EV estimate for ME2 was reduced by 24% (59 ct in model 6) and from model 7 (herd level mediator variables; ketosis cases per cow-year (KE) and displaced abomasum cases per cow-year (DA) and conception rate (CR)), where the estimate was reduced by 2% (77 ct in model 7). The latter reduction can only be related to herd dynamic effects that we did not want to remove from the estimate of the EV for the metritis trait.

To explore the implication of removing the indirect effect of metritis on longevity, we added mediator variables for replacement rate, involuntary culling and voluntary culling. This reduced the EV of metritis by 17% (64 ct in model 4). This reduction relates to the direct effect of metritis on culling and the corresponding structural changes, which we did not want to remove from the estimate of EV for the metritis trait. This also explains the higher R-square from this model (R-square increased from 0.51 to 0.72). There are arguments that longevity should have its own EV as a 'residual' cow problem. However, it can hardly be defined independently of herd dynamic effects.

Model 1 did not include a mediator variable for 'milk withdrawal per cow-year', despite the fact that this is modeled as a direct effect of metritis in the SimHerd model. If we had included milk withdrawal, it would reduce the estimated EV by 76 % (23 ct in model 5). This reduction is mainly, but not only, due to removing the direct effect of milk withdrawal after metritis. It is also to some extent due to structural herd effects, as the milk withdrawal variable was defined per cow-year.

In model 3 we estimated the EV for metritis both primiparous and multiparous cases to be 17 ct and 79 ct, respectively. As these two variables are highly correlated, it is quite arbitrary how the values are distributed between the two.

We compared the approach with multiple regression and mediator variables with the approach where we removed those relationships between metritis and the other traits in the SimHerd that correspond to our mediator variables. The estimate of EV for Me2 from this latter approach using a statistical model without mediator variables was 86 ct (model 8).

In the simulations we assumed treatment costs of €33.6 per metritis case. By changing this to zero, the estimated EV of Me2 was reduced in the order of 38 ct in the various models.

Production system. In this study, we addressed a certain production system. The approach in this study can be reproduced to estimate how the economic values differ between production systems. This can be done by setting up the SimHerd model to represent alternative production systems (like low-input, high tech and organic production systems) and thereby explore the relevant response surface for each trait in each system.

Conclusion

We found significant merit in using multiple regression and mediator variables to avoid double counting when economic values are calculated using stochastic herd simulation.

Literature Cited

Ettema, J.F., Østergaard, S., Sørensen, M. K. (2011). Animal, 5 (12): 1887-97.

Fairchild, A.J., MacKinnon, D.P. (2000). Prev. Sci. 10 (2): 87-99.

Groen, A.F., Steine, T., Colleau, J.J. et al. (1997). Livest. Prod. Sci. 49:1-21.

Kristensen, E.L., Østergaard, S., Krogh, M.A. and Enevoldsen, C. (2008). J. Dairy Sci. 91: 620-631

Nielsen, H.M. (2004). Ph.D thesis. The Agricultural Universty, Copenhagen. 156 pp.

Østergaard, S., Sørensen, J.T., Houe, H. (2003). Prev. Vet. Med. 58: 125-143.

Wolfova M. and Wolf J. (2013). Animal, 7 (9): 1401-1413.