

ESTIMATING THE AGE AND SEX COMPOSITION OF FAUNAL ASSEMBLAGES WITH BAYESIAN MULTILEVEL MIXTURE MODELS

Jesse Wolfhagen¹

¹Department of Anthropology, Purdue University, West Lafayette, Indiana, USA

email: jl.wolfhagen@gmail.com

R Markdown version last compiled on 2022-06-23

ABSTRACT

Understanding the age and sex composition of zooarchaeological assemblages can reveal details about past human hunting and herding strategies as well as past animal biometry and behavior. However, estimates of the sex ratio in an assemblage typically rely on a small subset of element portions, making it difficult to extrapolate these results to the entire assemblage. This paper describes a method to use zooarchaeological remains with standard biometric measurements to estimate the age and sex composition of the assemblage, focused on immature, adult-sized female, and adult-sized male specimens. The model uses a Bayesian framework to avoid the overfitting that plagues optimization-based methods, ensuring that the parameter estimates are biologically meaningful. The accuracy of the model is tested using simulated assemblages from known-age and sex specimens, showing that the model can accurately estimate the parameters and composition of assemblages. This includes the composition of the entire assemblage, not just the measured specimens. Two case studies also show how the model can be applied to archaeological scenarios. The first, focused on sheep from Neolithic Pinarbaşı B, highlight the importance of accounting for immature animals in an assemblage. The second, focused on four assemblages from 7th-6th Millennium BCE northwestern Anatolia, showcase how to compare biometric results from the model and how to use the model's results to estimate sex-specific fusion rates. This modeling framework provides a new avenue for investigating long-term trajectories in animal biometry alongside contextual analyses of past human choices in butchery and consumption.

Keywords: *Zooarchaeology, Biometry, Logarithmic size index (LSI), Domestication, Bayesian statistics.*

¹ 1. INTRODUCTION

² Different hunting and herding strategies target specific classes of animals among a herd that are determined
³ by the animal's age and sex (Dahl and Hjort 1976; Stiner 1990). In addition to human-driven goals, sex
⁴ differences in habitat use, diet quality, and reproductive capabilities among ungulate prey species contribute
⁵ to the susceptibility and desirability of males and females at different ages to human exploitation (Corti and
⁶ Shackleton 2002; Post et al. 2001; Ruckstuhl and Neuhaus 2002; Ruckstuhl 2007; Saïd et al. 2011). The com-
⁷ bination of these factors would thus affect the probability that bones from different classes of animals would
⁸ become incorporated into a zooarchaeological assemblage, even if taphonomic factors make it impossible to
⁹ reconstruct death assemblages exactly (Lyman 2008). The age and sex composition of zooarchaeological
¹⁰ assemblages can therefore reflect anthropologically-relevant aspects of past hunting strategies—like seasonal
¹¹ site use and scale of exploitation (Speth 2013)—or general management goals of past herding strategies (e.g.,
¹² Payne 1973; Redding 1984).

¹³ Reconstructing the age and sex composition of a zooarchaeological assemblage can enrich our under-
¹⁴ standing of past human-animal interactions by complementing mortality profiles and inter-assemblage com-
¹⁵ parisons. However, this task is complicated by the disaggregated nature of faunal assemblages. Because
¹⁶ articulated remains are rare, zooarchaeologists typically cannot relate elements that are morphologically dis-
¹⁷ tinct between the sexes (e.g., the pelvis) to other elements that can provide information about the animal's
¹⁸ age-at-death (e.g., limb bones or mandibles). We can, though, take advantage of the general pattern of
¹⁹ sexual dimorphism among ungulate taxa to use size differences in limb bones to distinguish between males
²⁰ and females. When combined with size index methods that allow researchers to relate measurements from
²¹ different elements together, this approach allows general descriptions of the sex ratio in an assemblage that
²² can be used to identify changes in these sex ratios or overall biometry over time (e.g., Grigson 1989; Arbuckle
²³ and Atici 2013).

²⁴ Unlike morphological differences, though, much greater overlap exists in the metrics of individual male
²⁵ and female animals, meaning that no clear thresholds can be identified between males and females without
²⁶ maintaining a large area of unidentifiability; this is compounded by the inclusion of immature animals
²⁷ from unfused specimens or from elements that exhibit post-fusion growth (Popkin et al. 2012). These
²⁸ complications make it difficult to describe the sex composition of an assemblage based solely on absolute
²⁹ metrical identifications of sex since large parts of the assemblage may be left unidentified, decreasing our
³⁰ ability to make reliable inferences about the entire assemblage.

³¹ Mixture modeling provides a way out of this dilemma by using probabilistic sex identifications rather than
³² absolute ones. By describing an assemblage of faunal measurements as a mixture of measurements from male

33 and female specimens—with their own parameters for average size (μ) and variability (σ)—a mixture model
34 allows researchers to not only describe parameters of the overall assemblage but to estimate the probabilities
35 that a specific specimen is male or female (Dong 1997; Monchot and Léchelle 2002). The model can also be
36 described using a “latent state” nomenclature: measured specimens come from female or male specimens,
37 but we cannot directly observe the identity of the specimens. If the specimen’s group identity could be
38 observed directly, the calculation of the mixture components would be trivial; since it cannot be observed,
39 however, one must model probabilities of group membership based on the parameters of each group. The
40 theoretical advantages of mixture modeling—along with the flexibility of the models themselves—explain the
41 increasing popularity of the method in large-scale zooarchaeological analyses focused on biometric change
42 (e.g., Helmer et al. 2005; Arbuckle et al. 2016).

43 However, the interpretation of mixture models is not necessarily straightforward. First, the very flexibility
44 of mixture modeling means that there is no guarantee that the ‘groups’ identified by the expectation-
45 maximization algorithm are biologically meaningful. There is no straightforward way to penalize results
46 where population parameters (μ and σ) are out of line with our understanding of these parameters from
47 known-sex populations. Second, by having more parameters to estimate mixture modeling can exacerbate
48 issues with measurement aggregation and size indices (Wolfhagen 2020). Fitting mixture models on a small
49 assemblage from a single set of measurements can produce ‘over-fitted’ results that are not biologically
50 meaningful, but combining different sets of measurements together with a size index erases the possibility of
51 allometry between a reference animal and the archaeological material, introducing another potential source
52 of size variation than sexual dimorphism. Finally, probabilistic identifications from mixture modeling still
53 do not resolve the complications of age on body size—the impact of post-fusion growth and the inclusion
54 of unfused specimens. Because these specimens may come from young animals that have not reached adult
55 body size, they are very likely to be confused as female animals by a mixture model that only considers
56 two possible groups. However, it is important to include these specimens in our models because unfused
57 specimens from later-fusing elements can still show sexual dimorphism, meaning we would exclude males
58 from our models by focusing only on fused elements if males are killed at a younger age than females (Zeder
59 and Hesse 2000).

60 This paper describes a Bayesian approach to mixture modeling of faunal measurements that addresses
61 these weaknesses of currently-applied mixture models. The model uses informative priors derived from a
62 reference population of known-age and sex individuals to constrain population parameter estimates to be
63 biologically interpretable (Popkin et al. 2012). It uses multilevel modeling to take advantage of partial
64 pooling to address aggregation issues and directly estimate parameters for each set of measurements in the
65 analysis (Gelman 2006a; Wolfhagen 2020). It also includes a third group in the mixture model—this group of

66 “immature” specimens is meant to capture specimens that died before reaching adult body size. The model
67 also emphasizes inference of the entire assemblage rather than just the measured specimens by incorporating
68 observations of the sex ratio (from morphological data) and the proportion of immature specimens (from
69 fusion data) to inform population parameters of the proportions of these different groups. The model is
70 used on sixteen simulated assemblages derived from the Popkin et al. (2012) Shetland sheep population to
71 test its ability to accurately estimate the age and sex composition of assemblages. Two archaeological case
72 studies then show the applicability of the model to archaeological assemblages for reconstructing the age and
73 sex composition of assemblages and to highlight the importance of incorporating immature specimens into
74 mixture modeling analyses.

75 **2. A BAYESIAN MULTILEVEL MIXTURE MODEL FOR FAUNAL MEA- 76 SUREMENTS**

77 The Bayesian model developed for this paper describes assemblages of faunal measurements as a mixture of
78 immature animals, (adult-sized) females, and (adult-sized) males that have distinct sizes. The model uses
79 multiple sets of measurements (e.g., humerus distal breadth “humerus Bd”, radius proximal breadth “radius
80 Bp”, abbreviations following Driesch (1976)), which are first converted into a log-size index, or LSI, values
81 with a natural logarithm base (Meadow 1999; Wolfhagen 2020). LSI observations from measurement sets are
82 grouped together within a specimen to create individuals grouped into defined “element portions” that serve
83 as the basis for the mixture model analysis (see Table 1 for key term definitions). Element portions relate to
84 categories used for element fusion (e.g., distal humerus) to relate biometry and mortality profiles; specimens
85 that contain multiple element portions—like complete limb bones—are grouped into the later-fusing element
86 portion (compare to “skeletal part type” in Breslawski 2022).

87 The multilevel structure of the model uses partial pooling to allow the mixture model parameters to
88 vary between element portions while resisting overfitting. These element-specific parameters are related to
89 each other through hyper-parameters, which describe the average value of the model parameters and the
90 variability of model parameters across element portions (Wolfhagen 2020). Prior distributions of mixture
91 model hyper-parameters that are related to biometric variability in animal populations are derived from the
92 Popkin et al. (2012) Shetland sheep population.

Table 1: Definitions of key terms used in this paper

Term	Definition
Element Portion	A complete or partial skeletal element defined by the zooarchaeologist, used as the foundation of the multilevel model (e.g., "distal humerus"). Model produces parameter estimates for all defined element portions, so element portions must be non-overlapping. Analogous to "skeletal part type" in Breslawski (2022).
Measurement Set	Specific type of observed measurement (e.g., "humerus distal breadth").
Measured Assemblage	Assemblage of measured specimens from a defined number of element portions of a specific taxon.
Modeled Assemblage	Assemblage of specimens from a defined number of element portions of a specific taxon. Includes measured and non-measured specimens, though all element portions must have some number of measured specimens. Measurability is assumed to be effectively random (i.e., unrelated to whether the specimen came from an immature, female, or male individual).
Full Assemblage	Assemblage of specimens from a defined number of element portions of a specific taxon. Includes element portions that do not have any observed measurements. Measurability is assumed to be effectively random (i.e., unrelated to whether a specimen came from an immature, female, or male individual).

