

High-impact Data Programs: A framework for expanded opportunities in genetic evaluation

Lane K Giess*

*Department of Animal Sciences, Colorado State University, Fort Collins, CO, USA 80523-1171

Introduction

Genetic evaluations for beef cattle have evolved significantly in the United States (Golden et al. 2009), where now the information systems for genetic predictions have been bolstered by millions of phenotypes, pedigree, and genomic data points. These data are paramount to the development of robust genetic predictions that allow producers to make more informed selection decisions. Many data points are readily available and relatively easy to collect, inherently lending greater statistical power to their genetic predictions. However, for the performance traits where the data collection is difficult, time consuming or where no existing data collection structure has been clearly defined, these genetic predictions lack informative power and accuracy on young animals.

One of the key goals for genetic evaluations should be to explore the existing data structure of phenotypic records obtained from producers and partner organizations and develop a thorough understanding of possible opportunities for improved genetic prediction and new trait development. These opportunities are not limited to the use of indicator traits to better the genetic prediction of an economically relevant trait (Golden et al. 2000), but rather may include the use of by-product data and leveraging associated opportunities from existing data collection programs. This framework explores the use of high-impact data programs for existing and future genetic evaluations for performance traits in beef cattle.

High-impact Data

High-impact data may be defined as the value placed on phenotypic information for which a genetic prediction suffers from a lack of quality inputs. An example of this is actual carcass

29 phenotypes, where the genetic predictions for marbling, carcass weight, ribeye area, fat
30 thickness, and tenderness suffer from a critical mass of data. Though ultrasound carcass data
31 provides a satiable alternative to bolster carcass genetic predictions, this may be considered low
32 value data due to sometimes poor genetic correlations of ultrasound carcass to actual carcass
33 traits (AGI, 2013). Angus Genetics Incorporated (2013) reported correlation estimates for actual
34 backfat to ultrasound backfat, actual marbling score to ultrasound marbling, actual ribeye area to
35 ultrasound ribeye area, and carcass to yearling weight of 0.65, 0.71, 0.65, and 0.75 respectively.
36 Few comparisons outside of structured studies have evaluated the correlation of ultrasound to
37 actual carcass data, so evaluating the genetic correlation between ultrasound data and the genetic
38 prediction provide additional insight. International Genetic Solutions (IGS) reports genetic
39 correlation between ultrasound marbling and marbling EPD, ultrasound fat and fat EPD,
40 ultrasound ribeye area and ribeye area EPD and post weaning gain and carcass weight as 0.77,
41 0.45, 0.52, 0.55 respectively (IGS, 2022).

42 The evaluation of what constitutes high-impact data may be debated among industry
43 organizations, however for the purposes of this framework, those phenotypes reported on fewer
44 than 20 percent of the animals enrolled in a given year are considered high-impact data. In
45 addition, traits that are currently being investigated within academic institutions and hold
46 promise for future genetic prediction development may also be considered high-impact data as
47 the need for critical mass is necessary. The process of evaluating new or novel high-impact data
48 is outlined by Garrick et al. (2014a). The following are a sample of notable systems contributing
49 opportunities for high-impact data collection.

50 **Whole Herd Reporting**

51 Inventory-based whole herd reporting (WHR) is a data program recommended by many industry
52 organizations which provides a structure for the collection of whole contemporary group
53 information on mature females and their offspring. Additionally, WHR provides added metrics
54 on a range of economically relevant traits – primarily reproductive efficiencies and longevity
55 (Hough and Ponder, 2001; Cammack et al., 2009), where stayability predictions see added
56 benefit because the inventory system provides the culling and calving performance data needed
57 for effective random regression modelling (Jamrozik et al. 2013).

58 The implementation of WHR programs amongst industry organizations vary, and the benefits
59 depend upon how WHR is accepted among users. For example, some organizations require
60 mandatory WHR from all users. While this forces every animal record enrolled in a breed
61 registry to originate from whole and complete contemporary groups, this may include the added
62 risk of inaccurate data reporting. The requirements of inventory reporting can be onerous to
63 many cattle producers and rather than collect every data point required on every animal, they
64 may opt to report inaccurate information to satisfy the system inventory reporting requirements.

65 Adversely, selective WHR systems provide producers the option to register an entire inventory
66 as well as selective enrollments. This method may more appropriately capture quality data points
67 from invested breeders yet continues to allow a portion of users the option to selectively report
68 the animals they deem fit for registration. This disrupts the evaluation of whole contemporary
69 group comparisons and loses the ability to capture a critical mass of fertility and longevity
70 measures.

