



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

## **Random Sampling of Beef Cattle for Genetic Testing: Optimal Sample Size Determination**

Nathanael M. Thompson  
Graduate Research Assistant  
Department of Agricultural Economics  
Oklahoma State University  
[nathan.thompson10@okstate.edu](mailto:nathan.thompson10@okstate.edu)

B. Wade Brorsen  
Regents Professor and A.J. and Susan Jacques Chair  
Department of Agricultural Economics  
Oklahoma State University  
[wade.brorsen@okstate.edu](mailto:wade.brorsen@okstate.edu)

Eric A. DeVuyst  
Professor  
Department of Agricultural Economics  
Oklahoma State University  
[eric.devuyst@okstate.edu](mailto:eric.devuyst@okstate.edu)

Jayson L. Lusk  
Regents Professor and Willard Sparks Endowed Chair  
Department of Agricultural Economics  
Oklahoma State University  
[jayson.lusk@okstate.edu](mailto:jayson.lusk@okstate.edu)

*Selected Paper prepared for presentation at the Southern Agricultural Economics Association's  
2016 Annual Meeting, San Antonio, Texas, February, 6-9, 2016*

*Copyright 2016 by N.M. Thompson B.W. Brorsen, E.A. DeVuyst, and J.L. Lusk. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

# **Random Sampling of Beef Cattle for Genetic Testing: Optimal Sample Size Determination**

## **Abstract**

Sample size is often dictated by budget and acceptable error bounds. However, there are many economic problems where sample size directly affects a benefit or loss function, and in these cases, sample size is an endogenous variable. We introduce an economic approach to sample size determination utilizing a Bayesian decision theoretic framework that balances the expected costs and benefits of sampling using a Bayesian prior distribution for the unknown parameters. To demonstrate the method for a relevant applied economics problem, we turn to randomly sampling beef cattle for genetic testing. A theoretical model is developed, and several simplifying assumptions are made to solve the problem analytically. Data from 101 pens (2,796 animals) of commercially-fed cattle are then used to evaluate this solution empirically. Results indicate that at the baseline parameter values an optimal sample size of  $n^* = 10$  out of 100 animals generate returns from sampling of nearly \$10/head, or a return-on-investment of 250%. Therefore, a large portion of the additional value for higher-quality cattle can be captured by testing a relatively small percentage of the lot. These results vary depending on the actual quality (or profitability) of a particular pen of cattle, the homogeneity within the pen, the variance of the buyer's subjective prior distribution of expected profit, and the per-head cost of genetic testing. Nonetheless, results suggest that random sampling has the potential to provide a context in which the benefits of genetic testing outweigh the costs, which has not generally been the case in previous research.

**Keywords:** Bayesian decision theory, beef cattle genetics, random sampling, sample size determination

The methods for determining sample size can be generally classified into two broad categories: frequentist and Bayesian (Adcock 1997). The debate between proponents of these two approaches has occupied entire issues of statistical journals (for example, *Journal of the Royal Statistical Society: Series D (The Statistician)* 46(2), 1997) and is ongoing. In practice, sample size has often been a function of budget and acceptable error bounds. Therefore, the commonly used frequentist approach determines sample size by specifying a null and alternative hypothesis for the parameter of interest and using predetermined specifications of size, power, variance, and a minimum detectable difference (Adcock 1997; Wilan 2008). However, there are many economic problems where sample size directly affects a benefit or loss function. In these cases, sample size is an endogenous variable that should be considered jointly with other choice variables in an optimization problem. In this article we introduce an economic approach to sample size determination utilizing a Bayesian decision theoretic framework. The Bayesian framework for sample size determination appears to be rarely used in economic research, but is arguably the theoretically sound approach to determining sampling for many applied economic problems. In addition to introducing the method, we develop a practical, currently relevant application of the model for randomly sampling beef cattle for genetic testing.

The issue of endogenous sample size was first identified by Grundy, Healy, and Rees (1956). In this seminal piece, the authors acknowledge that in order to determine the economically justifiable amount of experimentation, the costs of the experiment must be set against the potential benefits of the new process being evaluated. However, they point out that the main difficulty is that the expected benefits of the new process depend on the outcome of the experiment. Their model has since been generalized and further developed by Riffa and Schlaifer (1961) and, more recently, by Lindley (1997), and has come to be known as the Bayesian

decision theoretic approach to sample size determination. In this fully Bayesian framework, the economically-optimal sample size is determined using an objective function to balance the expected costs and benefits of sampling using a Bayesian prior distribution for the unknown parameters.

Despite its potential for application to a variety of research problems, to date, the use of this method has been largely limited to clinical trials in medical research (Gittins and Pezeshk 2000, 2002; O'Hagan and Stevens 2001; Kikuchi, Pezeshk, and Gittins 2008; Wilan 2008; Willan and Pinto 2005, 2006) and substantive tests in financial auditing (Smith 1976, 1979; Patterson, 1993; Laws and O'Hagan 2002). Still, there remain many research problems to which this method could, and likely should, be applied for determining the economically-optimal sample size. For example, in a variety of audit/inspection contexts (for example, internal bank audit, environmental regulation compliance audit, Food and Drug Administration [FDA] inspections, health inspections, etc.), it is much too costly to examine each individual unit of interest. However, there is significant value associated with obtaining a sufficient sample to identify potential losses or costly, undesirable outcomes. Therefore, the fully Bayesian approach allows the decision maker to determine the sample size that will balance the expected costs and benefits of sampling. In addition, a sampling error of  $\pm 3\%$  in many survey based research methods is dictated by convention. However, researchers rarely discuss, or even consider, the tradeoff between the cost and accuracy of a given survey question. Likewise, issues of quality control/statistical tolerancing or sampling related problems (for example, grain sampling, livestock sampling, soil sampling, etc.) could also benefit from this approach.

To demonstrate the method for a relevant applied economics problem, we turn to genetic testing for market livestock. Recent advancements in genomic technology have the potential to

generate value throughout the beef industry (Van Eenennaam and Drake 2012). Previous research has shown significant differences in the profitability of animals with different genetic profiles suggesting some merit in using the tests for selection (Lusk 2007; Thompson et al. 2014). However, economic evaluations of commercially available genetic marker panels have indicated that the value of this information for feedlot management is generally not enough to offset the current cost of genetic testing (about \$40/head; Igenity 2015). For example, the value of using genetic information for sorting feedlot cattle by optimal days-on-feed is less than \$3/head (DeVuyst et al. 2007; Lusk 2007; Lambert 2008; Thompson et al. 2014), and the value of using this information to selectively market (live weight, dressed weight, or grid pricing) fed cattle is less than \$13/head (Thompson et al. 2016).

Therefore, in order to achieve a scenario in which genetic testing is cost-effective, the value of genetic information must increase or the cost of testing must decrease. While animal scientists are continually progressing towards providing more accurate genetic markers that have the potential to increase the value of genetic information (for example, see Akanno et al. 2014), producers seeking to use this technology have no control over the pace at which these new variations are released. So, we introduce a strategy for reducing the overall cost of genetic testing that has been previously discussed, but has yet to be evaluated: random sampling (Thompson et al. 2014, 2016). That is, instead of testing each individual animal in a group of cattle, a random sample of animals could be tested to measure the genetic potential of the group. While appealing in theory, there is a thorny practical question: “What size sample should I take?” To answer this question we use a Bayesian decision theoretic approach to determine the economically-optimal sample size.

Unlike previous research evaluating the value of genetic information, which has generally focused on marker-assisted management at the feedlot stage, in this research we approach the problem from the feeder cattle producer's perspective. That is, producers who know that their feeder cattle have high-value genetics may want to try to convince buyers (for example, wheat stocker producers or feedlot owners) that their cattle are higher quality in order to receive a premium. However, in order to establish the actual genetic makeup of a lot of feeder cattle the seller must incur the cost of genetic testing. Therefore, the objective of this research is to determine if randomly sampling a group of feeder cattle for genetic potential is cost-effective and, if so, to determine the economically-optimal sample size. A general framework is introduced, and a theoretical model specific to randomly sampling feeder cattle for genetic testing is developed. Several simplifying assumptions are made to solve the problem analytically, and the solution is then evaluated empirically using data from 101 pens (2,976 animals) of commercially-fed cattle. After estimating the optimal sample size and the returns from sampling, sensitivity analysis is conducted to evaluate the robustness of these results to varying levels of quality and homogeneity of a particular lot of cattle. Our results suggest sampling could be a viable strategy to reduce the costs of genetic testing for beef cattle.