93 2.1 Mixture Model Likelihood

94 The central likelihood of the mixture model uses parameters that are specific to each element portion. These
95 parameters include the mixture proportions for the different components of immature animals, females, and
96 males (π_1, π_2, π_3), the average size for each component (μ_1, μ_2, μ_3), and the standard deviation for each
97 component ($\sigma_1, \sigma_2, \sigma_3$). For each element portion, immature animals are described with the first set of
98 parameters (π_1, μ_1, σ_1), adult-sized females with the second set of parameters (π_2, μ_2, σ_2), and adult-sized
99 males with the third set of parameters (π_3, μ_3, σ_3). This results in both a set of parameters that describe
100 the composition of the assemblage (of measurements from that element portion) and an equation to estimate
101 the probability that a particular specimen comes from an immature, adult female, or adult male individual.

$$\begin{aligned}
P(x|\pi_1, \pi_2, \pi_3, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3) = \\
& \pi_1 * \text{Normal}(x, \mu_1, \sigma_1) + \\
& \pi_2 * \text{Normal}(x, \mu_2, \sigma_2) + \\
& \pi_3 * \text{Normal}(x, \mu_3, \sigma_3)
\end{aligned} \tag{1}$$

102 In addition to a specimen's LSI value, the model needs two additional observed variables to address the
 103 potential presence of immature animals in the model. First, an indicator variable (Immature[specimen])
 104 describes whether the specimen could be from an immature animal based on the body part and the fusion
 105 characteristics (1 = potentially immature, 0 = cannot be immature). Data from known-age Shetland sheep
 106 show that specimens killed at younger than one year of age are significantly smaller than those killed at
 107 older ages, regardless of fusion status (Popkin et al. 2012). Thus, any measurement from an element with
 108 an unfused epiphysis or from an element that does not fuse or could fuse before one year of age is considered
 109 potentially immature. Measurements from specimens with fused epiphyses that fuse after one year of age
 110 are considered ineligible to be immature so the model does not consider that probability (it considers $\pi_1 = 0$
 111 for fitting that specimen).

112 Second, the proportion of specimens from an element portion that could be immature ($\text{proportion}_{\text{immature}}$)
 113 determines how to re-weight the mixture components (π_1 , π_2 , and π_3) for potentially-immature specimens
 114 from that element portion. The mixture components describe the entire assemblage for an element portion
 115 (a combination of potentially-immature and non-immature specimens), meaning that if $\pi_1 = 0.25$ we should
 116 expect 25% of the specimens to be from immature animals. If every specimen could be immature—say, for
 117 specimens from an early-fusing element—then the mixture components do not need to be re-weighted. If,
 118 however, only half of the specimens could be immature, then the mixture components must be re-weighted
 119 to ensure that the whole-assemblage proportions are correct. In such a case, we would expect half of the
 120 potentially-immature animals to be from immature animals if $\pi_1 = 0.25$ for the whole assemblage and we
 121 know that $\text{proportion}_{\text{immature}} = 0.50$ ($\frac{0.25}{0.50} = 0.50$); this same re-weighting would make it less likely that
 122 potentially-immature animals are from adult-sized female or adult-sized male animals. The code includes
 123 checks to ensure that π_1 can never exceed 1.00 after accounting for the proportion of immature specimens
 124 in cases where there are very few of such specimens and/or a relatively high expected assemblage-wide
 125 proportion of immature animals.

126 **2.2 Measurement Error and Observations**

127 The model estimates measurement error for different observed quantities that are used in the likelihood.
128 Measurements on both the archaeological specimens and the standard values used to calculate LSI values
129 are assumed to have a 1% measurement error (Popkin et al. 2012, fig. 6; Breslawski and Byers 2015). This
130 1% value comes from an evaluation of the Breslawski and Byers (2015) measurement data, where the average
131 standard deviation of repeated measurements on bison radius proximal breadth measurements was 1.1% the
132 average value of the measurement. This means that each measurement is given a standard deviation based on
133 the observed value, which is used to estimate the “modeled” measurement based on the observation we have.
134 These modeled measurements are then used to calculate the LSI value for that measurement ($LSI_{measurement}$).

135 Because specimens can have multiple measurements that are included in the mixture model on them (e.g.,
136 a distal humerus with both a Bd and BT observations or a complete radius with Bp and Bd observations),
137 the mixture model uses specimen-specific LSI values ($LSI_{specimen}$) that are related to $LSI_{measurement}$ values
138 in the same way. $LSI_{measurement}$ values are the “observations” with a standard deviation of 0.01 (in LSI_e
139 scale) based on intra-individual variation of LSI_e values for the Popkin et al. (2012) sheep using the *Ovis*
140 *orientalis* female standard animal (FMC 57951) from Uerpman and Uerpman (1994, Table 12).

141 *Observation Error Equations:*

$$\begin{aligned}\sigma_{measurement} &= \text{Measurement}_{observed} * 0.01 \\ \text{Measurement}_{observed} &\sim \text{Normal}(\text{Measurement}_{modeled}, \sigma_{measurement}) \\ \sigma_{reference} &= \text{Reference}_{observed} * 0.01 \\ \text{Reference}_{observed} &\sim \text{Normal}(\text{Reference}_{modeled}, \sigma_{reference}) \\ LSI_{measurement} &= \log_e(\text{Measurement}_{modeled}) - \log_e(\text{Reference}_{modeled}) \\ LSI_{measurement} &\sim \text{Normal}(LSI_{specimen}, 0.01)\end{aligned}\tag{2}$$

142 The model also uses observations of sex ratios and fusion rates to estimate different proportions of the
143 different components in the assemblage. These estimates are not directly used for element portion-specific
144 mixture model parameters but rather inform the hyper-parameters that describe the average value across
145 all element portions. These observations are part of a binomial process centered on the associated hyper-
146 parameters. As such, the measurement error in the observation is based on the total number of observed
147 specimens. The observation of the average proportion of immature specimens (μ_{π_1}) is based on the fusion
148 rate of proximal and middle phalanges, which fuse at around the same time as the estimated time that
149 animals reach adult body size in the Shetland sheep population (Popkin et al. 2012). The observation

150 of the average adult sex ratio—the proportion of females among mature animals $\left(\frac{\mu_{\pi_2}}{\mu_{\pi_2} + \mu_{\pi_3}}\right)$ —is based on
 151 the sex ratio of fused pelvises. In each case, the number of observable specimens (proximal or middle
 152 phalanges with observed fusion status, fused pelvises with estimated sexes) determines the measurement
 153 error using the binomial distribution. While this paper uses these observations for these hyper-parameters,
 154 other observations are possible and can be incorporated into the model in the same fashion.

155 *Demographic Observation Equations:*

$$\begin{aligned} N_{\text{unfused}} &\sim \text{Binomial}(N_{\text{ageable}}, \mu_{\pi_1}) \\ N_{\text{female}} &\sim \text{Binomial}(N_{\text{sexable}}, \frac{\mu_{\pi_2}}{\mu_{\pi_2} + \mu_{\pi_3}}) \end{aligned} \quad (3)$$

156 **2.3 Prior Distributions**

157 Prior distributions are central to Bayesian inference and describe one's prior beliefs in potential values of
 158 a model parameter. Prior distributions can be likened to a 'filter' from which parameter values are drawn
 159 to evaluate their fit with the data (Smith and Gelfand 1992). Several approaches exist for deciding how to
 160 describe this prior belief, ranging from 'objective' priors that provide equal weight to all possible values of
 161 a parameter to distinct distributions defined by a synthesis of previous or related research (Gelman 2006b).
 162 Objective priors poorly reflect our intuition about phenomena we are modeling, are inefficient, and can
 163 introduce errors into our analyses (Gabry et al. 2019); instead, 'weakly informative priors' or 'reference priors'
 164 use transformations of parameter values—like centering and scaling element portion-specific parameters—to
 165 describe variation in parameter values within reasonable values, with small deviations being more likely than
 166 large deviations (Gelman et al. 2008). Informative priors are derived from relevant knowledge, be it the
 167 results of earlier studies on the same subject, the quantification of expert opinion, or parameter values for
 168 related subjects (McCarthy and Masters 2005).

169 The prior distributions in this model are focused on describing previous beliefs about the value of the mix-
 170 ture model hyper-parameters, as the element portion-specific parameters are derived from these distributions.
 171 Mixture modeling performs well with unconstrained parameters values because it is more straightforward to
 172 estimate variation across element portions, meaning that constrained parameters—those where the range
 173 of possible values depends on the values of other parameters—must first be transformed into related uncon-
 174 strained parameters Betancourt (2017)]. The following sub-sections describe the necessary transformations
 175 for different sets of the mixture model parameters, describing the unconstrained parameters that can be
 176 modeled and the transformations that result in the mixture model parameters. While these sections use the
 177 mixture model parameter notations, prior distributions are for the 'central tendency' hyper-parameter for

178 the described unconstrained parameter.

179 It is important to remember that for all of these prior distributions are arbitrary choices made by the
180 researcher, regardless of whether the distributions are based on specific animal populations or on reference
181 priors. Other researchers could and should use different prior distributions to best reflect their intuition
182 about likely parameter values for particular case studies. This also highlight the importance of reporting
183 the prior distributions used in a Bayesian analysis to ensure replicability. Examining the implications of
184 different prior distributions is an important step in the development of Bayesian models, one that should be
185 regularly tested even before models are fit to datasets (Gelman, Vehtari, et al. 2020).

186 **2.3.1 Mixture Proportion Priors**

187 The three mixture components (π_1 , π_2 , π_3) are a three-value unit simplex, meaning that the values are
188 constrained as a group to sum up to one. This means that the simplex can be described by only two variables
189 because the third value cannot vary once those two values are known. The model uses two unconstrained
190 variables (θ_1 and θ_2) to describe the unit simplex of π values. These θ values are related back to π values
191 using the following ‘stick-breaking’ transformation by iteratively estimating the relative proportions of the
192 simplex taken up by each θ value (Team 2022, sec. 10.7).

193 *Stick-Breaking Transformations:*

$$\begin{aligned}\pi_1 &= \text{logit}^{-1}(\theta_1 + \log(0.5)) \\ \pi_2 &= (1 - \pi_1) * \text{logit}^{-1}(\theta_2 + \log(1)) \\ \pi_3 &= 1 - (\pi_1 + \pi_2)\end{aligned}\tag{4}$$

194 The distribution of potential π values in an assemblage is relatively broad—there is no reason to think
195 that any combination of the three mixture components cannot occur. There are, however, general expecta-
196 tions about the relative proportion of each group (immature, female, and male) in a zooarchaeological
197 assemblage that we can use to inform prior distributions of θ values. The θ_1 value can be directly related
198 to the π_1 value using the first line of the stick-breaking transformation, meaning that one can examine the
199 associated π_1 estimate for a given θ_1 value (Figure 1). While it is theoretically possible for all bones in
200 an assemblage to be from immature animals, ethnographically-recorded culling strategies and taphonomic
201 factors suggest that it is more likely that immature animals are a smaller component of the assemblage. A
202 ‘standard’ reference prior distribution, $\theta_1 \sim \text{Normal}(0, 2.5)$, only makes it slightly more likely that π_1 is below
203 50% than not (61%: Figure 1A), which does not fit our intuition about the proportion of immature animals.
204 Using a slightly off-center Normal distribution with a smaller standard deviation, $\theta_1 \sim \text{Normal}(-0.5, 1.5)$,

205 provides a distribution of potential π_1 values that better fit the expectation for the proportion of immature
 206 animals in an assemblage (79%: Figure 1B).