71 Regardless of implementation, WHR programs provide vast opportunities for new or improved
72 genetic prediction. Giess et al. (2021) outlined how enrollment, productivity and disposal codes
73 from the American Simmental Association were used in a logic-based system to generate heifer
74 pregnancy (HPG) phenotypes. This implementation of what may usually be considered by-
75 product information, provides added opportunity for the development of new genetic predictions
76 or inherently less onerous data collection programs for producers. Table 1 contains a list of data
77 points collected from WHR programs and existing or future uses for those records.

78 Table 1 suggests a small sampling of possibilities for the implementation of new or novel genetic
79 predictions, some of which are standalone predictions, whereas others augment an existing
80 prediction model improvement. In the case of Stayability, where the realized phenotype is
81 measured in multiple years or progeny (Martinez et al. 2005; Brigham et al. 2007; Jamrozik et al.
82 2013), optimization for the genetic prediction is extremely consequential due to the high
83 economic importance of the trait (Garrick, D. 2006; Garcia et al. 2015).

84 **Structured Sire Progeny Testing**

85 Structured sire progeny testing programs is the practice of placing registered sires in several
86 cooperating herds with the sole responsibility of collecting and reporting data on progeny. This is

87 no novel practice within the beef industry and the benefits of data obtained from progeny clearly
88 outweigh the value of an animal's own phenotype (Robertson and Rendel, 1950). This practice
89 utilizes artificial insemination to distribute genetics quickly and cheaply to participating
90 producers and leverages the use of proven reference sires to facilitate comparisons of young bulls
91 from different AI groups and years (Foulley et al. 1982). In recent years the prevalence of these
92 enterprises have dwindled due to difficulty of collecting data, high costs, and the rapid
93 acceptance of genomic capabilities as a substitute. Despite the lack of structured progeny testing
94 sites, these systems provide a vast amount of unbiased data for genetic prediction, and notably,
95 rapidly increase the accuracy of prediction on young sires through sheer amounts of data
96 collection.

97 The present framework is not meant to justify structured progeny testing, but rather explore its
98 uses when coupled with WHR and when such progeny testing programs demand emphasis for
99 rare and meaningful data collection. It is difficult to justify the cost of paying producers to
100 collect weight data since many producers are well attuned to collecting those measures, yet for
101 genetic evaluations suffering from a lack of phenotypes in a given trait, structured sire progeny
102 testing may be a viable option. Presently, many of the existing sire progeny testing programs
103 implemented among industry organizations put emphasis on the collection of actual carcass data.
104 As described earlier, carcass data is a valuable phenotype to pursue due to its difficulty of
105 collection and lack of substantial indicator metrics.

106 The inputs necessitated from personnel, costs, and time associated with the management of
107 structured sire progeny tests provide added justification to expand upon data collection emphasis
108 and pursue additional opportunities. While animals with actual carcass data reflects less than
109 approximately 3% of annual enrollments for most industry organizations (ASA Annual Report,
110 2022), even fewer are animals with actual carcass data and genomic information. Structured sire
111 progeny tests willing to add the additional cost of genotyping the progeny enrolled in the
112 program may provide increased value to the overall genetic evaluation. It's clear the value of
113 obtaining critical mass of phenotypes on animals with genomic markers increases the efficacy of
114 each genomic panel as well as increases the relative accuracy of non-genotyped individuals
115 (Garrick, D. J. 2011).

116 Carcass performance is closely tied to economic success, and thus emphasis should also be
117 placed on trait complexes influencing the amount of retail product being produced. One area of
118 opportunity is to address the consistent increase of feedlot death over the past 20 years. Figure 1
119 shows the number of feedlot deaths annually since 2000 (USDA, 2022).

120 There is developing evidence, suggesting bovine congestive heart failure (BCHF) in late-stage
121 feedlot cattle may be heritable and the risk of late-stage feedlot death might be improved through
122 genetic selection (Kukor et al. 2021). Specific markers have been investigated for association
123 with BCHF where ARRDC3 and NFIA variants showed small yet significant association with
124 the trait (Heaton et al. 2019; Heaton et al. 2022). Implementation of these markers in traditional
125 genetic evaluation systems should be done cautiously, as the selection for genetically superior
126 animals using only genomic indicators may prove ineffective without critical mass of phenotypes
127 and pedigree linkages available.