### **Conceptual Framework**

The objective of the Bayesian decision theoretic approach to economically-optimal sample size determination is to determine how large of a sample ( $n$ ) to take from a population to make an inference/decision about some feature of the population considering that both the costs and benefits of sampling are a function of  $n$  (Grundy, Healy, and Rees 1956; Riffa and Schlaifer 1961; Lindley 1997). The framework for the fully Bayesian treatment of this problem is set out in detail by Riffa and Schlaifer (1961) and was later updated by Lindley (1997). Both

descriptions approach the problem in temporal order: first the sample size  $n$  is chosen,  $n$  realizations of random quantity  $X$  are then collected  $(x_1, x_2, \dots, x_n)$  where the density of  $X$  is of a known form dependent on the unknown parameter  $\theta$ , the information from this sample is used to make decision  $d$  concerning some feature of the population, and finally the unknown parameter  $\theta$  is considered. The merit of this decision sequence is captured by specifying a utility function,  $u(n, x, d, \theta)$ .

Before performing the optimization the decision maker has a prior distribution of the unknown parameter  $\theta$ ,  $p_{prior}(\theta)$ . An application of Bayes rule to  $\theta$  and  $x$  gives the posterior distribution of  $\theta$  conditional on  $x$  and  $n$ ,  $p_{post}(\theta|x, n) = p(x|\theta, n)p_{prior}(\theta)/p(x|n)$ , where  $p(x|\theta, n) = \prod p(x_i|\theta)$  is the usual likelihood function and  $p(x|n) = \int p(x|\theta, n)p_{prior}(\theta)d\theta$  is a normalizing constant.

With the posterior distribution of  $\theta$  available, Riffa and Schlaifer's (1961) resolution of this problem is to proceed in reverse time order, taking expectations of utility over random variables  $\theta$  and  $x$ , and maximixing over choice variables  $d$  and  $n$ . Assuming that utility does not depend on  $x$  and is additive and linear in  $n$ , the utility function can be written as  $u(n, x, d, \theta) = u(d, \theta) - cn$ , where  $c$  is the cost in utiles of each additional observation (Riffa and Schlaifer 1961; Lindley 1997). Therefore, the objective function is (Lindley 1997):

$$(1) \quad \max_{n \geq 0} \left\{ \int_x \max_{d \geq 0} \left[ \int_{\theta} u(d, \theta) p_{post}(\theta|x, n) d\theta \right] p(x|n) dx - cn \right\}.$$

The problem can be solved by taking the expectation over  $\theta$  of the utility of  $d$ , given  $x$  and  $n$ , and then maximizing over the decision variable  $d$ . Subsequently, the expectation over  $x$  of this maximized value can be found using  $p(x|n)$ , and finally, this expectation can be maximized over  $n$  to answer the original question, "What size sample should I take?"



### *Application to Genetic Testing for Beef Cattle*

Although equation (1) offers a well-defined algorithm for solving the general form of the Bayesian decision theoretic approach to sample size determination, applications of this method are still limited. Most notably, several studies have used the fully Bayesian approach to determine the optimal sample size for clinical trials in medical research. Although these studies have been conducted from a variety of different perspectives, including societal/public health (Gittins and Pezeshk 2002; Wilan and Pinto 2005, 2006) and industry/pharmaceutical companies (Gittins and Pezeshk 2000, 2002; O'Hagan and Stevens 2001; Kikuchi, Pezeshk, and Gittins 2008; Wilan 2008), they focus on a single market participant. Conversely, there are two participants in the market for feeder cattle: buyers and sellers. Therefore, we extend the model to take into account both buyer and seller information.

Suppose  $\mathbf{X}_i$  ( $i = 1, 2, \dots, m$ ) is a column vector of molecular breeding values (MBVs) characterizing  $p$  economically relevant traits for the  $i$ th animal in a lot of  $m$  feeder cattle. Within a pen of cattle, this vector of genetic markers is assumed to be independent and identically distributed across animals with a multivariate normal density (Mrode 2014):<sup>1</sup>

---

<sup>1</sup> The independence assumption is a common simplifying assumption (Hoff 2009). However, in practice genetic markers are likely to be positively correlated across animals. For example, a lot of feeder cattle from the same ranch often have a high degree of relatedness with dams frequently being cousins and bulls being able to service 20-25 head. Holding all else constant, relaxing the independence assumption would likely result in a lower optimal sample size and higher returns from sampling as the information collected from each animal is more informative about the pen as a whole. Therefore, our model, which assumes independence, represents the

$$(2) \quad \mathbf{X}_i \sim iid \text{MVN}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}),$$

where the true value of  $\boldsymbol{\theta}$  is unknown and the  $p \times p$  variance-covariance matrix  $\boldsymbol{\Sigma}$  is known from previous experience. Both the buyer (feedlot) and the seller (producer) of a lot of feeder cattle have their own subjective prior distributions of the unknown parameter  $\boldsymbol{\theta}$ . For example, the seller's prior distribution of  $\boldsymbol{\theta}$  can be expressed as:

$$(3) \quad \boldsymbol{\theta} \sim \text{MVN}_p(\boldsymbol{\mu}_s, \mathbf{V}_s),$$

and the buyer's prior distribution can be written similarly as:

$$(4) \quad \boldsymbol{\theta} \sim \text{MVN}_p(\boldsymbol{\mu}_b, \mathbf{V}_b).$$

Given previous experience with their own cattle, the seller is expected to have a narrower distribution of  $\boldsymbol{\theta}$  than the buyer (i.e.,  $\mathbf{V}_s \leq \mathbf{V}_b$ ). Although buyers often have access to some information about how animals will perform, in the extreme case that the buyer is completely uninformed  $\mathbf{V}_b$  is characterized by the variance-covariance matrix of the MBVs between lots of cattle.

A model examining the consequences of asymmetric information between buyers and sellers was first introduced by Akerlof (1970) using the example of the market for used cars. This model has since been extended to a variety of topics, including asymmetric information in cattle auctions (Allen 1993; Chymis 2007). Results indicate that one way to alleviate the inefficiency created by this information gap is to introduce credible information to the decision problem. However, as indicated by Stigler (1961) a rational decision maker will only obtain additional information if the benefits outweigh the costs. Therefore, given the discrepancy between buyer and seller expectations of the genetic makeup of a lot of feeder cattle, sellers who

---

extreme case in which animals are completely unrelated, and as a result, the optimal sample sizes presented here are likely an upper bound on the true profit maximizing sample size.

know that their cattle have valued genetics may want to differentiate their cattle by convincing buyers that they are higher quality. One way to do this would be to use genetic testing to establish the actual genetic makeup of the lot. However, previous research has consistently found that testing each individual animal is not cost-effective. Therefore, the seller may choose to randomly sample a subset of the cattle. The question is how many cattle should the seller test to maximize returns?

An objective function characterizing the costs and benefits of randomly sampling  $n$  animals out of a lot of  $m$  feeder cattle is specified using a Bayesian prior distribution for the unknown vector of parameters  $\theta$ . The cost of testing is known to be  $\frac{c}{m}n$ , where  $c$  is the cost of commercial testing services (\$/head). The benefit of persuading the buyer that a pen of cattle has higher quality genetics than expected is characterized by the increase in the buyer's expected profit which is specified as a function of genetics,  $\pi(\theta)$ . For simplicity, the buyer is assumed to be risk neutral so that the seller receives the entire surplus profit created from testing. Moreover, in application the presence of a large number of buyers would result in the additional value to buyers being bid away given the reasonable assumption that the market for feeder cattle is perfectly competitive (Zhao, Du, and Hennessy, 2011).

As a result, the seller's objective function  $r(n)$ , which is the total expected benefit from the resulting improvement in the buyer's profit function minus the cost of testing, can be written as:

$$(5) \quad \max_{n \geq 0} r(n) = \int_X \left[ \int_{\theta} \pi(\theta) p_{post}(\theta|X, n) d\theta \right] p(X|n) dX \\ - \int_{\theta} \pi(\theta) p_{prior}(\theta) d\theta - \frac{c}{m}n,$$

where  $p_{post}(\boldsymbol{\theta}|\mathbf{X}, n)$  is the buyer's posterior distribution of  $\boldsymbol{\theta}$  conditional on the genetic information  $\mathbf{X}$  collected from a random sample of  $n$  animals and  $p_{prior}(\boldsymbol{\theta})$  is the buyer's prior distribution of  $\boldsymbol{\theta}$  given in equation (4). An application of Bayes rule indicates that the posterior distribution is proportional to the product of the likelihood function and the buyer's prior distribution of  $\boldsymbol{\theta}$ ,  $p_{post}(\boldsymbol{\theta}|\mathbf{X}, n) \propto p(\mathbf{X}|\boldsymbol{\theta})p_{prior}(\boldsymbol{\theta})$ . Therefore, in order to maximize the returns from sampling in equation (5), the seller is assumed to know the buyer's prior distribution of  $\boldsymbol{\theta}$ , which could be implied from the initial bid received.

