Preventive Veterinary Medicine 147 (2017) 163–171

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Implementing structural equation models to observational data from feedlot production systems

Elva Cha[a,b], Mike Sanderson[a,b], David Renter[a,b], Abigail Jager[c], Natalia Cernicchiaro[a,b], Nora M. Bello[b,c,*]

[a] Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS, USA
[b] Center for Outcomes Research and Epidemiology (CORE), College of Veterinary Medicine, Kansas State University, Manhattan, KS, USA
[c] Department of Statistics, College of Arts and Sciences, Kansas State University, Manhattan, KS, USA

ARTICLE INFO

Keywords:
Bovine respiratory disease
Structural equation modeling
Direct and indirect effects
Feedlot cattle

ABSTRACT

The objective of this study was to illustrate the implementation of a mixed-model-based structural equation modeling (SEM) approach to observational data in the context of feedlot production systems. Different from traditional multiple-trait models, SEMs allow assessment of potential causal interrelationships between outcomes and can effectively discriminate between direct and indirect effects. For illustration, we focused on feedlot performance and its relationship to health outcomes related to Bovine Respiratory Disease (BRD), which accounts for approximately 75% of morbidity and 50–80% of deaths in feedlots. Our data consisted of 1430 lots representing 178,983 cattle from 9 feedlot operations located across the US Great Plains. We explored functional links between arrival weight (AW; $i = 1$), BRD-related treatment costs (Trt$; as a proxy for health; $i = 2$) and average daily weight gain (ADG; as an indicator of productive performance $i = 3$), accounting for the fixed effect of sex and correlation patterns due to the clustering of lots within feedlots. We proposed competing plausible causal models based on expert knowledge. The best fitting model selected for inference supported direct effects of AW on ADG as well as indirect effects of AW on ADG mediated by Trt$. Direct effects from outcome *i'* to outcome *i* are quantified by the structural coefficient $\lambda_{ii'}$, such that every unit increase in kg/head of AW had a direct effect of increasing ADG by approximately (estimate ± standard error) $\hat{\lambda}_{31} = 0.002 \pm 0.0001$ kg/head/day and also a direct effect of reducing Trt$ by an estimated $\hat{\lambda}_{21} = \$0.08 \pm 0.006$ USD per head. In addition, every $1 USD spent on Trt$ directly decreased ADG by an estimated $\hat{\lambda}_{32} = 0.004 \pm 0.0006$ kg/head/day. From these estimates, we show how to compute the indirect, Trt$-mediated, effect of AW on ADG, as well as the overall effect of AW on ADG, including both direct and indirect effects. We further compared estimates of SEM-based effects with those obtained from standard linear regression mixed models and demonstrated the additional advantage of explicitly distinguishing direct and indirect components of an overall regression effect using SEMs. Understanding the direct and indirect mechanisms of interplay between health and performance outcomes may provide valuable insight into production systems.

1. Introduction

Structural equation models (SEM) constitute a multivariate statistical framework that can be used to explore the interrelated, and potentially causal, interplay between multiple outcomes in complex biological systems (Rosa et al., 2011). For this reason, SEMs can serve as valuable quantitative tools to enhance insight into agricultural production systems. Furthermore, under certain circumstances, SEMs can posit causal inference from the type of observational data that is often collected in commercial livestock operations, such as feedlots (Rosa and Valente, 2013). Despite being increasingly used in animal genetics and genomics genetics (de los Campos et al., 2006; Varona et al., 2007; de Maturana et al., 2009; Penagaricano et al., 2015a), SEMs have been implemented only sparingly in the context of veterinary epidemiology, primarily in dairy applications (Detilleux et al., 2012; Rehbein et al., 2013; Detilleux et al., 2015). A focus on modeling management practices, demographics and other environment-based factors is key to assess risk factors in terminal production systems that will not generate progenitors for the next generation, such as feedlots.

In feedlot production systems, Bovine Respiratory Disease (BRD) is a multifactorial syndrome with substantial economic impact, accounting for approximately 75% of feedlot morbidity, and 50–80% of feedlot

* Corresponding author at: 002 Dickens Hall, Manhattan, KS, 66506, USA.
E-mail address: nbello@ksu.edu (N.M. Bello).

http://dx.doi.org/10.1016/j.prevetmed.2017.09.002
Received 9 May 2017; Received in revised form 9 August 2017; Accepted 3 September 2017

deaths (Edwards, 1996; Chirase et al., 2001), costing the US beef industry approximately $4 billion annually (Griffin, 1997). This economic impact has likely increased in recent years and can be explained primarily by expenses associated with prevention, treatment, culling and death, in addition to negative ramifications to the health and well-being of millions of cattle (Martin et al., 1989; Smith, 1998; Gardner et al., 1999). Further, cattle affected with BRD show poor productive performance and impaired carcass quality characteristics, including quality and yield grades as well as average daily gain (ADG) and net returns (Wittum et al., 1996; Babcock et al., 2009; Cernicchiaro et al., 2013). Epidemiological risk factors for BRD are well documented. For example, management practices such as mixing gender groups, intermingling cattle from multiple sources and increasing shipping distance were associated with an increased risk of initial respiratory morbidity, whereas BRD-related risks were decreased for cattle with heavier arrival weights (Sanderson et al., 2008; Hay et al., 2014).