128 Young sire progeny tests provide an opportunity for data collection on heart and heart fat scores
129 on animals with already important data collection practices in place. Establishing a connection
130 between risk of feedlot death, pulmonary arterial pressure, heart score, carcass performance and
131 the genomic association with all traits provides a valuable opportunity for breed organizations to
132 invest in such programs.

133 **Nondescript Cattle Populations**

134 Traditional genetic evaluation has been predicated upon the use of leveraging sire-identified
135 breeding animals and associated data on themselves and their progeny to build genetic
136 predictions. This practice was improved with the use of genomic markers incorporated into
137 existing predictions and alongside non-genotyped animals (Garrick et al. 2014b).

138 Whereas genetic predictions are designed to predict merit for economically relevant traits in
139 commercial cattle production, most of the data being submitted to industry organizations are
140 collected on breeding stock, who are rarely exposed to the same level of rigor and selection
141 found in the commercial industry (Garrick and Golden 2009). This is primarily found in feeding
142 cattle, where few seedstock producers retain ownership and slaughter their own calves, and if
143 they do, the resulting effort provides data limited to animals deemed unsuitable for breeding.

144 This point is augmented by the fact nucleus breeders actively accelerate genetic progress by

145 reducing generation interval. While beneficial for increasing the rate of genetic progress, this
146 practice actively eliminates opportunity for data to be collected mature females including;
147 stayability, fertility and mature weight.

148 These tradeoffs distinguish an inherent weakness in existing genetic evaluations where
149 commercial animals are not represented in the genetic predictions for which commercial
150 producers are selecting their breeding animals.

151 Historically, the viability of using commercial populations of cattle for genetic prediction was
152 not possible since sire-identification was not widespread. However, determining relatedness of
153 commercial animals is possible through the use of genomic parent verification. In populations
154 where large numbers of parent animals are genotyped this can be accomplished, though perhaps
155 not cost-effectively. Relatedness in the absence of any pedigree information is supported (Tapio
156 et al. 2010) and developing linkages between non-pedigreed and sire-identified populations
157 provide opportunity for rare data collection.

158 Individually genomically sampling animals may prove to be too costly to be implemented at a
159 large enough scale to truly benefit genetic prediction. The prospect of genomic pooling provides
160 an innovative opportunity to cut costs drastically while facilitating large volumes of data
161 collection on commercial populations (Reverter et al. 2016). Genotyping technology is in its
162 infancy among commercial producers, yet readily available and implemented in a growing
163 population of nucleus breeders. Should genomic pooling provide the necessary catalyst to link
164 commercial data to seedstock populations, the industry will benefit from more powerful and
165 unbiased predictions.

166 These commercial populations provide the greatest opportunity for the collection of rare and
167 meaningful data already discussed. Developing genomic association with impactful traits from
168 non-descript populations has shown some merit in the absence of pedigree information.

169 However, for the most effective genetic tools, the use of robust traditional genetic evaluation is
170 more appropriate. Investigation into opportunities for the evaluation of genomically-sampled
171 non-descript cattle alongside traditional registered populations may prove useful, if viable.

172 **Structure for High-Impact Data Programs**

173 As described in Garrick et al. (2011), phenotyping is now the limiting factor in expanding the
174 offering of traits routinely recorded in genetic evaluation systems. Small breed registry's may not
175 have the resources nor technical ability to greatly modify their genetic evaluation and will suffer
176 from the lack of high-impact data programs funneling rare and meaningful phenotypes into their
177 databases. It is with this in mind the necessity of collaboration among industry organizations
178 provides an opportunity for enterprises to share resources and implement shared high-impact
179 data programs.

180 As the opportunities for high-impact data programs are endless, beef cattle registry services and
181 industry organizations should develop a protocol for evaluating their existing genetic evaluation
182 services and consider unique solutions to build in added statistical power to predictions with rare
183 or novel data points. The development of high-impact data programs may facilitate the mass
184 accumulation of desirable data points. The following structure provides a starting point to
185 evaluate the use of high-impact data programs among industry organizations.