If, for simplicity of exposition, we assume that the buyer's profit function in equation (5) is linear in the MBVs,  $\pi = \alpha + \boldsymbol{\beta}\mathbf{X}$ , the multivariate distribution of genetics in equation (2) can be transformed into a univariate normal distribution of profit per head within a pen of cattle:

$$(6) \quad \pi_i \sim iid N(\alpha + \boldsymbol{\beta}\boldsymbol{\theta}, \boldsymbol{\beta}\boldsymbol{\Sigma}\boldsymbol{\beta}'),$$

where the mean is still unknown. Similarly, buyer and seller subjective prior distributions of genetics can also be transformed into prior distributions of profit. For example, the seller's prior distribution of expected profit is a linear transformation of the prior distribution of genetics in equation (3):

$$(7) \quad \alpha + \boldsymbol{\beta}\boldsymbol{\theta} \sim N(\alpha + \boldsymbol{\beta}\boldsymbol{\mu}_s, \boldsymbol{\beta}\mathbf{V}_s\boldsymbol{\beta}'),$$

and the linear transformation of the prior distribution of genetics in equation (4) returns the buyer's prior distribution of expected profit:

$$(8) \quad \alpha + \boldsymbol{\beta}\boldsymbol{\theta} \sim N(\alpha + \boldsymbol{\beta}\boldsymbol{\mu}_b, \boldsymbol{\beta}\mathbf{V}_b\boldsymbol{\beta}').$$

Replacing distributions of genetics with distributions of profit, the objective function in equation (5) can be rewritten as:

$$(9) \quad \max_{n \geq 0} r(n) = \int_{\bar{\pi}} \left[ \int_{\alpha + \beta \theta} (\alpha + \beta \theta) p_{post}(\alpha + \beta \theta | \bar{\pi}, n) d(\alpha + \beta \theta) \right] p(\bar{\pi} | n) d\bar{\pi} \\ - \int_{\alpha + \beta \theta} (\alpha + \beta \theta) p_{prior}(\alpha + \beta \theta) d(\alpha + \beta \theta) - \frac{c}{m} n,$$

where  $p_{post}(\alpha + \beta \theta | \bar{\pi}, n)$  is the buyer's posterior distribution of profit conditional on the sufficient statistic  $\bar{\pi} = \frac{1}{n} \sum_{i=1}^n \pi_i$  and sample size  $n$  and  $p_{prior}(\alpha + \beta \theta)$  is the buyer's prior distribution of expected profit in equation (8). Given that both the likelihood function and the buyer's prior distribution of expected profit are normally distributed, the posterior distribution is also well known to be normally distributed as (Hoff 2009):

$$(10) \quad \alpha + \beta \theta | \bar{\pi}, n \sim N \left( \frac{(\beta \Sigma \beta')(\alpha + \beta \mu_b) + n(\beta V_b \beta') \bar{\pi}}{(\beta \Sigma \beta') + n(\beta V_b \beta')}, \frac{(\beta \Sigma \beta')(\beta V_b \beta')}{(\beta \Sigma \beta') + n(\beta V_b \beta')} \right).$$

From this familiar solution we can see that when  $n = 0$  the mean and variance of the posterior distribution reduce to the buyer's prior distribution of profit in equation (8). However, at values of  $n > 0$  the mean and variance of the posterior distribution converge towards the sample mean and sample variance.

Because of the linearity assumption of the profit function, the integrals in equation (9) can be evaluated by replacing the random parameters with their expected values. For example, the integrals with respect to  $\alpha + \beta \theta$  can be evaluated by replacing the unknown values of profit with the expected values of the buyer's posterior and prior distributions of profit, respectively:

$$(11) \quad \max_{n \geq 0} r(n) = \int_{\bar{\pi}} \left[ \frac{(\beta \Sigma \beta')(\alpha + \beta \mu_b) + n(\beta V_b \beta') \bar{\pi}}{(\beta \Sigma \beta') + n(\beta V_b \beta')} \right] p(\bar{\pi} | n) d\bar{\pi} \\ - (\alpha + \beta \mu_b) - \frac{c}{m} n.$$

Given that the objective function is from the seller's perspective, the density of the sample mean is known to be  $\bar{\pi}|n \sim N\left(\alpha + \beta\mu_s, \frac{1}{n}(\beta V_s \beta')\right)$ . Therefore, the integral with respect to  $\bar{\pi}$  can also be evaluated by replacing  $\bar{\pi}$  with its expected value,  $\alpha + \beta\mu_s$ . As a result, equation (11) can be rewritten as:

$$(12) \quad \max_{n \geq 0} r(n) = \frac{(\beta \Sigma \beta')(\alpha + \beta\mu_b) + n(\beta V_b \beta')(\alpha + \beta\mu_s)}{(\beta \Sigma \beta') + n(\beta V_b \beta')} - (\alpha + \beta\mu_b) - \frac{c}{m}n.$$

In practice, sample size  $n$  is an integer. However, if we treat  $n$  as continuous for ease of exposition, we can maximize  $r(n)$  by letting  $\frac{dr(n)}{dn} = 0$ . Taking the derivative of equation (12) and solving for  $n$ , the optimal sample size,  $n^*$ , is:

$$(13) \quad n^* = \frac{\sqrt{(\beta \Sigma \beta')(\beta V_b \beta')(\beta\mu_s - \beta\mu_b)} - (\beta \Sigma \beta')\sqrt{\frac{c}{m}}}{(\beta V_b \beta')\sqrt{\frac{c}{m}}}.$$

The second order condition indicates that the solution in equation (13) maximizes the seller's returns from sampling for values of  $\beta\mu_s > \beta\mu_b$ :

$$(14) \quad \frac{d^2 r(n)}{dn^2} = \frac{-2(\beta \Sigma \beta')(\beta V_b \beta')^2(\beta\mu_s - \beta\mu_b)}{[(\beta \Sigma \beta') + n(\beta V_b \beta')]^3} < 0 \quad \forall \beta\mu_s > \beta\mu_b.$$

That is, as long as the seller's prior expectation of profit is higher than the buyer's prior expectation of profit,  $n^* = \text{argmax}\{r(n)\}$ . This makes sense given that a seller who knows that the quality of their cattle is lower than the buyer's prior expectation has no incentive to test.

#### *Alternative Applications for Genetic Testing of Beef Cattle*

The framework described above is just one of many applications of this method to genetic testing for beef cattle. For example, similar to the scenario described above, a producer could use a sample of genetic information to determine whether or not to retain ownership of a lot of feeder cattle. Another potential alternative is that a feedlot may want to differentially manage cattle

based on genetics, or marker-assisted management, but cannot collect genetic information until after they have purchased a lot of feeder cattle. In this case, the benefit portion of the objective function will be the improvement in feedlot profit from improved management decisions, including how cattle are fed, how technologies such as implants and beta agonists are used, and how cattle are marketed (Van Eenennaam and Drake 2012), and the cost portion will remain the same. Therefore, the feeder now chooses some feedlot management decision variable ( $d$ ; e.g., marketing method and/or days-on-feed) and the number of animals to be sampled ( $n$ ), and the objective function in equation (5) can be re-specified as:

$$(15) \quad \max_{n \geq 0} \left\{ \int_{\mathbf{X}} \max_{d \geq 0} \left[ \int_{\boldsymbol{\theta}} \pi(d, \boldsymbol{\theta}) p_{post}(\boldsymbol{\theta} | \mathbf{X}, n) d\boldsymbol{\theta} \right] p(\mathbf{X} | n) d\mathbf{X} \right. \\ \left. - \int_{\boldsymbol{\theta}} \pi(d, \boldsymbol{\theta}) p_{prior}(\boldsymbol{\theta}) d\boldsymbol{\theta} - \frac{c}{m} n \right\},$$

where profit is now a function of the feedlot management decision variable  $d$  and genetics  $\boldsymbol{\theta}$ ,  $p_{post}(\boldsymbol{\theta} | \mathbf{X}, n)$  is the feeder's posterior distribution of  $\boldsymbol{\theta}$  conditional on the genetic information  $\mathbf{X}$  collected from a random sample of  $n$  animals, and  $p_{prior}(\boldsymbol{\theta})$  is the feeder's prior distribution of  $\boldsymbol{\theta}$ . While it is important to acknowledge alternative applications of this model, for simplicity in this paper we focus on the original context discussed above of a seller trying to convince a buyer that their cattle are higher quality in order to receive a premium.

## Data

Data were provided by Neogen, the parent company of commercial testing service Igenity, for 2,976 commercially-fed cattle from a single feedlot in Iowa. Cattle represented year-round placements in the years 2007 and 2008. At placement, animals were weighed and a hair sample or tissue punch from ear tag application was collected for genetic testing. Genetic information was provided in the form of MBVs for the following seven traits: marbling, yield grade, rib-eye

area (in<sup>2</sup>), hot-carcass weight (lbs.), average daily gain (lbs./day), tenderness (lbs. of Warner-Bratzler shear force [WBSF]), and days-on-feed (days). Each of these markers, except hot-carcass weight and days-on-feed, have been found to be significantly correlated with the traits they are designed to predict in independent validations (DeVuyst et al. 2011; National Beef Cattle Evaluation Consortium 2015). Molecular breeding values are a continuous representation of an animal's genetic potential to express a given trait. Similar to expected progeny differences (EPDs), MBVs are reported in the units of the trait they represent. However, they are interpreted as the “relative differences expected in animals across breeds compared to their contemporaries” (Igenity 2013, p. 2). For example, if two animals exposed to the same environmental and management conditions have marbling MBVs of -100 and 100, respectively, we would expect, on average, that these two animals' marbling scores would differ by 200 units ( $100 - [-100] = 200$ ). Additional live-animal characteristics for days-on-feed, sex, and hide color were also provided, and carcass performance measurements, including calculated yield grade, marbling score, and hot-carcass weight, were collected at slaughter. Summary statistics for carcass performance, live-animal characteristics, and MBVs are reported in table 1.