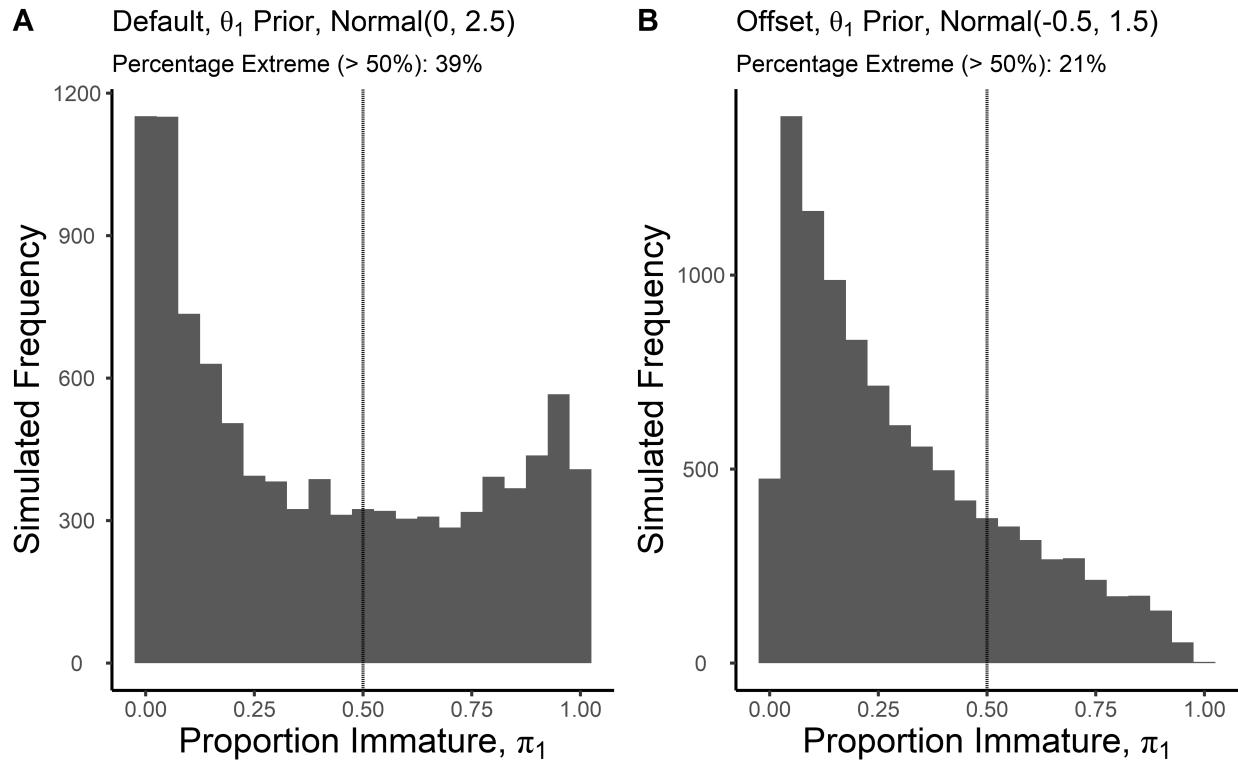


Figure 1: Comparison of π_1 Distributions from Different θ_1 Distributions

207 Within the stick-breaking transformation, θ_2 relates to the relative proportions of π_2 and π_3 after π_1
 208 has been estimated, which is effectively the adult sex ratio. Just as we could examine the expected π_1
 209 estimates from a distribution of θ_1 values, we can thus use expected adult sex ratios ($\frac{\pi_2}{\pi_2 + \pi_3}$) estimates from
 210 a particular prior distribution for θ_2 (Figure 2). In this case, we have no reason to, *a priori*, believe that
 211 the adult sex ratio skews towards males or females, though we may believe that extremely imbalanced sex
 212 ratios, however defined, are not likely. Again, a standard reference prior definition, $\theta_2 \sim \text{Normal}(0, 2.5)$,
 213 produces estimates that may not fit our expectation—in this case, extreme sex ratios (< 10% or > 90%
 214 female) are only about half as likely as sex ratios within those bounds (38%: Figure 2A). Decreasing the
 215 standard deviation, $\theta_2 \sim \text{Normal}(0, 1.5)$, makes these extreme sex ratios about half as likely (14%: Figure
 216 2B).

217 *Prior Distribution Definitions for θ Hyper-parameters:*

$$\begin{aligned} \theta_1 &\sim \text{Normal}(-0.5, 1.5) \\ \theta_2 &\sim \text{Normal}(0.0, 1.5) \end{aligned} \tag{5}$$

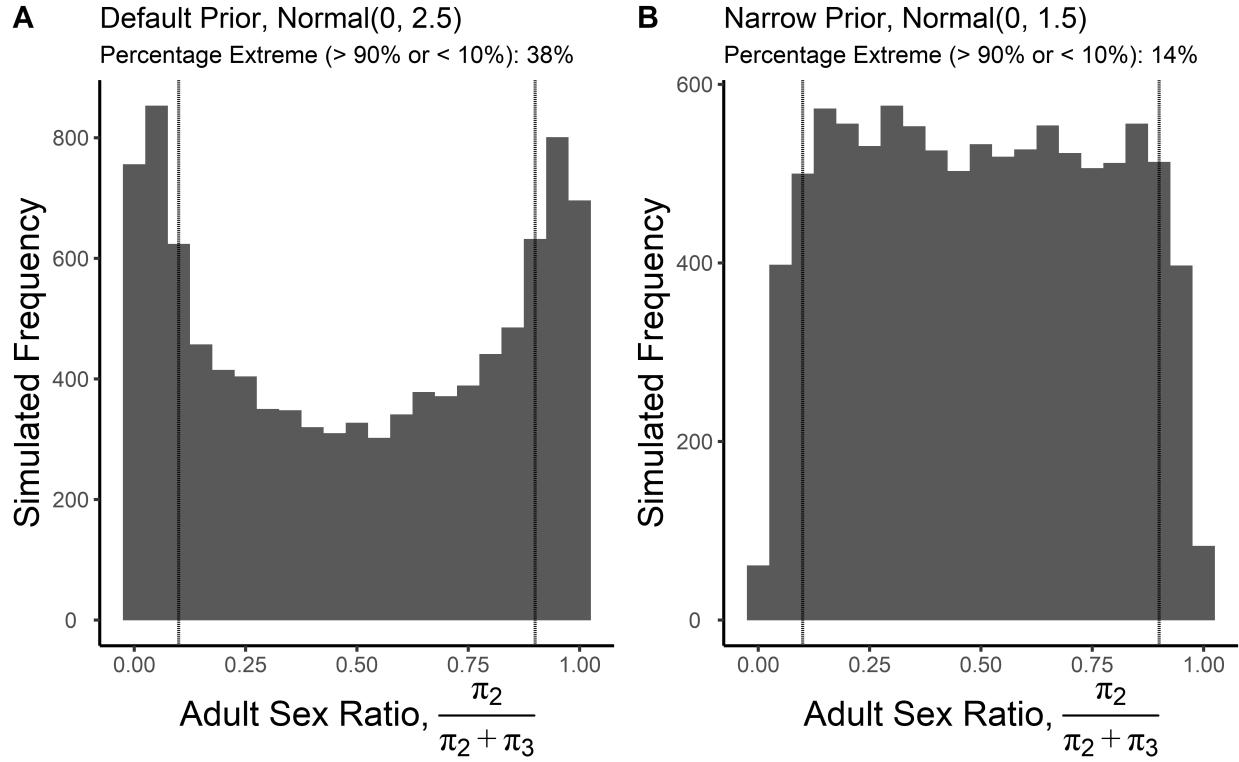


Figure 2: Comparison of Adult Sex Ratios $\frac{\pi_2}{\pi_2 + \pi_3}$ Distributions from Different θ_2 Distributions

2.3.2 Average Body Size Priors

While the average body sizes of the different components (μ_1, μ_2, μ_3) are not intrinsically linked in the same way that π values are, the model still requires some structure to aid interpretability. Bayesian mixture models that are fit using Markov Chain Monte Carlo (MCMC) methods, like the model in this paper, can suffer from an issue called “label switching” if μ values are not related to one another (Jasra, Holmes, and Stephens 2005). This describes situations where some iterations of the model switch what specimens it effectively measures—that is, an instance where μ_1 centers on female animals while μ_2 centers on immature animals. To avoid label switching, the average body sizes are strictly ordered, meaning that $\mu_1 < \mu_2 < \mu_3$ must be maintained. Note that this only affects average values, individual immature specimens can still be larger than female specimens or male specimens and individual female specimens can be larger than male specimens. This is done by only estimating μ_2 (average LSI_e value for females) directly while estimating the average LSI_e value for immature and male animals with offsets (δ_1, δ_2). The δ values must be positive to maintain the ordering of the μ values, so each δ is modeled in a log-transformed space.

$$\begin{aligned}\mu_1 &= \mu_2 - \delta_1 \\ \mu_3 &= \mu_2 + \delta_2\end{aligned}\tag{6}$$

232 Conceptually, this expression of animal body size defines female animals as the generic “body size” that
 233 is subject to various selective pressures, with the offset for male animals reflecting sex-specific pressures on
 234 males. This interpretation of body size broadly fits the general pressures affecting adult body size in females
 235 and males across many ungulate taxa, including domestic herd animals (Tchernov and Horwitz 1991; Pérez-
 236 Barbería, Gordon, and Pagel 2002). The body size offset between immature animals and adult-sized females
 237 (δ_1) is admittedly an *ad hoc* definition rather than one under strict biological constraints, as it can be affected
 238 by the age immature animals reach before being killed (Gillis et al. 2014). The computational advantages of
 239 this definition arguably outweigh the awkwardness of the definition, however. Further, evaluation of δ_1 and
 240 δ_2 values across different sites could conceivably highlight variation in the timing of the killing of immature
 241 animals (δ_1) and the degree of adult sexual dimorphism (δ_2); both variables can be related to models of
 242 hunting intensity, animal domestication, and herd management (Gillis et al. 2014; Zeder and Hesse 2000;
 243 Marom and Bar-Oz 2013).