Modern statistical methods have proven critical to advance such epidemiological insight. Specifically, linear mixed models (Stroup, 2013) have enabled joint assessment of multiple risk factors while accounting for the often-hierarchical nature of the data and ensuing correlation patterns. These patterns are due to the clustered nature of feedlot observations whereby cattle are clustered within pens, which in turn are clustered within operations and even geographical regions (Dohoo et al., 2001; Dohoo et al., 2009). Indeed, standard linear mixed models are well suited to study the behavior of single-response processes under these conditions (Stroup, 2013). However, these models are unable to partition overall effects and discriminate indirect effects through mediating variables. This limitation impairs insight into the mechanisms by which performance and health outcomes might influence each other within animal production systems (Rosa and Valente, 2013). It is under this umbrella that SEMs provide an opportunity for additional insight.

First introduced by Wright (1934) in the context of path analysis, SEMs were initially used mostly in the social sciences and economics (Wright, 1934; Wei et al., 2003; Kwon and Bessler, 2011; Toma et al., 2013). Briefly, the workings of SEMs are characterized by two fundamental entities rooted in the interphase between graph theory and probability theory (Pearl, 2009). First, the hypothesized causal processes between outcomes are articulated in graphical form using directed graphs. Next, graphical connections are expressed probabilistically in the form of a statistical model that specifies the joint probability distribution needed to support estimation and inference on the hypothesized process. The link from graphical theory and causal structures to probabilistic modeling and statistical association relies on the concept of directed-separation (or d-separation) (Pearl, 2009), such that, under certain non-trivial assumptions, directed graphs can be aligned 1-to-1 with the expected conditional independencies in the joint probability distribution specified by a statistical model (Pearl, 2009). It was only recently that SEMs were adapted to a mixed models framework (Gianola and Sorensen, 2004), thereby accommodating correlation patterns between observations, as induced by data architecture (e.g., clustering, nesting) as well as the potential for multilevel correlations between outcomes, as is commonly found in animal agriculture data.

Our objective in this study was to illustrate the implementation of a mixed-model-based structural equation modeling framework in the context of feedlot production medicine. In particular, we assess potential causal interrelationships between health and performance outcomes of feedlot cattle and show how SEMs can distinguishing between direct and indirect components of an overall relationship between outcomes, thereby enabling quantification of mediating effects. Further, we compare SEM-based results and their interpretation with those obtained from standard linear mixed models methodology.

2. Materials and methods

2.1. Data description

For consistency with our previous work (Babcock et al., 2009; Cernicchiaro et al., 2012) and in order to focus on the target population of commercial cattle at risk for BRD, we define lots as cohorts of cattle that were similarly managed and implement the following inclusion criteria in this study: 1) lots of *Bos taurus* cattle excluding Holstein breed, 2) lots consisting of at least 20 cattle, 3) lots of either steers or heifers (no mixtures), 4) lots with average arrival weight between 136 and 398 kg/head, 5) lots with at least 45 days on feed (DOF) recorded, and 6) lots with positive recorded average daily weight gains (i.e., > 0 kg/head/day). After editing, the data used for analyses consisted of complete observations (i.e. no missing data) collected on 1430 lots representing 178,983 cattle from 9 feedlots located across the US Great Plains during 2014–2015. All feedlots were managed similarly in open drylot pens and fed a diet common to feedlots in the Great Plains.

For each lot, outcomes of interest included y_1 = arrival weight (AW; expressed in kg/head), y_2 = BRD-related treatment costs (Trt$; expressed in USD/head) and y_3 = average daily gain (ADG; expressed in kg/head/day), calculated as follows: AW = arrival weight of the lot/number of head in the lot at arrival; Trt$ = (lot-level summation of all BRD-related treatment costs incurred on individual animal and lot-level treatments)/number of head in the lot at arrival; ADG = [(Final weight of the lot – arrival weight of the lot)/(total number of head-days on feed for the lot)]. Descriptive statistics for each outcome are presented in Table 1. The outcome labeled Trt$ was considered a proxy for morbidity and mortality risk to BRD. In this context of these data, lots with lower values of Trt$ can be reasonably assumed to have fewer cattle that received treatment for BRD. Although no standardized case definition for BRD was used across feedlots, all feedlots were under common management and utilized the same consulting veterinary group who were responsible for training feedlot crews on detection of BRD. An individual animal was treated for BRD if it presented clinical signs attributable to respiratory disease including labored breathing, nasal or ocular discharge, depression and decreased abdominal fill. Animals identified with these signs were individually examined and treated. Treatment protocols for BRD differed across feedlots but included macrolides, penicillins, oxytetracyclines, flouroquinolones, florfenicol, and neomycin.

Table 1
Descriptive statistics for data used in this study, corresponding to 1430 lots from 9 feedlots located across the US Great Plains during 2014–2015 and representing 178,983 cattle.

Lot-level variables	N	Mean	Min	25th percentile	Median	75th percentile	Max
Arrival weight (AW; kg/head)	1430	320.9	144.3	290.3	332.7	363.6	397.6
BRD treatment costs (Trt$;USD/head)	1430	7.4	0.0	0.0	0.5	4.4	91.7
Average daily gain (ADG; kg/head/day)	1430	1.50	0.06	1.40	1.54	1.70	2.30
Lot size (number of head)	1430	125.2	20.0	73.0	105.0	167.0	607.0
Days on feed (days)	1430	159.2	45.0	135.8	150.3	174.1	320.3

BRD: Bovine respiratory disease.