The data consist of 101 contemporary groups, which are defined as groups of animals that had an equal opportunity to perform: same sex, managed alike, and exposed to the same feed resources. These groups ranged in size from 11 to 69 animals with an average group size of 29 animals. While most of the groups are expected to be from a single producer, the extent to which some cattle were comingled from different herds was not recorded. Therefore, the results of our analysis are conditional on the data that we use, and may underestimate the value of testing for cattle that are known to have homogeneous genetics. Sensitivity analysis is done to provide some context for the results presented here.



## Procedures

Using the data described above we estimate the parameters needed to obtain the optimal sample size in equation (13) and the returns from sampling in equation (12). A brief description of each of the parameters along with their baseline values and a range of parameter values for sensitivity analysis are reported in table 2. The procedures for how these parameters and their ranges were estimated from the data are discussed below.

### *Expected Profit*

The intercept and slope coefficients in the buyer's profit equation,  $\alpha$  and  $\beta$ , are estimated using a mixed model regression of feedlot profit on the MBVs. Prior to estimating these parameters an estimate of profit for each animal in the sample was generated using grid pricing:

$$(16) \quad \pi = P(YG, QG, HCW) \times HCW - PC(PWT, SEX) - FC(DOF) - YC(DOF) \\ - IC(PC, DOF),$$

where  $P$  is the grid price which is a function of actual yield grade ( $YG$ ), quality grade ( $QG$ ), and hot-carcass weight ( $HCW$ ),  $PC$  is the purchase cost of feeder cattle which is a function of placement weight ( $PWT$ ) and sex ( $SEX$ ),  $FC$  is feed cost,  $YC$  is yardage cost, and  $IC$  is interest cost on the purchase of feeder cattle which are all a function of days-on-feed ( $DOF$ ). Fed cattle prices, including grid premiums and discounts, and feeder cattle prices were simple averages of the weekly prices reported by the USDA Agricultural Marketing Service (AMS) for the 2014 marketing year and were obtained from the Livestock Marketing Information Center (LMIC) spreadsheets (LMIC 2015; USDA AMS 2015). The prices used in this analysis are reported in table 3. Observations of feed intake were not available. Therefore, a standardized estimate of feed intake was generated for each animal in the sample using the dry matter intake model from the National Research Council's (NRC) *Nutrient Requirements of Beef Cattle* (NRC 2000). For

examples of the National Research Council's dry matter intake model see Lusk (2007) or Thompson et al. (2014). Additional information needed to evaluate profit includes a dry matter feed cost of \$230/ton (\$0.12/lb.), yardage cost of \$0.40/day, and a 7% interest rate on the purchase of feeder cattle (Lardy 2013).

Using equation (16), estimates of profit for the animals in our sample are generally negative. This is contradictory to Tonsor (2015), which reported that feedlot net returns in Kansas were positive for most of 2014. However, negative returns to cattle feeding are common with average net returns over the past 13 years being -\$26.77 and -\$15.44/head for steers and heifers, respectively (Tonsor 2015). The discrepancy between our estimates of profit and those reported by Tonsor (2015) are likely due to differences in the input and output prices used in this analysis and those observed when cattle in our sample were actually being fed.

A mixed model regression equation of profit on MBVs and other live-animal characteristics is then estimated as:

$$(17) \quad \pi_{ij} = \alpha_0 + \alpha_1 PWT_{ij} + \alpha_2 DOF_{ij} + \alpha_3 DOF_{ij}^2 + \alpha_4 PWT_{ij} DOF_{ij} + \alpha_5 STR_{ij} \\ + \alpha_6 BLK_{ij} + \boldsymbol{\beta} \mathbf{X}_{ij} + v_j + \varepsilon_{ij},$$

where  $\pi_{ij}$  is the estimated feedlot profit for the  $i$ th animal in the  $j$ th contemporary group,  $PWT_{ij}$  is placement weight,  $DOF_{ij}$  is days-on-feed,  $STR_{ij}$  is a dummy variable equal to one if the animal was a steer and zero otherwise,  $BLK_{ij}$  is a dummy variable equal to one if the animal was black-hided and zero otherwise,  $\mathbf{X}_{ij}$  is a  $7 \times 1$  vector of MBVs characterizing marbling, yield grade, rib-eye area, hot-carcass weight, average daily gain, tenderness, and days-on-feed,  $v_j \sim N(0, \sigma_v^2)$  is a contemporary group random effect, and  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  is a random error term.

The model is estimated using Proc Mixed in SAS (SAS Institute Inc. 2013). D'Agostino-Pearson  $K^2$  omnibus test for skewness and kurtosis and a conditional variance test identified

evidence of nonnormality and static heteroskedasticity. Sandwich estimators of the standard errors were estimated to obtain estimates of standard errors that were consistent in the presence of nonnormality and static heteroskedasticity (White 1982). Given the large sample size, asymptotic properties are relevant, and the small sample biases common with generalized method of moments estimators should be of little concern.

Results from the estimation of equation (17) are reported in table 4. Live-animal characteristics generally exhibited the expected relationships with the only notable result being that, on average, fed cattle profit for steers was \$75/head less than heifers. While previous research has discussed the potential for heifers to generate higher returns (Williams et al. 1993; Tonsor 2015), steers are usually expected to generate higher fed cattle profit. In this case, the discrepancy between feedlot profit for steers and heifers is a result of the prices used. Although feeder cattle prices include a premium for steers regardless of the weight class, the average dressed fed cattle prices used as the base for the grid were nearly identical for steers and heifers. Fixed effects for live-animal characteristics are not of interest in this study, so these variables are set to their mean values and absorbed into the intercept,  $\alpha = \alpha_0 + \alpha_1 \overline{PWT} + \alpha_2 \overline{DOF} + \alpha_3 \overline{DOF}^2 + \alpha_4 \overline{PWT} \times \overline{DOF} + \alpha_5 \overline{STR} + \alpha_6 \overline{BLK}$ . Thus, the profit equation can be written as a linear function of the MBVs,  $\pi = \alpha + \beta X$ .

Each of the MBV effects, except for the rib-eye area MBV, is positive. This is consistent with expectations given that higher MBVs correspond with more favorable outcomes for the traits they characterize. The marbling and hot-carcass weight MBVs were the only markers to significantly influence fed cattle profitability. The negative effect of the rib-eye area MBV is likely due to genetic correlations among the MBVs included in our model. Specifically, the known inverse relationship between rib-eye area and marbling is likely the primary driver of this

result (DeVuyst et al. 2011). That is, more favorable rib-eye area outcomes are often accompanied by less favorable outcomes for marbling. Therefore, for the grid used in this analysis, the premiums associated with more favorable rib-eye area (yield grade) outcomes are not enough to offset the lower premiums, or higher discounts, associated with less favorable marbling (quality grade) outcomes.

Buyer and seller expectations of profit are estimated from the data using equation (17). We assume that the buyer's expected value of profit,  $\alpha + \beta\mu_b$ , is equal to the mean profit observed in our data. Given linearity of the profit function, this is equivalent to profit at the mean values of the seven MBVs. If the seller's expectation of profit is less than or equal to the buyer's prior expectation of profit then the optimal sample size is zero. That is, there is no incentive to test if sellers know that their cattle will have below-average profitability. However, for any combination of MBVs that  $\alpha + \beta\mu_s > \alpha + \beta\mu_b$  there is a potential benefit to testing. Therefore, the baseline value for the seller's expectation of profit is arbitrarily assumed to be equal to the 75<sup>th</sup> percentile of profit observed in our sample, and sensitivity analysis is conducted for values of  $\alpha + \beta\mu_s$  ranging from the mean to the maximum profit observed in the sample to determine the effect of the quality of a particular set of cattle on optimal sample size and the returns from sampling.

#### *Variance of Profit Between and Within Pens*

The variances of the MBVs *between* and *within* lots of cattle (i.e., the diagonal elements of  $\mathbf{V}_b$  and  $\mathbf{\Sigma}$ , respectively) can be estimated using a random effects model for each of the seven MBVs:

$$(18) \quad MBV_{ijk} = \mu_k + g_{jk} + e_{ijk},$$

where the dependent variable is the  $k$ th MBV for the  $i$ th animal in the  $j$ th contemporary group,  $\mu_k$  is the mean for the  $k$ th MBV,  $g_{jk} \sim N(0, \tau_k^2)$  is a contemporary group random effect where  $\tau_k^2$

is the variance of the  $k$ th MBV *between* groups of cattle, and  $e_{ijk} \sim N(0, \sigma_k^2)$  is a random error term where  $\sigma_k^2$  is the variance of the  $k$ th MBV *within* groups. Models for each of the seven MBVs were estimated independently using Proc Mixed in SAS (SAS Institute Inc. 2013).