244 Prior distributions for these size-related parameters (μ_2 , δ_1 , δ_2) are based on LSI_e data from 356 known-
 245 age and known-sex Shetland sheep described in Popkin et al. (2012). Castrated individuals were included as
 246 males and animals killed under one year of age were considered “immature” specimens, leaving 48 immature
 247 animals, 164 female animals, and 144 male animals. LSI_e values are calculated using the *Ovis orientalis*
 248 female standard animal (FMC 57951) from Uerpmann and Uerpmann (1994: Table 12). Table 2 shows
 249 the measurements included in the LSI_e simulation analyses. From the 2848 element portions from this
 250 population, 150 immature, female, and male element portions were randomly drawn to create an assemblage
 251 of 450 element portions. A Bayesian multilevel mixture model was fit to these known-identity specimens to
 252 create estimates of the relevant biologically-constrained hyper-parameters (μ_2 , δ_1 , δ_2 , σ_1 , σ_2 , σ_3). The results
 253 of this analysis are then used as the foundation for prior distributions of the relevant hyper-parameters in
 254 the archaeological model where group identities are unknown.

Table 2: Measurements included in the Simulation Analyses

Element	Measurement
Scapula	GLP
Humerus	Bd
Humerus	BT
Radius	Bp
Radius	Bd
Metacarpus	Bp
Metacarpus	Bd
Femur	Bd
Tibia	Bd
Astragalus	Bd
Metatarsus	Bp
Metatarsus	Bd

255 The prior distributions used in this Bayesian multilevel mixture model on the reference population are
256 more straightforward. While the same transformations to create unconstrained parameters are necessary
257 (e.g., modeling average size as μ_2 with offsets for immature and male animals), the definition of these prior
258 distributions can be broadly described as weakly-informative priors (Gelman et al. 2008). These weakly-
259 informative prior distributions are reasonable in this case—and not in the archaeological case—because all
260 the mixture model parameters have direct observations rather than relying on latent state estimations. That
261 is, parameters like the size difference between males and females (δ_2) and the size variability in female animals
262 (σ_2) can be directly observed because the group identities of every specimen are known. With these direct
263 observations, the prior distributions have a more muted influence on the resulting posterior distributions.
264 This does not mean that the prior distributions have no effect, however, which is why objective priors can
265 have undesirable impacts on modeling results (Gabry et al. 2019).

266 Figure 3 shows the posterior distributions of the model hyper-parameters for both average body size (μ_2 ,
267 δ_1, δ_2) and size variability ($\sigma_1, \sigma_2, \sigma_3$) based on the Shetland sheep sample. These distributions are associated
268 with proposed prior distribution definitions for the same hyper-parameters for archaeological applications of
269 the model when identity is unknown. Note that these hyper-parameters describe averaged estimates across

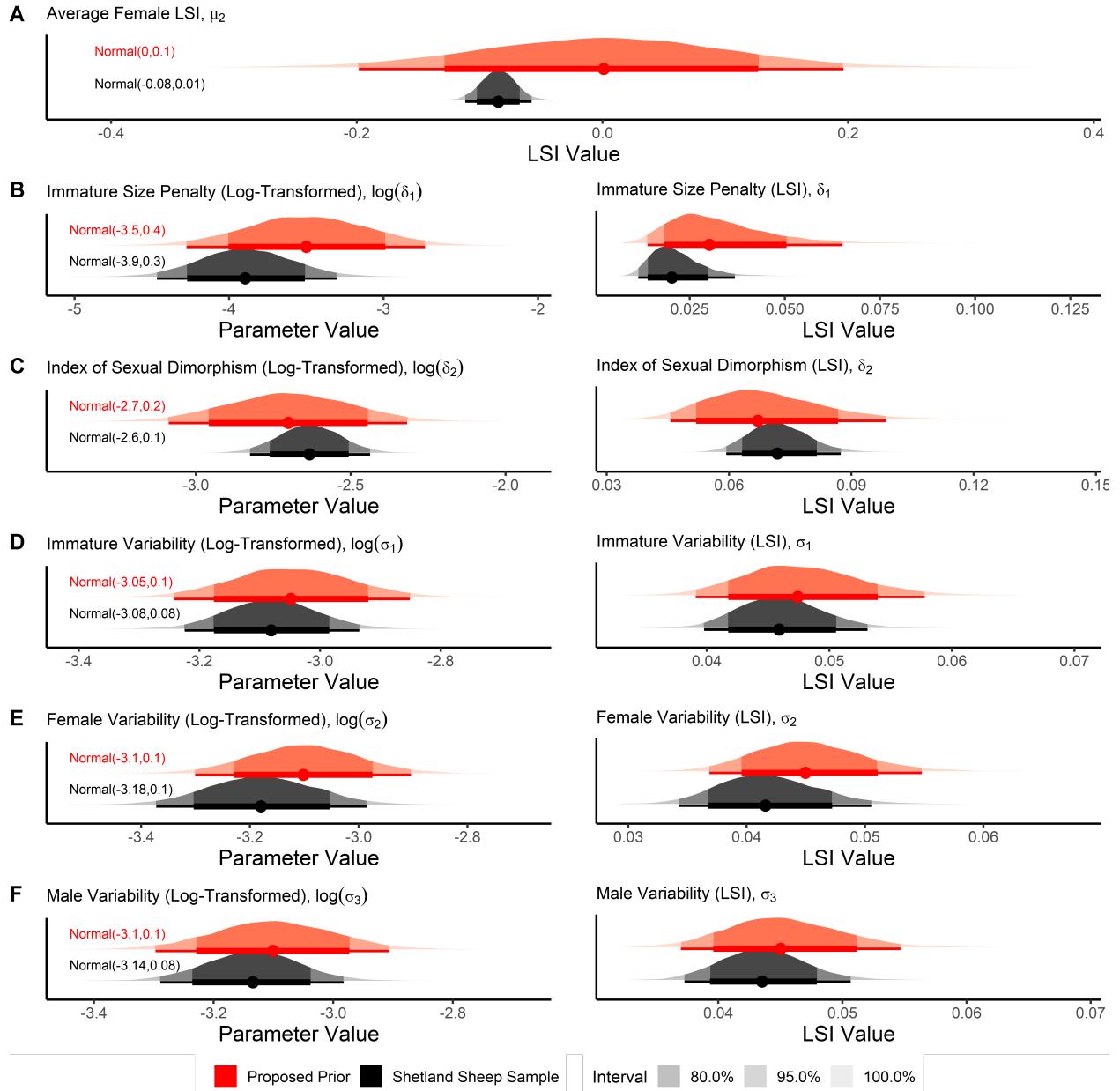


Figure 3: Posterior Distributions of Model Hyper-parameters from Sample of Known-Identity Specimens and Proposed Prior Distributions for Archaeological Mixture Model Applications

270 element portions in the model. In general, the posterior distributions from the sample of known-identity
271 Shetland sheep are narrower than the associated prior distribution proposals. This narrowness is due in
272 part to the large sample size of the sampled assemblage but also to the fact that every element portion is
273 represented by the same individuals, something that is extremely unlikely in archaeological scenarios. Thus,
274 care should be taken before directly translating these results into prior distributions for an archaeological
275 scenario.

276 The average LSI_e value for females (μ_2) is likely to vary across contexts in reaction to different selective
277 pressures, both anthropogenic and ecological (e.g., Davis 1981; Manning et al. 2015; Wright and Viner-
278 Daniels 2015). While the posterior distribution is extremely focused on this specific population, there is no
279 reason to think that this value should be centered at any particular value since that relates to the standard
280 animal used (Meadow 1999; Wolfhagen 2020). Therefore, the prior distribution used in archaeological models
281 for μ_2 uses a larger standard deviation, $\mu_2 \sim \text{Normal}(0, 0.1)$, to encompass likely LSI_e values (Figure 3A).
282 Under this definition, there is a 95% probability that the μ_2 value lies within the range of -0.20 and 0.20 on
283 the LSI_e scale, which translates to roughly 82-122% the size of the standard animal's measurement.

284 For the average size difference between immature and female animals (δ_1), the narrowness of the posterior
285 distribution likely reflects the fact that immature animals in the sample cover a narrow age range. Animals
286 killed under one year of age span only 36 days and the youngest animals are nearly half a year old (178-214
287 days: Popkin et al. 2012). Thus while the posterior results provide a useful starting point for estimating
288 this offset, there is a good potential for larger δ_1 values (i.e., greater size differences between immature and
289 adult-sized female animals) in other contexts that could include animals killed at a younger age (Figure 3B).
290 To capture this possibility, the archaeological model uses a prior distribution with a larger standard deviation
291 and a slightly higher center, $\log \delta_1 \sim \text{Normal}(-3.5, 0.4)$, which results in an average size difference of 0.03 on
292 the LSI_e scale and a 95% probability that the size difference is between 0.01 and 0.07. This translates into
293 expecting the average body size of immature animals in an assemblage being 3% smaller than the average
294 body size of adult-sized female animals, but also plausibly believing that this size difference could range from
295 1-7% smaller.

296 The average size difference between adult males and females (δ_2), also known as the index of sexual
297 dimorphism (Fernández and Monchot 2007), is likely to be under stricter biological control than the other
298 "average body size" parameters in the model. This does not mean that this difference could not vary between
299 contexts, however. Some models of animal domestication argue that initial domestication removed sexual
300 selective pressures on male body size, reducing sexual dimorphism (e.g., Tchernov and Horwitz 1991). In
301 a similar fashion, specialized hunting strategies could also reduce sexual dimorphism by targeting large-
302 bodied males, for example (Zeder 2012; Proaktor, Coulson, and Milner-Gulland 2007; Milner, Nilsen, and

303 Andreassen 2007). Again, the posterior distribution of the extent of sexual dimorphism in the Shetland
 304 sheep population provides a useful starting point to describe a prior distribution for the model (Figure
 305 3C). Increasing the standard deviation of the distribution slightly, $\log \delta_2 \sim \text{Normal}(-2.7, 0.1)$, produces a
 306 distribution centered at 0.07 LSI_e units with a 95% probability that the value is between 0.06-0.08, translating
 307 to the average male being 6-9% larger than the average female relative to a standard measurement. The
 308 smaller standard deviation in the prior distribution of δ_2 than for δ_1 reflects our understanding that the extent
 309 of sexual dimorphism, as a biological phenomenon, is less likely to have extreme values than the average size
 310 difference between immature and female animals, since δ_2 is unaffected by the specific age structure of the
 311 assemblage.

312 *Prior Distribution Definitions for μ and δ Hyper-parameters:*

$$\begin{aligned}\mu_2 &\sim \text{Normal}(0, 0.1) \\ \log \delta_1 &\sim \text{Normal}(-3.50, 0.4) \\ \log \delta_2 &\sim \text{Normal}(-2.70, 0.1)\end{aligned}\tag{7}$$

313 **2.3.3 Size Variability Priors**

314 The LSI_e size variability of animals within a group ($\sigma_1, \sigma_2, \sigma_3$) is a key variable in the Bayesian mixture
 315 model. The values of these standard deviation parameters play a major role in ensuring that the mixture
 316 components reflect biological entities rather than overfitting to specific sample noise. Previous research
 317 into size variability suggests that these σ values should be relatively stable across elements. Coefficients
 318 of variance (CVs) for raw mammal bone measurements from a single sex have been found to be relatively
 319 consistent (Davis 1996; Popkin et al. 2012). When transforming these measurements using a logarithm, this
 320 produces consistent standard deviations of the transformed measurement values (Wolfhagen 2020: Figure
 321 1).