- 186 1. Critically evaluate all existing predictions and data influences implemented in the genetic
187 evaluation.
 - 188 a. Take into consideration:
 - 189 i. Average accuracy values
 - 190 ii. Percentage of phenotypes/genotypes reported to animals enrolled annually
 - 191 iii. Percentage of animals genotyped with phenotypes
 - 192 iv. Genomic progeny equivalents
- 193 2. Critically evaluate all existing data collection initiatives and programs offered to users of
194 the genetic evaluation.
 - 195 a. Are existing data programs high impact?
 - 196 b. Do the existing data programs further the mission and goal of the organization?
- 197 3. Evaluate the merit and cost of implementing a high-impact data program for users of the
198 genetic evaluation.
 - 199 a. Take into consideration:
 - 200 i. Cost of data collection
 - 201 1. Phenotype subsidy
 - 202 2. Genotype subsidy

- 203 ii. Volume of participants
- 204 iii. Staff management and travel
- 205 4. Clearly identify goals for which phenotypes, genotypes and research traits you wish to be
- 206 emphasized.
- 207 a. Take into consideration:
- 208 i. Weaknesses in the genetic evaluation
- 209 1. Percentage of phenotypes/genotypes reported to animals enrolled
- 210 annually
- 211 2. Average accuracy values for traits
- 212 5. Survey users to determine the ideal demographic and isolate candidate herds or
- 213 populations to include in the proposed high-impact data program.
- 214 a. Identify the number of participants
- 215 i. Size of cow herd
- 216 b. Familiarity with data collection
- 217 c. Education of unfamiliar trait collection
- 218 6. Develop protocol and operating policy to set parameters for data collection and facilitate
- 219 the success of the high-impact data collection program.
- 220 7. Implement the high-impact data program with the intent to review and modify as needed.

221 **Conclusion**

222 Genetic evaluations primarily focus on the development of economically relevant traits which
223 allow producers to actively select for improvement in areas of their breeding program directly
224 affecting their bottom-line. While these predictions are paramount to the success of the beef
225 business, a special emphasis on the quality of prediction should also be taken into consideration.
226 The absence of a critical evaluation and validation of existing predictions and lack of well-
227 defined goals in genetic evaluation systems results in lost opportunity to place special emphasis
228 for creative solutions. It is with this outlined framework, that genetic evaluations may start to
229 evaluate their current data programs and how extra information may be captured for relevant
230 traits.

231 However, as industry genetic evaluations continue to scale in volume with admixed populations,
232 there is a need to fuel specific trait predictions with a higher volume of quality phenotypes,

233 genotypes and associated indicator traits. Research and development efforts into novel traits
234 plays a critical role in improving these evaluations, through understanding emerging trait
235 complexes and correlated inputs. High-impact data programs provide a nexus of data collection
236 opportunities where rare and meaningful data points can be accumulated on whole contemporary
237 groups.

238 By building a framework for evaluating the merit of existing genetic predictions and developing
239 creative and unique data programs to collect effective indicator traits, genomic information and
240 research inputs it is possible to build a more interconnected and powerful genetic evaluation for
241 beef cattle.