Estimates of the between and within variance from the random effects models are reported in table 5. The variance of the MBVs within contemporary groups was generally higher than the variance between groups. As discussed earlier, this result could indicate that some of the cattle in our sample were comingled for feeding. That is, cattle from different sources with differing genetics were fed in the same contemporary group resulting in higher within pen variability than may be experienced by a single producer with a lot of cattle with homogenous genetics. Nonetheless, the estimated variances, along with the correlation matrix of the seven MBVs, can be used to calculate the covariances, or off-diagonal elements, of  $\mathbf{V}_b$  and  $\mathbf{\Sigma}$  using the known relationship between them,  $Cov(x, y) = \rho_{x,y}\sigma_x\sigma_y$ . Variance-covariance matrices of genetics are then converted to scalar estimates of the variance of profit within and between lots of cattle using the vector of parameter values  $\boldsymbol{\beta}$ .

In addition to their baseline values, the variance of profit between  $(\boldsymbol{\beta}\mathbf{V}_b\boldsymbol{\beta}')$  and within  $(\boldsymbol{\beta}\mathbf{\Sigma}\boldsymbol{\beta}')$  lots of cattle is subjected to sensitivity analysis to determine how the homogeneity or heterogeneity of a particular lot of cattle influences optimal sample size and the returns from sampling. The upper and lower bounds for the sensitivity analysis of both  $\boldsymbol{\beta}\mathbf{V}_b\boldsymbol{\beta}'$  and  $\boldsymbol{\beta}\mathbf{\Sigma}\boldsymbol{\beta}'$  were approximately set to equal the maximum and minimum variance of profit observed for the contemporary groups in our data.

### *Cost of Genetic Testing*

The final component to evaluating the optimal sample size in equation (13) is the per-head cost of genetic testing,  $\frac{c}{m}$ . The current cost of commercially available genetic testing services is  $c =$

\$40/head (Igenity 2015). Therefore, assuming the data collected is applied to a pen of  $m = 100$  animals, the cost of testing an additional animal is \$0.40/head. In addition to this baseline value, sensitivity analysis is conducted for costs of testing ranging from \$0.10/head to \$1.00/head.

These values are evaluated to account for differing costs of testing due to smaller or larger lots of cattle and to account for the potential decreasing cost of genetic testing services.

## **Results**

The optimal sample size and returns from sampling are first evaluated at the baseline parameter values in table 2. A plot of the expected value of sample information (EVSI), total cost, and expected net gain (ENG), as a function of sample size ( $n$ ), are reported in figure 1. The EVSI is increasing sharply for small sample sizes as the genetic information collected from each additional animal contains valuable information. However, as sample size increases above  $n = 20$  the EVSI levels off at values of \$16-\$18/head. Given that the cost of testing each additional animal is the same, \$0.40/head, the total cost function is linear with respect to sample size. The difference in the EVSI and total cost is the ENG, or the returns from sampling. Results indicate that in this case an optimal sample size of  $n^* = 10$  maximized the ENG. Despite only testing 10% of the animals, the returns from sampling are \$9.79/head compared with not testing any animals and accepting the buyer's prior expectation of the genetic makeup of the pen.

Aggregating across the assumed pen size of  $m = 100$  animals, that is a net return of approximately \$1,000 from a \$400 investment, or a 250% return-on-investment. These results indicate that random sampling has the potential to provide a context in which the benefits of genetic testing outweigh the costs, which has not generally been the case in previous research where each individual animal was tested and the results were used to sort cattle into management groups.

Sensitivity analysis is conducted to evaluate the robustness of these results with respect to the parameters  $\alpha + \beta\mu_s$ ,  $\beta\Sigma\beta'$ ,  $\beta V_b\beta'$ , and  $\frac{c}{m}$ . As expected, holding all else constant, optimal sample size and the returns from sampling are both increasing with respect to the seller's expectation of profit ( $\alpha + \beta\mu_s$ ; figure 2, panel i). That is, sellers with higher quality (more profitable) cattle have an incentive to test more animals. However, consider that the optimal sample size for a lot of cattle with a genetic makeup similar to the most profitable animals in our sample is only about  $n^* = 24$ . Although the returns from testing for such animals are quite high, nearly \$60/head, the marginal return to testing additional animals beyond  $n = 24$  is not enough to offset the marginal cost of obtaining additional information. Therefore, at the baseline values for the variance of profit between and within lots of cattle and the cost of testing, the benefits of genetic testing can be captured by testing a relatively small portion of a lot of feeder cattle (less than 25 out of 100 animals for a set of very high-quality cattle).

Optimal sample size is also increasing with the variance of profit *within* a lot of cattle ( $\beta\Sigma\beta'$ ; figure 2, panel ii). However, the returns from sampling are simultaneously decreasing. That is, the additional noise associated with more heterogeneous lots of cattle make it more difficult to identify improvements in the actual quality of a particular set of cattle. As a result, more animals must be tested in order to “convince” buyers that a particular set of cattle is actually higher quality. In addition, the marginal return to testing each additional animal is decreased relative to more homogeneous lots of cattle resulting in decreased returns from sampling. For example, a set of cattle with variability similar to the most heterogeneous lot of cattle observed in our sample has an optimal sample size of  $n^* = 12$ , but the returns from sampling are only about \$4/head. On the other hand, producers selling feeder cattle that are known to have very homogeneous genetics may be able to capture a large portion of the

additional value by testing a relatively small percentage of the cattle. For example, the returns from sampling for a set of cattle with variability similar to the most homogeneous group of cattle in our sample is \$15/head with an optimal sample size of just  $n^* = 5$ .

In the case that the buyer is completely uninformed, the variance of profit *between* lots of cattle is used to characterize the variance of the buyer's subjective prior distribution of expected profit for a lot of feeder cattle. Contrary to within variability described above, optimal sample size is decreasing with the variance of the buyer's prior distribution of expected profit ( $\beta V_b \beta'$ ), and the returns from sampling are increasing (figure 2, panel iii). This indicates that as the buyer's prior distribution of expected profit becomes more uninformed (i.e., more diffuse), it is easier to persuade their opinion away from their prior expectation towards the seller's expectation of the actual profitability of the pen. As a result, fewer animals need to be tested to “convince” the buyer that the cattle are actually higher quality, and the marginal returns from testing each additional animal becomes more valuable. For example, if the buyer is completely uninformed (high variance for the prior distribution of profit) the optimal sample size is  $n^* = 4$ , and the returns from sampling are over \$16/head. However, for buyers with very narrow prior distributions of expected profit the optimal sample size is  $n^* = 11$  and the returns from sampling are just \$8/head.

Lastly, as would be expected, the per-head cost of testing  $\left(\frac{c}{m}\right)$  is inversely related to optimal sample size and the returns from sampling (figure 2, panel iv). That is, as the marginal cost of testing increases, the ENG of testing each additional animal is reduced, and as a result, optimal sample size will decrease. The cost of testing is made up of two components: the cost of genetic testing services ( $c$ ) and the size of the pen to which the information collected is being applied ( $m$ ). Therefore, a lower per-head cost of testing could be the result of either a reduction



in the cost of the test due to technological advancements or an increase in the number of animals to which the decision is applied. Either way, results indicate that the ability to reduce the cost of testing would enable sellers to test more animals allowing them to achieve closer to the full value of their cattle. For example, the ability to cut the baseline value of per-head testing cost in half, \$0.20/head, increases the optimal sample size to  $n^* = 15$  and the returns from testing to \$12/head. Conversely, doubling the per-head cost of testing, \$0.80/head, leads to an optimal sample size of  $n^* = 6$  animals and returns from sampling of less than \$7/head.

## **Conclusions**

In this article, we introduce an economic approach to sample size determination utilizing a Bayesian decision theoretic framework (Grundy, Healy, and Rees, 1956; Riffa and Schlaifer 1961; Lindley 1997). To date, few economic studies explicitly consider the endogeneity of sample size and so this method is scarcely used in economic research. However, this method is a theoretically sound approach to determining sample size for many economic problems and should be considered in a wide range of economic modeling problems. For example, consider a generic scenario in which a new policy is to be implemented for the betterment of society. Prior to its implementation, the policy maker must determine the optimal level of the policy through an assessment of its effect on the population. In most cases, it would be too costly to take a census. However, there is significant value associated with obtaining a sufficient sample to determine the level of the policy that will maximize the overall benefits. Therefore, sample size is an endogenous variable that needs to be considered jointly with other decision variables by utilizing an objective function to balance the expected costs and benefits of sampling.