322 While σ parameters from different groups within the model are not directly related to each other, the
 323 values still need some transformations to be modeled consistently by the multilevel model. These values must
 324 be positive, which conflicts with the multilevel model's need for unconstrained variables. To achieve this, the
 325 model uses the same log-transformation technique used for size offsets to create an unconstrained parameter,
 326 $\log \sigma$, that is then transformed to actual σ values after estimating variation across element portions.

327 As in the average body size parameters, prior distributions for the size variability model parameters
 328 are developed from the Bayesian model of known-identity Shetland sheep measurements. The resulting σ
 329 hyper-parameters provide a baseline for establishing hyper-parameter prior distributions in archaeological
 330 cases. Figure 3D-F shows the posterior distributions of these σ hyper-parameters in both the log-transformed

331 values and associated LSI_e values. Average size variability within an element portion for immature animals
 332 (σ_1) is higher, on average, than for females (σ_2) and males (σ_3). The immature category includes both male
 333 and female animals, so larger size variability makes sense; again, it is possible that σ_1 is relatively low in
 334 this population relative to other contexts given the narrow age range of immature animals in the Shetland
 335 sheep population. Unlike the average body size parameters, there are not compelling reasons to believe
 336 that size variability parameters for females and males (σ_2 and σ_3) should vary widely in different contexts
 337 given the consistency of coefficients of variation in mammals broadly (Davis 1996). Thus, the results of
 338 this analysis are used for the prior distributions of $\log \sigma_2$ and $\log \sigma_3$, while the prior distribution of $\log \sigma_1$
 339 is given an increased standard deviation and slightly increased average value. Overall, however, these prior
 340 distributions suggest that the average size variability within an element portion is between 0.04-0.06 for
 341 females and males and is between 0.04-0.05 for immature animals. Note that even though σ_2 and σ_3 have
 342 the same prior distributions, these values can still vary from each other in different contexts.

343 *Prior Distribution Definitions for σ Hyper-parameters:*

$$\begin{aligned}
 \log \sigma_1 &\sim \text{Normal}(-3.05, 0.1) \\
 \log \sigma_2 &\sim \text{Normal}(-3.10, 0.1) \\
 \log \sigma_3 &\sim \text{Normal}(-3.10, 0.1)
 \end{aligned} \tag{8}$$

344 **2.4 Multilevel Structure for Element Portions**

345 The previous section described prior distributions that describe the *average* value for different mixture model
 346 parameters across all element portions. To create parameter estimates that are specific to different element
 347 portions, it is necessary to estimate the *variation* around these average values that different parameters
 348 can have among different element portions. The model uses a Multivariate Normal definition of the model
 349 parameters to allow for correlations between different parameters; effectively, the possibility that multiple
 350 model parameters will covary from element portion to element portion. To do this, each hyper-parameter
 351 has an associated σ_{element} parameter that describes inter-element variation in parameter values. The model
 352 uses a non-centered parameterization, wherein the Multivariate Normal distribution is centered at zero to
 353 calculate offsets, ν_{element} , that are added to the average hyper-parameters to calculate model parameters for
 354 each element portion. This definition provides computational stability and makes it more straightforward
 355 to incorporate other levels of multilevel structure (see Section 2.5).

356 Equations for Defining Inter-Element Variation (Multilevel Modeling):

$$\nu_{\text{element}} \sim \text{MultivariateNormal} \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma_{\text{element}} \right)$$

$$\Sigma_{\text{element}} = \begin{pmatrix} \sigma_{\text{element}}[1] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\text{element}}[8] \end{pmatrix} \rho_{\text{element}} \begin{pmatrix} \sigma_{\text{element}}[1] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\text{element}}[8] \end{pmatrix}$$

$$\rho_{\text{element}} = LKJcorr(2)$$

$$\theta_1[\text{element}] = \theta_1 + \nu_{\text{element}}[1] \quad (9)$$

$$\theta_2[\text{element}] = \theta_2 + \nu_{\text{element}}[2]$$

$$\mu_2[\text{element}] = \mu_2 + \nu_{\text{element}}[3]$$

$$\log \delta_1[\text{element}] = \log \delta_1 + \nu_{\text{element}}[4]$$

$$\log \delta_2[\text{element}] = \log \delta_2 + \nu_{\text{element}}[5]$$

$$\log \sigma_1[\text{element}] = \log \sigma_1 + \nu_{\text{element}}[6]$$

$$\log \sigma_2[\text{element}] = \log \sigma_2 + \nu_{\text{element}}[7]$$

$$\log \sigma_3[\text{element}] = \log \sigma_3 + \nu_{\text{element}}[8]$$

357 Prior distributions for the σ_{element} values are weakly-informative priors based on the scale of the parameter
 358 and the expectation for variation in the parameter values among element portions. For example, there is
 359 likely more variation in θ parameters—that govern the relative composition of immature, female, and male
 360 animals—among element portions than variation in σ parameters that govern size variability within each
 361 group. The scale of different parameters also affects the expected spread of values; for instance, μ_2 values
 362 are on the direct LSI_e scale while δ and σ parameters are on the log-scale of LSI_e differences. The prior
 363 distributions for σ_{element} parameters associated with μ , δ , and σ parameters are based on results of the
 364 known-specimen model, erring on the side of more variability for most parameters. Increased variability in
 365 the value of $\log \delta_1$ allows for greater size variation across elements, which makes intuitive sense because δ_1 is
 366 affected by age composition and biology rather than strictly biology.

367 *Prior Distribution Definitions for σ_{element} Parameters (Inter-Element Variation):*

$$\begin{aligned}\sigma_{\text{element}}[1, 2] &\sim \text{Half-Normal}(0, 1) \\ \sigma_{\text{element}}[3] &\sim \text{Half-Normal}(0, 0.1) \\ \sigma_{\text{element}}[4] &\sim \text{Half-Normal}(0, 0.5) \\ \sigma_{\text{element}}[5] &\sim \text{Half-Normal}(0, 0.25) \\ \sigma_{\text{element}}[6, 7, 8] &\sim \text{Half-Normal}(0, 0.25)\end{aligned}\tag{10}$$

368 **2.4.1 Extending the Multilevel Analysis to Multiple Sites**

369 The multilevel structure used to allow variation in parameter estimates across element portions can also be
370 expanded to create multisite models that can directly compare sex-specific biometric estimates alongside the
371 age/sex composition of different assemblages. Such comparisons can highlight variation in herd management
372 strategies or diachronic body size trends related to population turnover (e.g., Arbuckle and Atici 2013;
373 Arbuckle et al. 2016). To do this, an additional multilevel structure can be applied to the same mixture
374 model parameters, using σ_{site} rather than σ_{element} parameters. However, an additional set of multilevel
375 structure parameters, $\sigma_{\text{interaction}}$, are also necessary to ensure that elemental variation is different at different
376 sites (e.g., the difference between μ_2 for the distal humerus and μ_2 for the distal radius is not necessarily the
377 same at different sites). Again, weakly-informative priors are appropriate for both sets of parameters. Each
378 additional term is included in the sum to create specific mixture model parameter values.

379 *Example of Parameter Definition for Inter-Site and Inter-Element Variation:*

$$\theta_1[\text{Site, Element}] = \theta_1 + \nu_{\text{site}}[\text{Site}] + \nu_{\text{element}}[\text{Element}] + \nu_{\text{interaction}}[\text{Site, Element}]$$

380 The inclusion of multiple sites changes the definition of the ‘grand mean’ variable (θ_1 in the example
381 equation) from a site-level estimate to an overall mean across the sites and elements. These parameter esti-
382 mates thus describe the average composition of the entire set of assemblages. The details of the assemblages
383 included in the analyses would affect how useful these estimates are for interpretation. Assemblage-specific
384 estimates can be calculated for each model parameter by adding the relevant ν_{site} estimate to the ‘grand
385 mean’ parameter, which would again act to describe the average composition of the assemblage regardless
386 of its elemental composition.

387 **2.5 Using Model Results to Estimate Composition and Sex-Specific Fusion Rates**

388 The results of the Bayesian multilevel mixture model include specimen-specific membership probabilities
389 (π_{specimen}) based on the mixture model parameters. While these membership probabilities can be used to
390 calculate “critical size limits” where the largest membership probability shifts from one group to another (e.g.,
391 Monchot and Léchelle 2002), they can also be used to simulate assemblages of known-group specimens to
392 examine age/sex-stratified estimates of body part representation and sex-stratified fusion rates. Membership
393 probabilities (π_{specimen}) are used to simulate the specimen’s identity by sampling from the probabilities using
394 a multinomial distribution; in each posterior sample, a single simulated assemblage is created, resulting in a
395 distribution of simulated assemblages with known age/sex assignments (Crema 2011). The characteristics of
396 these assemblages can then be used to summarize the overall assemblage or identify differences in composition
397 based on element types, fusion states, sub-assemblage features, or other pertinent factors that a researcher
398 is interested in examining in relation to the composition of the assemblage.

399 There are important extensions that need to be considered when analyzing composition based on mixture
400 modeling results. First, the Bayesian multilevel mixture model estimates the composition and biometry of a
401 taxon’s *measurement* assemblage. In most cases, however, our interest as zooarchaeologists is the composition
402 of the taxon’s *entire* (or *modeled*) assemblage. Generally, this is done by assuming that “measurability”—the
403 probability that a specimen has preserved body parts that allow for biometric measurements—is random; that
404 is, that whatever factors impacting whether a specimen is preserved well enough to have intact measurements
405 is unrelated to the age and sex category of the specimen. This is informally applied when analysts use the
406 results of a biometric analysis to describe an assemblage. The modeling results here can be used to formalize
407 this relationship by stating that we believe that the element portion-specific model estimates, particularly
408 the mixture proportions π , equally describe the measured and the non-measured specimens from the element
409 portion. That is, while we can estimate the membership probabilities for a measured assemblage using the
410 mixture model, our best estimate for the membership probabilities of a non-measured specimen is the element
411 portion-specific mixture proportions π . This provides a way to include all specimens from a modeled element
412 into estimates of composition and fusion rates, with the same caveats that fusion status can preclude the
413 probability that a specimen could be from an immature animal.