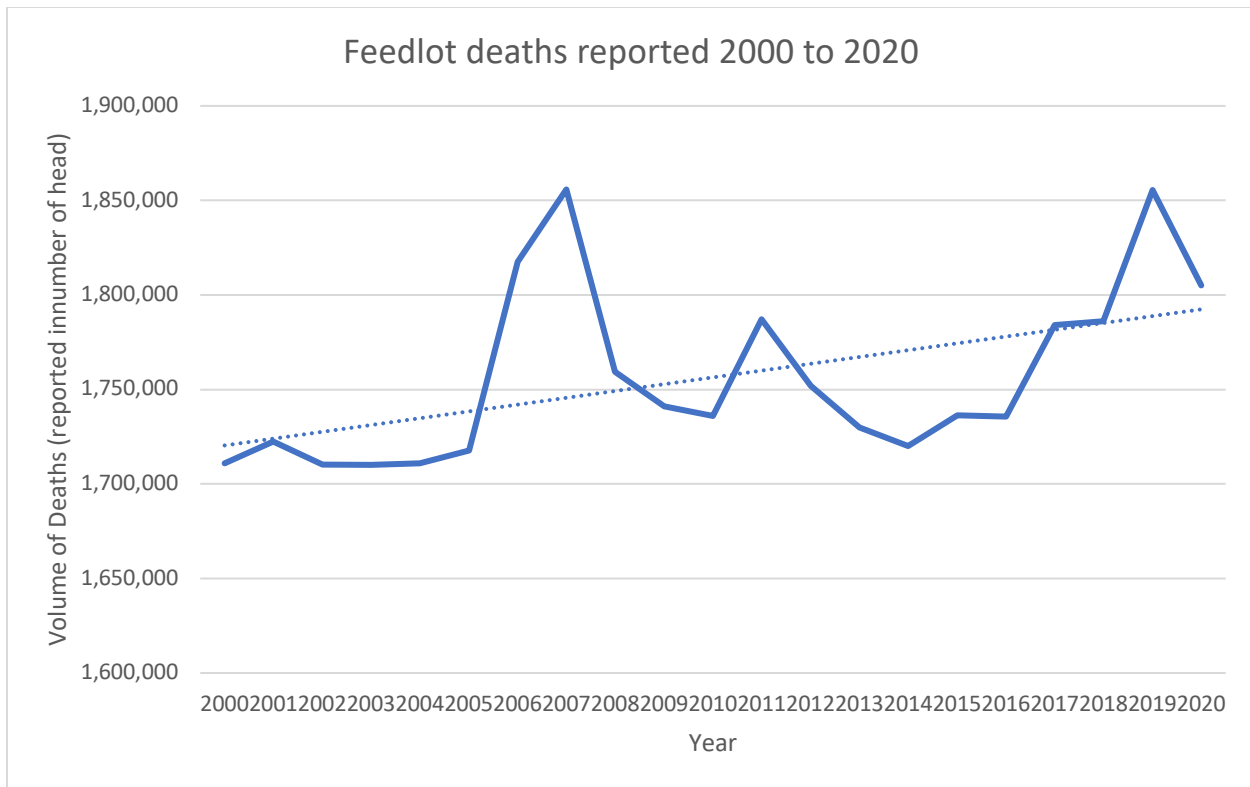
242

243 Table 1: A sampling of proposed and existing uses for data collected from whole herd reporting
 244 programs.

Whole Herd Reporting Data	Existing Genetic Prediction Uses	Future Genetic Prediction Uses
Cow Inventory	Whole contemporary group evaluation: Stayability Heifer pregnancy Maternal calving ease Maternal weaning weight Maternal effects for existing predictions Mature weight	Sustained cow fertility Maternal effects for existing predictions Stayability to older ages Gestation Length Maintenance cost
Calf Records	Whole contemporary group evaluation: Growth Maternal calving ease Milk Average daily gain	Pre-weaning survival Post-weaning survival Days to weaning
Enrollment/Productivity/Disposal Codes	Adjustment of cow contemporary group	Heifer pregnancy In-utero survival Respiratory health Udder quality Docility Maintenance Lifetime value Augment existing predictions: Feet and legs Stayability

245

246 Figure 1: Graph of annual beef cattle feedlot death loss since 2000 – excluding calves



247

248

249 Literature Cited:

250 AGI. 2013. FAQ – Ultrasound vs genomic data vs actual carcass data.

251 https://www.cuplab.com/Files/content/ultrasound_FAQ_1-4-18.pdf (Accessed April 11,
252 2022)

253 ASA Annual Report. 2022. American Simmental Association Annual Report.

254 https://issuu.com/asapublications/docs/annual_report_2021 (Accessed April 12, 2022)

255 Brigham, B. W., S. E. Speidel, R. M. Enns, and D. J. Garrick. 2007 Stayability to alternate ages.

256 Western Section Proceedings. Vol 58:27-30

257 Foulley, J. L., L. R. Schaeffer, H. Song, and J. W. Wilton. 1982. Progeny group size in an

258 organized progeny test program of AI beef bulls using reference sires. Can. J. Anim. Sci

259 63.1:17-26

260 Garcia, J., D. Anderson, A. Herring, and D. Riley. 2015. Economic analysis selecting for cow

261 longevity. Southern Ag. Econ. Assoc. 2015 Annual Meeting. No. 1375-2016-109490.

262 Garrick, D. J. 2006. Evaluation, interpretation, and relevance of Stayability in beef cattle genetic
263 improvement. *J. Anim. Sci.* 84(Suppl 2):50

264 Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the
265 United States. *Gen. Sel. Evol.* 43:17 <http://www.gsejournal.org/content/43/1/17>
266 (Accessed April 12, 2022)

267 Garrick, D. J., and B. L. Golden. 2014a. Producing and using genetic evaluations in the United
268 States beef industry of today. *J. Anim. Sci.* 87(E. Suppl.):E11-E18. doi:10.2527/jas.2008-
269 1431

270 Garrick, D. J., J. C. M. Dekkers, B. L. Golden, and R. L. Fernando. 2014b. Bayesian prediction
271 combining genotyped and non-genotyped individuals. 10th World Congress on Genetics
272 Applied to Livestock Production.