To demonstrate the method for a relevant applied economics problem, we turn to the problem of asymmetric information in the market for feeder cattle. A theoretical model is

developed characterizing a scenario in which sellers (producers) who know that their feeder cattle have high-value genetics are trying to convince buyers (feedlots) that their cattle are higher quality in order to receive a premium above the prevailing market price. However, in order to establish the actual genetic makeup of a lot of feeder cattle, the seller must incur the cost of genetic testing. Using the Bayesian sampling model we determine the economically-optimal sample size and the returns from sampling.

Results from this example indicate that the marginal benefit to testing is high for small sample sizes as the genetic information collected from each additional animal contains valuable information. However, as the sample size increases the marginal expected value of sample information diminishes quickly, indicating that a large portion of the additional value for higher-quality cattle can be estimated by testing a relatively small percentage of a lot of feeder cattle. For example, at the baseline parameter values the optimal sample size is  $n^* = 10$  animals out of a lot of 100 feeder cattle and the returns from sampling are nearly \$10/head. Aggregating across the assumed pen size of 100 animals, that is a net return of approximately \$1,000 from a \$400 investment, or a 250% return-on-investment. Sensitivity analysis was conducted to provide some context, and results indicated that the optimal sample size and returns from sampling may increase or decrease depending on the actual quality (or profitability) of a particular pen of cattle, the homogeneity within the pen, the variance of the buyer's subjective prior distribution of expected profit, and the per-head cost of genetic testing. The only scenario evaluated in which the returns from sampling were not positive was if the seller's expectation of profit was less than or equal to the buyer's prior expectation of profit, in which case the optimal sample size is  $n^* = 0$ .

Nonetheless, results suggest that random sampling has the potential to provide a context in which the benefits of genetic testing outweigh the costs, which has not generally been the case in previous research. Previous research demonstrates that genetic differences generate measurable differences in fed cattle profitability. However, the per-head difference in profit has been shown to be less than genetic testing costs. So prior to this research, genetic testing of feeder cattle for feedlot placement decisions has not been considered economically advisable. Our sampling method demonstrates that economically-optimal sampling of feeder cattle can improve profitability. This is the first research to demonstrate the potential for genetic testing of feeder cattle to improve net return from testing. So, this research represents an important contribution to the literature evaluating the economic value of genetic testing for beef cattle and potentially more importantly to the beef cattle industry. While demonstrated for a feeder calf sale, the method can be similarly applied for a fed calf sale to packers or boxed beef sale to high-end retailers.

When interpreting the results presented here consider that we only include the direct cost of genetic testing. However, there may be additional management and handling costs associated with collecting samples for genetic testing. As indicated in previous research, these additional costs could be mitigated, or even eliminated, if samples were collected at a point when animals are already being handled (Koontz et al. 2008). Perhaps the more troublesome issue is that if the seller is the one making the sampling decision, will they truly select a “random sample?” That is, there is an incentive for the seller to selectively sample animals known to be higher quality. This raises the question of whether third-party verification would be needed to ensure the cattle being tested are really a random sample. This is similar to the third-party verification associated with value-added feeder cattle sales (Chymis et al. 2007; Williams et al. 2012), and brings about the

potential for some additional costs that may need to be considered in order for the application described here to succeed. However, using third-party verified value-added feeder calf sales as the metric, this cost is likely much less than estimated returns from genetic sampling.

It is also important to note that the static context described here does not lead to changes in the product form, and as a result, does not appear to generate any additional value. Instead, value is redistributed between buyers and sellers. However, by better aligning prices received by feeder calf producers with traits desired by feedlot operators, economically-optimal genetic testing has the potential to reduce misaligned incentives and improve the overall profitability of the beef industry. By reducing asymmetric information, cow-calf producers with genetics valued by feedlot operators are compensated for those genetics. This provides the incentive to produce more cattle that perform better in feedlots with higher value carcasses, so profits for the beef sector improve. While the model presented here assumes that profits accrue to the cow-calf producer for tractability, the likely outcome is a division of profits from sampling to both buyer and seller. In addition, if feedlot operators were subsequently able to use the genetic information provided by sellers to improve feedlot management decisions, including how cattle are fed, how technologies such as implants and beta agonists are used, and how cattle are marketed (Van Eenennaam and Drake 2012), they could generate additional value by changing the product form (Koontz et al. 2008). This additional information would lead to reduced inefficiencies associated with cattle feeding and could potentially create social welfare gains through improved quality and consistency of beef products.

## References

- Adcock, C.J. 1997. Sample Size Determination: A Review. *Journal of the Royal Statistical Society: Series D (The Statistician)* 46(2):261-283.
- Akanno, E.C., G.S. Plastow, B.W. Woodward, S. Bauck, H. Okut, X.L. Wu, C. Sun, J.L. Aalhus, S.S. Moore, S.P. Miller, Z. Wang, and J.A. Basarab. 2014. Reliability of Molecular Breeding Values for Warner-Bratzler Shear Force and Carcass Traits of Beef Cattle – An Independent Validation Study. *Journal of Animal Science* 92(7):2896-2904.
- Akerlof, G.A. 1970. The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84(3):488-500.
- Allen, D.W. 1993. Pot-Bellies, Cattle Breeds and Revealing Signals. *Economic Inquiry* 31(3):481-487.
- Chymis, A.G. H.S James, Jr., S. Konduru, V.L. Pierce, and R.L. Larson. 2007. Asymmetric Information in Cattle Auctions: The Problem of Revaccinations. *Agricultural Economics* 36(1):79-88.
- DeVuyst, E.A., J.R. Bullinger, M.L. Bauer, P.T. Berg, and D.M. Larson. 2007. An Economic Analysis of Genetic Information: Leptin Genotyping in Fed Cattle. *Journal of Agricultural and Resource Economics* 32(2):291-305.
- DeVuyst, E.A., J.T. Biermacher, J.L. Lusk, R.G. Mateescu, J.B. Blanton Jr., J.S. Swigert, B.J. Cook, and R.R. Reuter. 2011. Relationships between Fed Cattle Traits and Igenity Panel Scores. *Journal of Animal Science* 89(5):1260-1269.
- Gittins, J.C., and H. Pezeshk. 2000. How Large Should a Clinical Trial Be? *Journal of the Royal Statistical Society: Series D (The Statistician)* 49(2):177-187.

- Gittins, J.C., and H. Pezeshk. 2002. A Decision Theoretic Approach to Sample Size Determination in Clinical Trials. *Journal of Biopharmaceutical Statistics* 12(4):535-551.
- Grundy, P.M., M.J.R. Healy, and D.H. Rees. 1956. Economic Choice of the Amount of Experimentation. *Journal of the Royal Statistical Society: Series B (Methodological)* 18(1):32-55.
- Hoff, P.D. 2009. *A First Course in Bayesian Statistical Methods*. Dordrecht: Springer.
- Igenity. 2013. Igenity Results Key for Beef. Available online at <http://www.igenity.com/pdfs/forms/AG048IgenityResultsKeyForBeef.pdf>.
- Igenity. 2015. Order Form for Beef. Available online at [http://www.neogen.com/Genomics/pdf/SubmissionForms/AG088\\_IgenityOrderFormBeef.pdf](http://www.neogen.com/Genomics/pdf/SubmissionForms/AG088_IgenityOrderFormBeef.pdf).
- Kikuchi, T., H. Pezeshk, and J.C. Gittins. 2008. A Bayesian Cost-Benefit Approach to the Determination of Sample Size in Clinical Trials. *Statistics in Medicine* 27(1):68-82.
- Koontz, S.R., D.L. Hoag, J.R. Brethour, and J. Walker. 2008. Production Inefficiency in Fed Cattle Marketing and the Value of Sorting Pens into Alternative Marketing Groups Using Ultrasound Technology. *Journal of Agricultural and Resource Economics* 40(3):895-912.
- Lardy, G. 2013. A Cow-Calf Producer's Guide to Custom Feeding. Extension Publication AS-1162, North Dakota State University Extension Service, Fargo, ND. Available online at <http://www.ag.ndsu.edu/pubs/ansci/beef/as1162.pdf>.
- Lambert, D.K. 2008. The Expected Utility of Genetic Information in Beef Cattle Production. *Agricultural Systems* 99(1):44-52.
- Laws, D.J., and A. O'Hagan. 2002. A Hierarchical Bayes Model for Multilocation Auditing. *Journal of the Royal Statistical Society: Series D (The Statistician)* 51(4):431-450.
- Lindley, D.V. 1997. The Choice of Sample Size. *Journal of the Royal Statistical Society: Series D (The Statistician)* 46(2):129-138.