414 The second extension focuses on the multilevel structure of the model, specifically the assumption that
415 the overall mixture model hyper-parameters for *modeled* element portions are equally valid for *unmodeled*
416 element portions. Because the model produces estimates of the “average” hyper-parameters and the vari-
417 ability of these parameters across element portions (σ_{element}), one could use these data to estimate the
418 element portion-specific mixture model parameter values of a completely unmodeled element portion (McEl-

419 reath 2020; Gelman, Carlin, et al. 2020). These parameter estimates, then, could be used to estimate
420 π_{specimen} membership probabilities for the unmodeled (and unmeasured) specimens to serve as the baseline
421 for simulated assemblages.

422 These extensions may seem like an extreme departure from the mixture model results, but they are a
423 simply a formalization of the implicit assumptions analysts adopt when using the results of a biometric
424 analysis to describe whole assemblages. If anything, these formalizations would be expected to make esti-
425 mates of age/sex composition for an assemblage *less extreme* than using only the results of the measured
426 assemblage to describe the entire assemblage. This is because of the uncertainty in the parameter values
427 used to create π_{specimen} values for unmeasured and unmodeled specimens, which will likely be much greater
428 than the uncertainty in the π_{specimen} values for measured specimens. Thus, extending the mixture model
429 results to unmeasured and unmodeled specimens creates a more faithful estimate of the overall assemblage,
430 avoiding overfitting to the smaller measurement assemblage.

431 **2.6 Computational Details of the Bayesian Model**

432 The Bayesian multilevel mixture model is written in Stan, version 2.21.0 (Team 2022). All analyses in
433 this paper use R version 4.1.2 (2021-11-01), in Rstudio 2022.2.3.492 (Prairie Trillium) (R Core Team 2022;
434 RStudio Team 2022). The analytical scripts use the following packages: “data.table” version 1.14.2, “parallel”
435 version 4.1.2, and “doParallel” version 1.0.17 for data aggregation and multi-core processing (Dowle and
436 Srinivasan 2021; Corporation and Weston 2022), “cmdstanr” version 0.4.0 and “rstan” version 2.21.3 for
437 running the Bayesian models in the R environment and summarizing the posterior distributions of the
438 models (Gabry and Cešnovar 2022; Stan Development Team 2021), “zoolog” version 0.4.1 for standard
439 animal measurements (Pozo et al. 2021), and “Cairo” version 1.5.15, “ggplot2” version 3.3.5, “ggdist”
440 version 3.1.1, “ggpubr” version 0.4.0, “rnaturalearth” version 0.1.0, “rnaturalearthdata” version 0.1.0, and
441 “sf” version 1.0.7 to create visualizations (Urbanek and Horner 2022; Wickham 2016; Kay 2022; Kassambara
442 2020; South 2017a, 2017b; Pebesma 2018).

443 The model Stan code and analytical R code necessary to replicate and apply the analyses in this paper
444 are freely available in a GitHub page (<https://github.com/wolfhagenj/ZooarchMixMod>) and Open Science
445 Framework page (<https://osf.io/4h9w6/>). The project includes a copy of the Shetland sheep data file from
446 the supplemental files published in Popkin et al. (2012) and archaeological datasets for the case studies down-
447 loaded from OpenContext (Dataset 2006, 2013, 2014; Buitenhuis 2013). The analytical code includes two
448 script files—a script for replication and one for application. The R markdown file (“ZooarchMixMod.Rmd”)
449 file replicates the entire analytical workflow of the paper, with a specific seed set to ensure exact replicability
450 of the submitted manuscript. Another set of scripts to standardize the analytical workflow for faunal datasets

451 structured like the OpenContext faunal datasets used in these case studies, see the GitHub for more details.
452 All scripts (R and Stan) are released under the MIT license and figures are released as CC-BY to encourage
453 reuse and reproducibility (Marwick 2017; Marwick and Pilaar Birch 2018).

454 3. TESTING THE BAYESIAN MULTILEVEL MIXTURE MODEL

455 Two sets of tests are used to evaluate different aspects of the Bayesian multilevel mixture model. First,
456 the accuracy of the model's ability to reconstruct the age and sex composition of assemblages is tested by
457 using simulated faunal assemblages of known age and sex from the Shetland sheep population. This test
458 evaluates both the single-assemblage model and the multi-assemblage model. Second, two archaeological
459 case studies showcase the applicability of the model to archaeological data and the added insights gained
460 from adopting Bayesian multilevel mixture models. The simulated assemblage case study and the single
461 assemblage archaeological case study use sheep (*Ovis aries*) measurements, with standard measurements
462 coming from a female wild sheep (*Ovis orientalis* FMC 57951: Uerpmann and Uerpmann 1994: Table 12).
463 The multiple assemblage case study uses cattle (*Bos taurus*) measurements, with standard measurements
464 coming from a wild female aurochs (*Bos primigenius* "Ullerslev": Degerbøl 1970). Two measurements of the
465 standard cow (Scapula GLP: 89 mm; and Calcaneus GB: 46 mm) were not included in the 'zoolog' output
466 and were included manually, drawn from the referenced source.

467 3.1 Simulated Assemblages

468 A series of simulated assemblages of known-age and sex composition are created from the Shetland sheep
469 population by randomly drawing element portions (and all associated measurements) from the total assem-
470 blage without replacement. The first test, using a single-assemblage model, uses 150 element portions from
471 the Shetland sheep population where every element portion has an equal probability of being selected. There
472 is no guarantee, however, that the element portions have equal representation or even that all element por-
473 tions are present in the simulated assemblage, which better approximates archaeological assemblages. The
474 result of this first simulation produces an assemblage of 231 measurements from individual animals. The
475 second test creates 15 simulated assemblages using the same procedure that are analyzed in a single multi-
476 assemblage model. To further test the model's flexibility, these assemblages vary in sample size and some
477 are manipulated to vary in average body size and expected composition from the original Shetland sheep
478 population. Demographic observations for phalanx fusion rates and pelvis sex ratios were also simulated
479 from the Shetland sheep population using the same underlying probabilities as the measurement assem-
480 blages. Table 3 describes the sample sizes of the measurement assemblages, including any manipulations to

Table 3: Group Composition of the Simulated Measurement Assemblages (Element Portions)

Assemblage	Demographics	Size	Immature	Female	Male	Total
Single Assemblage	13% Immature, 46% Female, 40% Male	1.00	23	80	47	150
Site 01	13% Immature, 46% Female, 40% Male	1.00	2	11	17	30
Site 02	13% Immature, 46% Female, 40% Male	1.00	3	3	4	10
Site 03	13% Immature, 46% Female, 40% Male	1.20	4	13	13	30
Site 04	13% Immature, 46% Female, 40% Male	0.80	3	11	16	30
Site 05	13% Immature, 46% Female, 40% Male	1.20*	4	13	13	30
Site 06	20% Immature, 70% Female, 10% Male	1.00	5	24	1	30
Site 07	20% Immature, 70% Female, 10% Male	1.00	0	8	2	10
Site 08	20% Immature, 70% Female, 10% Male	1.20	4	23	3	30
Site 09	20% Immature, 70% Female, 10% Male	0.80	4	25	1	30
Site 10	20% Immature, 70% Female, 10% Male	1.20*	11	17	2	30
Site 11	5% Immature, 35% Female, 60% Male	1.00	2	10	18	30
Site 12	5% Immature, 35% Female, 60% Male	1.00	0	3	7	10
Site 13	5% Immature, 35% Female, 60% Male	1.20	1	17	12	30
Site 14	5% Immature, 35% Female, 60% Male	0.80	3	11	16	30
Site 15	5% Immature, 35% Female, 60% Male	1.20*	2	12	16	30

Note:

Demographics in the Single Assemblage and Sites 01-05 reflect original Shetland sheep composition

* Size increased for males only

481 the measurement values. The specific elemental composition and measurements of the assemblages, along
 482 with the simulated demographic observations, used in both simulations can be recovered from the replication
 483 script with the recorded random seed (see also Supplemental Tables 1-3); using another random seed would
 484 provide a conceptual replication of new assemblages drawn from the same underlying populations.

485 For each model test, accuracy is evaluated in two ways. First, the posterior distributions of the mixture
 486 model hyper-parameters are compared to the known values for the population (“parametric accuracy”).
 487 The model hyper-parameters are first transformed into estimates of composition (π), body size (μ) and size
 488 variability (σ) rather than the unconstrained variables that were modeled directly. These known values are
 489 calculated from the entire Shetland sheep population from which the assemblages were sampled (Popkin
 490 et al. 2012), including any manipulations of composition or size as described in Table 3. Second, the
 491 mixture model results are used to estimate the age and sex composition of the relevant measured and
 492 modeled assemblages—estimates of the number of immature, female, and male specimens for each element
 493 portion—that are compared to the actual composition of the assemblages (“compositional accuracy”). For
 494 the multi-assemblage model, compositional accuracy is evaluated both overall (all the assemblages combined)
 495 and for each assemblage individually. As described in Section 2.5, modeled assemblages are the measured and
 496 unmeasured specimens from the relevant element portions that are modeled with the mixture model. The
 497 measurement assemblage is considered a sample from this modeled assemblage, under the assumption that

498 measurability—sufficient preservation of anatomical structures to be measured—is unrelated to the age and
 499 sex category of the specimen. Thus, just as the parametric accuracy is compared against the “population”
 500 parameters from which the measured assemblages were sampled, so to the compositional accuracy can be
 501 compared against the “modeled” assemblage that includes non-measured specimens sampled from the same
 502 underlying population.

503 Modeled assemblages were created for the single-assemblage simulation and the multi-assemblage sim-
 504 ulation by assuming that measured specimens represent 20% of the overall assemblage and sampling more
 505 specimens from the Shetland sheep population to create the remaining 80% of the assemblage. For example,
 506 in the single-assemblage simulation with 150 measured specimens, this means sampling 600 more specimens
 507 from the Shetland sheep population to create a total modeled assemblage of 750 specimens. Specimens could
 508 not be repeatedly sampled, though multiple specimens could be from the same individual. As described in
 509 Section 2.5, unmeasured specimens use the relevant π parameters for the element portion. For the multi-
 510 site simulation, this potentially includes element portions where there are no relevant measurements in the
 511 specific assemblage.

512 Because the “grand mean” parameters in the multisite simulation no longer represent the same thing as
 513 in the single assemblage model (see Section 2.4.1), the prior distributions must also be changed to reflect
 514 different expectations. Again, the goal of these prior distribution definitions is to prevent extreme overfitting
 515 so that parameter estimates are biologically feasible (Gelman et al. 2008). In general, the centers of the
 516 distributions stayed the same but the uncertainty was increased, reflecting the fact that there’s less certainty
 517 about what may be a biologically feasible value to describe multiple populations, especially if there is size
 518 variation expected between the assemblages.