273 Giess, L. K., M. G. Thomas, S. E. Speidel, M. M. Culbertson, W. R. Shafer, S. C. McGuire, and
274 R. M. Enns. 2021. Whole herd reporting data from the American Simmental Association
275 as a data source for heifer pregnancy phenotypes. *Western Section Proceedings. T. Anim.*
276 *Sci.* 5:S199-S203 doi.org/10.1093/tas/txab152

277 Golden, B. L., D. J. Garrick, S. Newman, and R. M. Enns. 2000. Economically relevant traits: A
278 framework for the next generation of EPDs. *Proceedings of the 32nd Research*
279 *Symposium and Annual Meeting of the Beef Improvement Federation.*

280 Golden, B. L., D. J. Garrick, and L. L. Benyshek. 2009. Milestones in beef cattle genetic
281 evaluation. *J. Anim. Sci.* 87(E. Suppl.):E3-E10 doi: 10.2527/jas.2008-1430

282 Heaton M. P., A. S Bassett, K. J. Whitman, G. M. Krasfur, S. I. Lee, J. M Carlson, H. J Clark, H.
283 R. Smith, M. C. Pelster, V. Basnayake, D. M. Grotelueschen, and B. L. Vander Ley.
284 2019. Evaluation of EPAS1 variants for association with bovine congestive heart failure.
285 *F1000Research*, 8

286 Heaton, M. P., G. P. Harhay, A. S. Bassett, H. J. Clark, J. M. Carlson, E. E Jobman, H. R. Sadd,
287 M. C. Pelster, A. M. Workman, L. A. Kuehn, T. S. Kalbfleisch, H. Piscatelli, M. Carrie,
288 G. M. Krasfur, D. M. Grotelueschen, and B. L. Vander Ley. 2022. *F1000Research* 11:
289 385

290 IGS. M. M. Culbertson. 2022. Personal communication on genetic correlations used for carcass
291 predictions within the IGS multi-breed genetic evaluation powered by BOLT. April 11,
292 2022.

293 Jamrozik J., S. McGrath. R. A. Kemp, and S. P. Miller. 2013. Estimates of genetic parameters
294 for Stayability to consecutive calvings of Canadian Simmentals by random regression
295 models. *J. Anim. Sci.* 91:3634-3643 doi:10.2527/jas2012-6126

296 Kukor, I. M., M. G. Thomas, R. M. Enns, T. Holt, S. E. Speidel, M. A. Cleveland. B. P. Holland,
297 A. B. Word, and G. B. Ellis. 2021. Sire differences within heart and heart fat score in
298 beef cattle. *Western Section Proceedings. T. Anim. Sci.* 5:S149-S153
299 <https://doi.org/10.1093/tas/txab147>

300 Martinez, G. E., R. M. Koch, L. V. Cundiff, K. E. Gregory, S. D. Kachman, and L. D. Van
301 Vleck. 2005. Genetic parameters for stayability, stayability at calving, and stayability at
302 weaning to specified ages for Hereford cows. *J. Anim. Sci.* 83:2033-2042.

303 Reverter, A., L. R. Porto-Neto, M. R. S. Fortes, R. McCulloch, R. E. Lyons, S. Moore, D. Nicol,
304 J. Henshall, and S. A. Lehnert. Genomic analysis of tropical beef cattle fertility based on
305 genotyping pools of Brahman cows with unknown pedigree. *J. Anim. Sci.* 94
306 doi:10.2527/jas2016-0675

307 Robertson, A., and J. M. Rendel. 1950. The use of progeny testing with artificial insemination in
308 dairy cattle. *J. Genetics.* 50.1:21-31

309 Tapio, I., M. Tapio, M. Li. R. Popov. Z. Ivanova, and J. Kantanen. 2010. Estimation of
310 relatedness among non-pedigreed Yakutian cryo-bank bulls using molecular data:
311 implications for conservation and breed management. *Gen. Sel. Evo.* 42:28
312 <http://www.gsejournal.org/content/42/1/28> (Accessed April 12, 2022)

313 USDA. 2022. Economic Research Service – National Agricultural Statistics Service quickstats
314 on cattle, (excl calves) - loss, death, measured in head since 2020.
315 <https://quickstats.nass.usda.gov/> (Accessed April 10, 2022)