- Livestock Marketing Information Center. 2015. Spreadsheets: Fed Cattle Prices. Available online at <http://www.lmic.info/>.
- Lusk, J.L. 2007. Economic Value of Selecting and Marketing Cattle by Leptin Genotype. *Journal of Agricultural and Resource Economics* 32(2):306-329.
- Mrode, R.A. 2014. *Linear Models for the Prediction of Animal Breeding Values*. 3<sup>rd</sup> ed. Wallingford: CABI.
- National Beef Cattle Evaluation Consortium. 2015. Commercial Genetic Test Validation. Available online at <http://www.nbcec.org/validation/igenity.html>.
- National Research Council (NRC). 2000. *Nutrient Requirements of Beef Cattle*, 7<sup>th</sup> ed. Washington D.C.: National Academy Press.
- O'Hagan, A., and J.W. Stevens. 2001. Bayesian Assessment of Sample Size for Clinical Trials of Cost-Effectiveness. *Medical Decision Making* 21(3):219-230.
- Patterson, E.R. 1993. Strategic Sample Size Choice in Auditing. *Journal of Accounting Research* 31(2):272-293.
- Riffa, H., and R. Schlaifer. 1961. *Applied Statistical Decision Theory*. Boston: Harvard University Graduate School of Business Administration.
- SAS Institute Inc. *SAS Systems for Windows v.6.1*. Cary, NC: SAS Institute Inc., 2013.
- Smith, T.M.F. 1976. *Statistical Sampling for Accountants*. London: Accountancy Age Books.
- Smith, T.F.M. 1979. Statistical Sampling in Auditing: A Statistician's Viewpoint. *Journal of the Royal Statistical Society: Series D (The Statistician)* 28(4):267-280.
- Stigler, G.J. 1961. The Economics of Information. *The Journal of Political Economy* 69(3):213-225.

- Thompson, N.M., E.A. DeVuyst, B.W. Brorsen, and J.L. Lusk. 2014. Value of Genetic Information for Management and Selection of Feedlot Cattle. *Journal of Agricultural and Resource Economics* 39(1):139-155.
- Thompson, N.M., E.A. DeVuyst, B.W. Brorsen, and J.L. Lusk. 2016. Using Genetic Testing to Improve Fed Cattle Marketing Decisions. *Journal of Agricultural and Resource Economics*, in press.
- Tonsor, G.T. 2015. Historical and Projected Kansas Feedlot Net Returns: September 2015. Publication AM-GTT-KFR-9.2015, Kansas State University, Department of Agricultural Economics, Manhattan, KS. Available online at [http://www.agmanager.info/livestock/marketing/outlook/newsletters/FinishingReturns/FeedlotReturns\\_Sep\\_15.pdf](http://www.agmanager.info/livestock/marketing/outlook/newsletters/FinishingReturns/FeedlotReturns_Sep_15.pdf).
- U.S. Department of Agriculture, Agricultural Marketing Service (USDA AMS). 2015. Livestock Cattle Reports. Available online at <http://www.ams.usda.gov/AMSV1.0/ams.fetchTemplateData.do?template=TemplateP&page=Cattle>.
- Van Eenennaam, A.L., and D.J. Drake. 2012. Where in the Beef-Cattle Supply Chain Might DNA Tests Generate Value? *Animal Production Science* 52(2-3):185-196.
- White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50(1):1-25.
- Wilan, A.R. 2008. Optimal Sample Size Determination from an Industry Perspective Based on the Expected Value of Information. *Clinical Trials* 5(6):587-594.
- Wilan, A.R., and E.M. Pinto. 2005. The Value of Information and Optimal Clinical Trial Design. *Statistics in Medicine* 24(12):1971-1806.
- Wilan, A.R., and E.M. Pinto. 2006. Correction: The Value of Information and Optimal Clinical Trial Design. *Statistics in Medicine* 25(4):720.



- Williams, C.L., M.R. Langemeier, J. Mintert, and T.C. Schroeder. 1993. Profitability Differences between Steers and Heifers. Extension Publication MF-1075, Kansas State University Cooperative Extension Service, Manhattan, KS. Available online at <http://riskmgt.uwagec.org/MarketRisk/ProfitabilityDifferencesBetweenSteers&Heifers.pdf>.
- Williams, G.S., K.C. Raper, E.A. DeVuyst, D. Peel, and D. McKinney. 2012. Determinants of Price Differentials in Oklahoma Value-Added Feeder Cattle Auctions. *Journal of Agricultural and Resource Economics* 37(1):114-127.
- Zhao, H., X. Du, and D.A. Hennessy. 2011. Pass-Through in United States Beef Cattle Prices: A Test of Ricardian Rent Theory. *Empirical Economics* 40(2):497-508.

**Table 1. Summary Statistics for Carcass Performance, Live-Animal Characteristics, and Molecular Breeding Values ( $n = 2,976$ ).**

Variable	Mean	Standard Deviation	Minimum	Maximum
Carcass performance				
Marbling score	414.43	70.87	250.00	830.00
Yield grade	2.97	0.58	0.31	5.10
Hot-carcass weight, cwt.	7.20	0.70	4.58	9.83
Live-animal characteristics				
Placement weight, cwt.	6.55	1.24	2.94	11.16
Days-on-feed, days	171.51	29.09	106.00	238.00
Steer <sup>a</sup>	0.74			
Black <sup>b</sup>	0.78			
Molecular breeding values (MBV)				
Marbling MBV	-16.44	26.94	-119.37	68.26
Yield grade MBV	-0.06	0.07	-0.28	0.20
Rib-eye MBV, in <sup>2</sup>	-0.78	0.47	-2.16	1.38
Hot-carcass weight MBV, lbs.	28.43	9.06	-15.57	55.91
Average daily gain MBV, lbs./day	0.20	0.10	-0.12	0.48
Tenderness MBV, lbs. of WBSF <sup>c</sup>	-1.49	1.52	-5.90	2.92
Days-on-feed MBV, days	-2.83	2.95	-14.28	8.35

Note: Molecular breeding values (MBVs) are reported in the units of the trait and reflect the differences expected in animals across breeds compared to their contemporaries (Igenity, 2013). Therefore, mean MBVs offer little insight. Instead, the range of MBVs is more informative. For example, the range of average daily gain MBV suggests that the animal with the highest genetic potential for average daily gain in the sample would be expected, on average, to gain approximately 0.60 lbs./day more than the animal with the lowest genetic potential for average daily gain ( $0.48 - [-0.12] = 0.60$ ).

<sup>a</sup> Steer is a dummy variable equal to one if the animal was a steer and zero otherwise.

<sup>b</sup> Black is a dummy variable equal to one if the animal was black-hided and zero otherwise.

<sup>c</sup> Warner-Bratzler shear force.

**Table 2. Parameter Definitions, Baseline Values, and Ranges for Sensitivity Analysis**

Parameter	Definition	Baseline Value	Range for Sensitivity Analysis
$\alpha + \beta\mu_s$	Seller's expectation of profit (\$/head)	-\$63.16	-\$82.30 to -\$3.74
$\alpha + \beta\mu_b$	Buyer's expectation of profit (\$/head)	-\$82.30	—
$\beta\Sigma\beta'$	Variance of profit <i>within</i> a lot of cattle	608.57	100 to 2000
$\beta V_b \beta'$	Variance of the buyer's subjective prior distribution of expected profit (variance of profit <i>between</i> lots of cattle)	157.01	100 to 2000
$\frac{c}{m}$	Cost of genetic testing (\$/head)	\$0.40	\$0.10 to \$1.00

**Table 3. Average 2014 Base Price and Yield Grade, Quality Grade, and Hot-Carcass Weight Premiums and Discounts for Grid Pricing**

Grid Component	Premium/(Discount)
	\$/cwt.
Base price	
Steers	\$244.22
Heifers	\$244.21
Quality grade adjustment	
Prime	\$19.26
Choice	\$0.00
Select	(\$8.63)
Standard	(\$20.84)
Yield grade adjustment	
1.0-2.0	\$4.58
2.0-2.5	\$2.25
2.5-3.0	\$2.13
3.0-4.0	\$0.00
4.0-5.0	(\$8.63)
>5.0	(\$13.64)
Hot-carcass weight adjustment	
400-500	(\$25.42)
500-550	(\$22.19)
550-600	(\$2.93)
600-900	\$0.00
900-1000	(\$0.24)
1000-1050	(\$2.27)
>1050	(\$23.24)

Source: USDA Agricultural Marketing Service (AMS) reports LM\_CT150 and LM\_CT169 obtained from the Livestock Marketing Information Center (LMIC) spreadsheets (LMIC, 2015; USDA AMS, 2015).

**Table 4. Feedlot Profit Mixed Model Regression Estimates ( $n = 2,976$ )**

Variable	Coefficient	Standard Error <sup>a</sup>
Intercept	-1151.33***	353.24
Placement weight	103.82***	25.15
Days-on-feed	8.08**	3.35
Days-on-feed squared	-0.01	0.01
Placement weight $\times$ days-on-feed	-0.60***	0.14
Steer <sup>b</sup>	-75.25***	10.88
Black <sup>c</sup>	2.08	7.58
Marbling MBV <sup>d</sup>	0.44***	0.12
Yield grade MBV	17.54	55.70
Rib-eye area MBV	-23.40**	9.58
Hot-carcass weight MBV	1.55***	0.31
Average daily gain MBV	27.21	32.32
Tenderness MBV	1.86	2.10
Days-on-feed MBV	0.30	0.84
quasi- $R^{2e}$	0.29	

Note: Dependent variable is estimated feedlot profit (\$/head) from equation (16). Single, double, and triple asterisks (\*, \*\*, \*\*\*) indicate significance at the 10%, 5%, and 1% level.