519 *Prior Distribution Definitions for the Multisite Simulation Model Hyper-Parameters*

$$\theta_1 \sim \text{Normal}(-0.5, 1.5)$$

$$\theta_2 \sim \text{Normal}(0.0, 1.5)$$

$$\mu_2 \sim \text{Normal}(0.0, 0.2)$$

$$\log \delta_1 \sim \text{Normal}(-3.5, 0.5) \tag{11}$$

$$\log \delta_2 \sim \text{Normal}(-2.7, 0.5)$$

$$\log \sigma_1 \sim \text{Normal}(-3.05, 0.25)$$

$$\log \sigma_2 \sim \text{Normal}(-3.10, 0.2)$$

$$\log \sigma_3 \sim \text{Normal}(-3.10, 0.2)$$

520 **3.2 Archaeological Case Studies**

521 The Bayesian multilevel mixture model is applied to two archaeological case studies to showcase the utility
522 of the model for both interpreting a single assemblage and examining multiple assemblages. In both case
523 studies, the sheep and cattle measurements have been previously published on OpenContext and the general
524 zooarchaeological summaries of the assemblages have been published, as well (Dataset 2006, 2013, 2014;
525 Buitenhuis 2013, 2008; Carruthers 2005; Gerritsen and Özböl 2019; Gourichon and Helmer 2008). Again,
526 LSI_e values are calculated using the same standard animal as the simulation analysis for the single assem-
527 blage analysis, the *Ovis orientalis* female standard animal (FMC 57951) from Uerpmann and Uerpmann
528 (1994: Table 12) and the *Bos primigenius* female standard animal (Ullerslev: Degerbøl 1970; Grigson 1989),
529 operationalized through ‘zoolog’ functions (Pozo et al. 2021). Alongside metric data, the OpenContext
530 faunal tables provide demographic data that can be used to observe relevant estimates of the age and sex
531 composition of the assemblages. The goal of applying the mixture model to these assemblages, then, is to
532 use the metric data to improve these estimates of the age and sex composition of the assemblage, biometric
533 estimates, and sex-specific fusion rates.

534 **3.2.1 Single Assemblage: Biometric Analysis of Sheep from 7th Millennium BCE Central
535 Anatolia (Pinarbaşı B)**

536 The site of Pinarbaşı, located in the Konya Plain of central Turkey, consists of a series of rock shelter
537 and open-air sites at the foothills of the Karadag volcanic region and Lake Hotamis and associated wetlands
538 (Baird et al. 2013; Kabukcu 2017). This case study examines the Pinarbaşı B late Neolithic occupation,
539 which is dated to the second half of the 7th millennium BCE and includes a large number of domesticated
540 sheep and goat remains (?). Carruthers (2005) analyzed fauna from the 1994-1995 excavations by Trevor
541 Watkins (Watkins 1996), interpreting the presence of fetal sheep remains and other juvenile remains in the
542 assemblage as evidence for herders penning sheep on-site. The Neolithic assemblage was thus described as
543 the result of seasonal occupation by sheep and goat herders during the lambing season and the fall, with
544 culling in the spring possibly focused on young males (Carruthers 2005). This analysis makes several claims
545 that can be evaluated with the Bayesian multilevel mixture model: the dominance of immature remains, a
546 female-dominated adult sex ratio, and sex-specific differences in fusion rates for later-fusing elements.

547 The Bayesian multilevel mixture model for the late Neolithic Pinarbaşı B assemblage uses 44 sheep
548 measurements from 44 specimens (see Table 4). In addition to these measurements, the observed proportion
549 of immature animals from unfused first and second phalanges is 59 / 62 (95%), including specimens identified
550 to sheep and to sheep/goat. There are 0 observed sheep (or sheep/goat) pelvis bones with sex identifications;
551 this is entered into the model by having an observed adult sex ratio for the assemblage of 0 / 0 (females /

552 females + males). All data come from the Pinarbaşı faunal assemblage uploaded to OpenContext, focusing
 553 only on specimens in the Site B Neolithic contexts (Dataset 2006). The Pinarbaşı B sheep model uses the
 554 same prior distribution definitions for the model hyper-parameters as the single assemblage simulation since
 555 both models, even though the sheep body sizes likely differ between the two populations, showcasing the
 556 flexibility of the standard prior distribution definitions.

Table 4: Elemental Composition of the Pinarbaşı B Assemblage

Element Portion	Measurement	N
Astragalus	Bd	10
Calcaneus	GB	9
Humerus	Bd	1
Metacarpal (Distal)	Bd	2
Metatarsal (Distal)	Bd	4
Metatarsal (Proximal)	Bp	2
First Phalanx	Bp	9
Radius (Distal)	Bd	2
Tibia (Distal)	Bd	3
Tibia (Proximal)	Bp	2

557 **3.2.2 Multiple Assemblages: Biometric Analysis of Cattle from 7th-6th Millennium BCE North-
 558 west Anatolia (Barçın Höyük, İlpınar Höyük, Menteşe Höyük)**

559 Understanding the development of Neolithic communities in northwestern Anatolia has long been of in-
 560 terest for researchers interested in studying the spread of agricultural lifeways from southwest Asia into
 561 Europe (e.g., Çakırlar 2013; Karul 2019; Özdoğan 2011, 2019). Agricultural communities first appear in the
 562 Marmara region in the mid-seventh millennium BCE in sites like Barçın Höyük (Gerritsen and Özböl 2019;
 563 Karul 2019). The domestic animal economies of these Late Neolithic and Early Chalcolithic communities
 564 appears to be focused on cattle and caprine (sheep and goat) herding, rather than pig husbandry (Buiten-
 565 huis 2008; Çakırlar 2013; Gourichon and Helmer 2008). Milk residues on pottery recovered from these sites
 566 suggest that these communities regularly consumed milk, potentially orienting herd management strategies
 567 of sheep, goats, and particularly cattle to specialize in milk production (Evershed et al. 2008; Thissen et al.
 568 2010).

569 Four archaeological components from three sites are used in this case study, located near Lake Iznik
 570 and on the Yenisehir Plain in the Bursa province of Turkey (Figure 4). The Neolithic layers from Barçın
 571 Höyük (Phase VI) is the earliest of these assemblages, with occupation roughly from 6500-6000 cal BCE;
 572 excavations revealed a subsistence economy focused on cereal agriculture and the herding of cattle, sheep,
 573 and goat (Dataset 2013; Gerritsen and Özböl 2019). Menteşe Höyük is located approximately five km west
 574 of Barçın Höyük on the Yenisehir Plain; the three Neolithic layers at the site date to 5800-5600 cal BCE

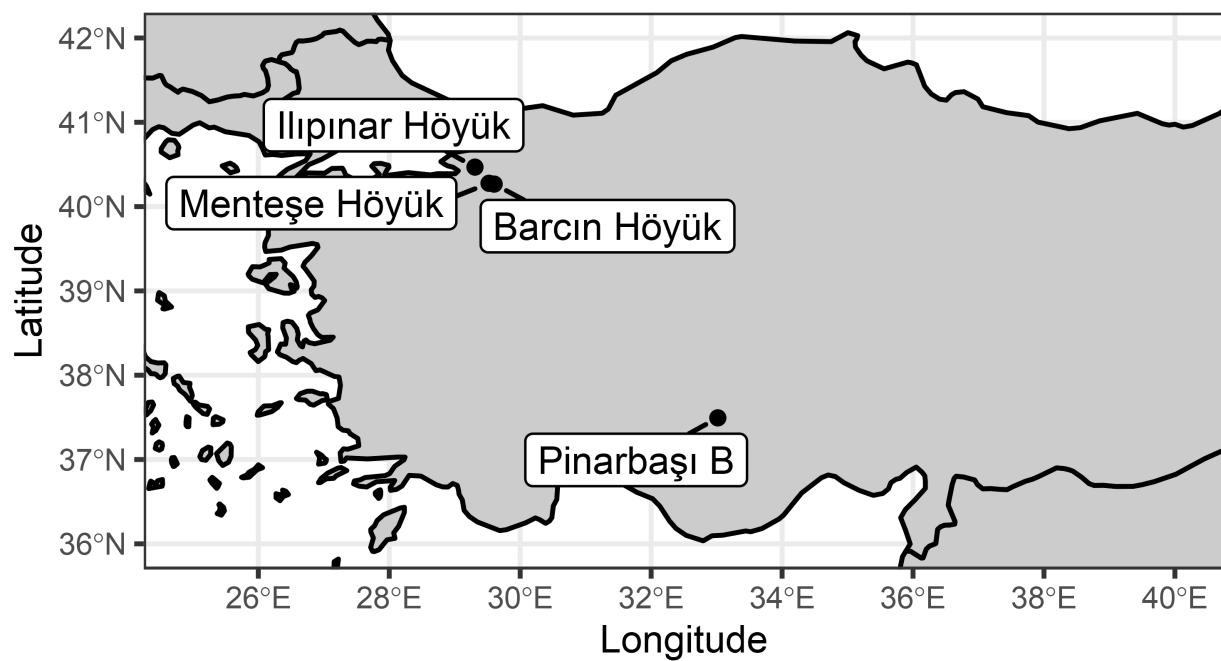


Figure 4: Map of archaeological sites included in this analysis

575 (Dataset 2014; Roodenberg et al. 2003). Previous faunal analysis of the Neolithic assemblage identified
576 animal economies that shifted from predominantly cattle to sheep herding over the course of the occupation
577 (Gourichon and Helmer 2008). İlipinar Höyük is located near Lake İznik, separated from the Yenisehir Plain
578 by a mountain ridge (Roodenberg 2012a). The Neolithic/Early Chalcolithic occupation of the site spanned
579 6200-5400 cal BCE (Buitenhuis 2013); I divided the assemblage into two sub-assemblages (Neolithic İlipinar
580 = Phases X-VII, 6000-5700 cal BCE; Chalcolithic İlipinar = Phases VI-V, 5600-5400 cal BCE), marked by
581 the introduction of mudbrick architecture and expanded storage (Roodenberg 2012a, 2012b). Sheep and
582 goat are common in the earlier assemblages of the site, with cattle becoming predominate in later phases of
583 the site (Buitenhuis 2008; Roodenberg 2012a). Notably for this biometric analysis, Buitenhuis (2008) notes
584 that cattle body sizes are stable throughout the Neolithic layers.