<sup>a</sup> Standard errors are estimated using “sandwich estimators” to obtain estimates of standard errors that are consistent in the presence of nonnormality and static heteroskedasticity (White, 1982).

<sup>b</sup> Steer is a dummy variable equal to one if the animal was a steer and zero otherwise.

<sup>c</sup> Black is a dummy variable equal to one if the animal was black-hided and zero otherwise.

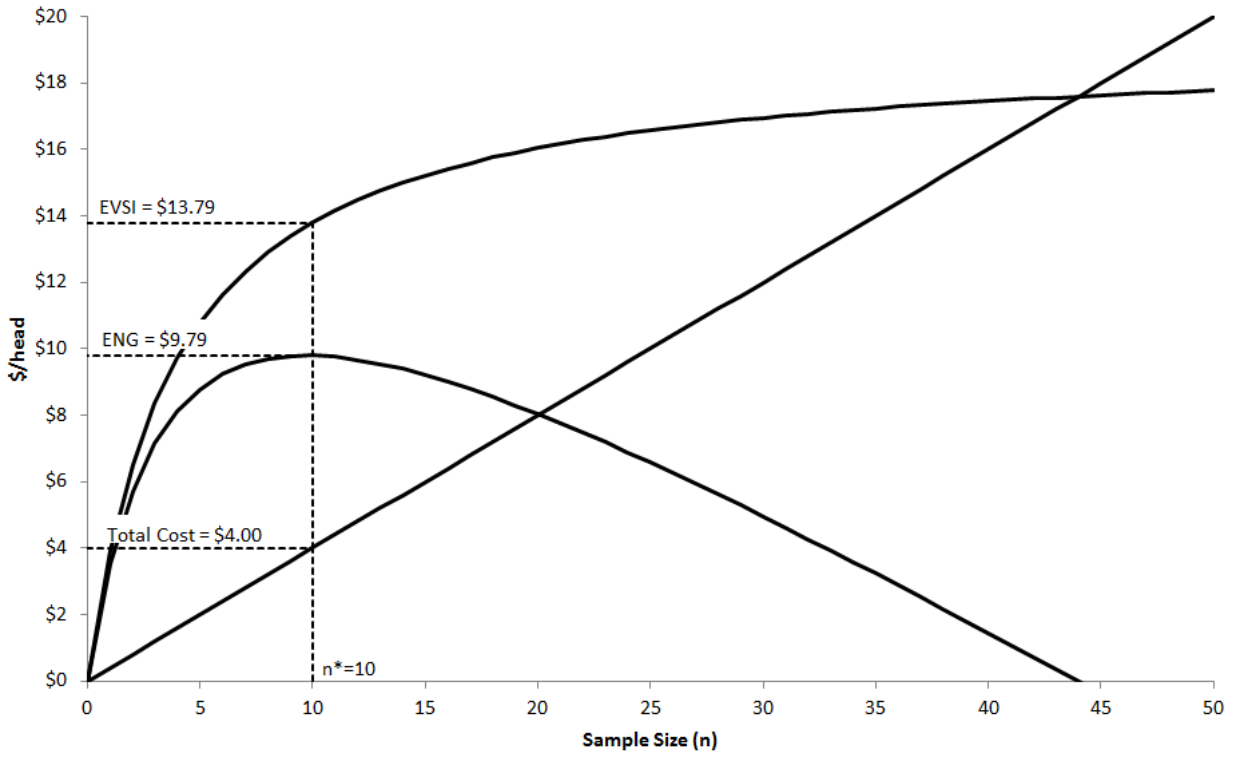
<sup>d</sup> Molecular breeding value.

<sup>e</sup> The quasi- $R^2$  is calculated as the squared correlation of the actual and predicted values including the contemporary group random effect.

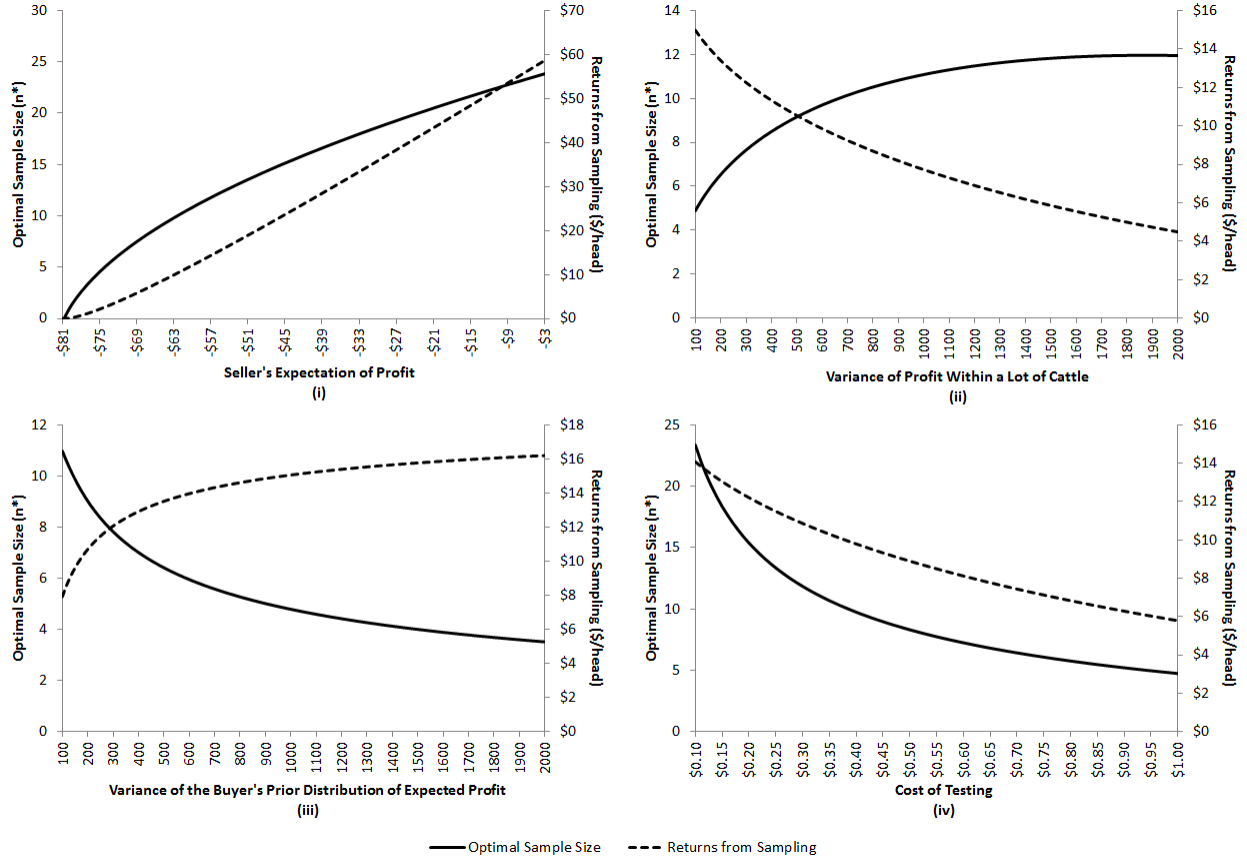
**Table 5. Random Effects Model Estimates of Variance of Molecular Breeding Values  
Between and Within Lots of Cattle (*n* = 2,976)**

Molecular Breeding Value	<i>Between</i> Variance	<i>Within</i> Variance
Marbling	191.730***	543.420***
Yield grade	0.001***	0.004***
Rib-eye area	0.046***	0.173***
Hot-carcass weight	11.184***	72.080***
Average daily gain	0.001***	0.008***
Tenderness	1.072***	1.332***
Days-on-feed	0.297***	8.425***

Note: Single, double, and triple asterisks (\*, \*\*, \*\*\*) indicate significance at the 10%, 5%, and 1% level.



**Figure 1. Expected value of sample information (EVSI), expected net gain (ENG), and total cost as a function of sample size ( $n$ ) at the baseline parameter values**



**Figure 2. Sensitivity analysis of optimal sample size and the returns from sampling with respect to (i) the seller's expectation of profit ( $\alpha + \beta\mu_s$ ), (ii) the variance of profit within a lot of cattle ( $\beta\Sigma\beta'$ ), (iii) the variance of the buyer's prior distribution of expected profit ( $\beta V_b\beta'$ ), and (iv) the per-head cost of genetic testing ( $\frac{c}{m}$ )**