585 The northwest Anatolian cattle bone assemblages consists of 614 measured specimens spread unevenly
586 across the four components (Barçın Höyük N = 67, Menteşe Höyük N = 45, Neolithic İlipinar N = 249, Chal-
587 colithic İlipinar N = 253). All measured *Bos* remains were included in the analysis, rather than separating out
588 those identified as aurochs (*Bos primigenius*, N = 3) or identified only to *Bos* spp. (N = 134) in the İlipinar
589 Höyük dataset; all specimens were only labeled as “Bos” in the Menteşe Höyük dataset. Table 5 shows the
590 measurement composition of the four northwest Anatolian assemblages. Demographic observations of the
591 proportion of immature animals and the adult sex ratio for each assemblage describe these assemblage-level
592 parameters. For the four northwest Anatolian assemblages, estimates of the assemblage-level proportion of
593 immature specimens based on the fusion rates of proximal and middle phalanges for cattle specimens are 28
594 / 87 (32%) for Barçın Höyük, 28 / 184 (15%) for Neolithic İlipinar, 8 / 25 (32%) for Menteşe Höyük, and 9 /
595 89 (10%) for Chalcolithic İlipinar. The observed adult sex ratios (females / females + males) based on cattle
596 pelvis morphology are 3 / 4 (75%) for Barçın Höyük, 0 / 0 for Neolithic İlipinar, 0 / 0 for Menteşe Höyük,
597 and 3 / 5 (60%) for Chalcolithic İlipinar. As in the Pinarbaşı B example, observations of 0 / 0 impart no
598 information onto the prior distribution of the adult sex ratio. All demographic and measurement data come
599 from the OpenContext datasets (Buitenhuis 2013; Dataset 2013, 2014); the ‘R script for replication’ includes
600 the steps for data processing and analysis.

601 Previous syntheses of the Late Neolithic and Early Chalcolithic animal economies in northwest Anatolia
602 provide several assumptions about the age and sex structure of cattle bone assemblages that can be evaluated
603 with the results of the Bayesian multilevel mixture model. First, the general cultural continuity of the
604 assemblages suggests that the biometry and composition of cattle bone assemblages may be similar at
605 the sites, having been produced by similar processes (e.g., Çakırlar 2013; Özdogan 2019). Second, the
606 widespread evidence of milk consumption from pottery residue analyses from these sites and others in the
607 region (Evershed et al. 2008; Thissen et al. 2010) suggest that cattle were managed for milk production

Table 5: Elemental Composition of the Northwest Anatolian Cattle Assemblages

Element Portion	Barcın Höyük	Ilıpınar Höyük (Late Neolithic/Transitional)	Ilıpınar Höyük (Early Chalcolithic)	Menteşe Höyük
Astragalus	5	19	14	15
Calcaneus	4	13	13	1
Femur (Distal)	0	5	3	0
Femur (Proximal)	0	4	4	0
Humerus	0	17	39	2
Metacarpal (Distal)	2	6	28	0
Metacarpal (Proximal)	3	19	9	3
Metatarsal (Distal)	5	6	16	2
Metatarsal (Proximal)	4	10	10	1
First Phalanx	16	76	35	8
Second Phalanx	21	49	29	14
Radius (Distal)	1	8	22	0
Radius (Proximal)	2	10	35	0
Scapula	0	9	14	0
Tibia (Distal)	4	9	21	0
Tibia (Proximal)	0	2	4	0
Total	67	262	296	46

608 (Gourichon and Helmer 2008; Roodenberg 2012a); thus, one may expect that each assemblage has a female-
609 dominated sex ratio and potentially higher fusion rates for later-fusing elements among females than males.
610 The multilevel modeling results can be used to evaluate the feasibility of these assumptions by examining
611 posterior distributions of relevant parameters and simulations of sex-specific fusion rates.

612 Because the model is a multisite model and deals with a different taxon than the original simulations,
613 the prior distributions for the model hyper-parameters are again redefined to reflect different expectations
614 of biological feasibility. While the multisite simulation provides useful prior distribution definitions for most
615 of the parameters, two other parameters (average body size of females μ_2 and index of sexual dimorphism
616 $\log(\delta_2)$) should be further changed because of different expectations modeling cattle rather than sheep. The
617 change in the prior distribution definition of μ_2 reflects the fact that the standard measurements for cattle
618 come from an aurochs female (Degerbøl 1970), which is expected to be larger than the domestic cattle
619 females in the assemblages. Cattle are expected to be more sexually dimorphic than sheep, which is reflected
620 in increasing the average expected value of $\log(\delta_2)$, resulting in an expectation of 0.14 LSI_e units between
621 males and females on average. This is slightly lower than index of sexual dimorphism seen in the Degerbøl
622 (1970) aurochs specimens [Grigson (1989), Figure 2, which uses LSI₁₀; the equivalent size difference is 0.06
623 on the LSI₁₀ scale], though domestic cattle may be expected to be less sexually dimorphic than their wild
624 counterparts (e.g., Tchernov and Horwitz 1991).

$$\begin{aligned}
\theta_1 &\sim \text{Normal}(-0.5, 1.5) \\
\theta_2 &\sim \text{Normal}(0.0, 1.5) \\
\mu_2 &\sim \text{Normal}(-0.1, 0.1) \\
\log \delta_1 &\sim \text{Normal}(-3.5, 0.5) \\
\log \delta_2 &\sim \text{Normal}(-2.0, 0.5) \\
\log \sigma_1 &\sim \text{Normal}(-3.05, 0.25) \\
\log \sigma_2 &\sim \text{Normal}(-3.10, 0.2) \\
\log \sigma_3 &\sim \text{Normal}(-3.10, 0.2)
\end{aligned} \tag{12}$$

626

4. RESULTS

627

4.1 Simulated Assemblages

628 Bayesian models work by updating prior information with new data to produce posterior distributions of
 629 parameters of interest (Otárola-Castillo et al. 2022). Thus, the difference between a model parameter's prior
 630 and posterior distribution shows the amount that the model "learns" from the data-if the data do not provide
 631 relevant information on a parameter's potential values, then the posterior distribution will resemble the prior
 632 distribution. Figure 5 compares the prior and posterior distributions of the main model hyper-parameters
 633 for the single assemblage simulation. The results show that the data provides much more information about
 634 the likely values of the two demographic parameters (the proportion of immature animals, π_1 , and the
 635 adult sex ratio, $\frac{\pi_2}{\pi_2 + \pi_3}$) and the average body size for females (μ_2). This is largely to be expected, as the
 636 prior distribution definitions were weakly-informative priors (Gelman et al. 2008), but also shows how these
 637 choices did not appear to severely influence the resulting posterior distributions.

638 The prior distribution definitions for the size offsets (δ_1 and δ_2) and the size variability estimates (σ_1 , σ_2 ,
 639 and σ_3) have a lot more overlap between the prior distributions and their respective posterior distributions.
 640 This overlap stresses the importance of using a Bayesian framework, particularly one relying on informative
 641 prior distributions, to produce meaningful parameter estimates from zooarchaeological data. But it also
 642 highlights the interpretive weight given to the reference population. However, the overlap is not necessarily
 643 a drawback of the model, as again the prior distribution definitions were designed as informative priors,
 644 specifically to ensure that the resulting parameter estimates would be biologically feasible. Further, the
 645 simulated population also has the same underlying biological population (the Shetland sheep population)
 646 that was used to develop the prior distributions, so it is possible that this overlap reflects that fact.

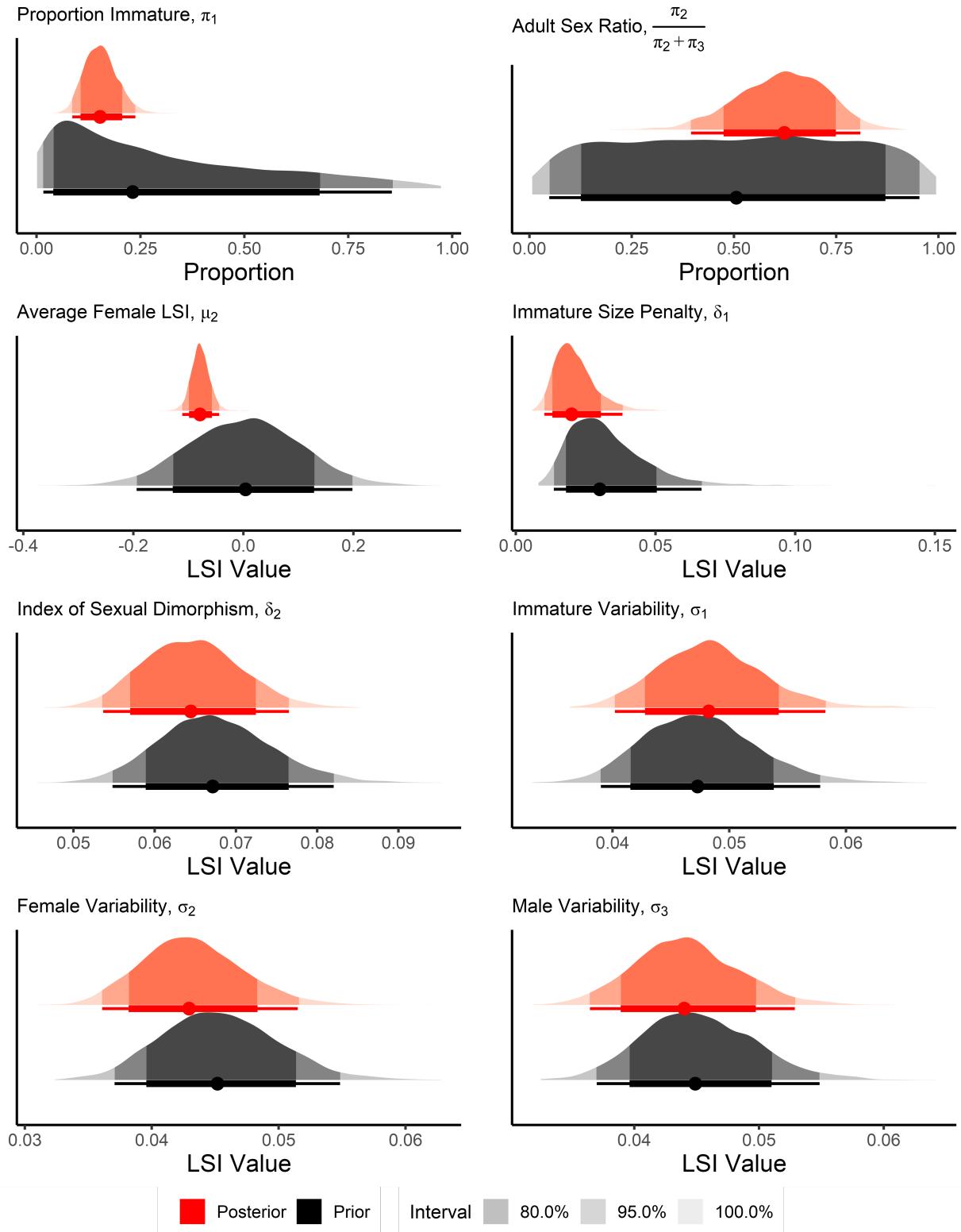


Figure 5: Prior-Posterior Comparison of Single Assemblage Simulation Model Hyper-Parameters

647 Parametric accuracy for the simulations relies on relating the posterior distributions for “assemblage-
 648 level” parameters to the known values for the overall Shetland sheep population (including any relevant
 649 modifications). Ideally, the 80% and 95% credible intervals from these posterior distributions should contain
 650 the true population value 80% and 95% of the time, respectively. An overfit model would contain the
 651 population value less frequently than the stated interval, producing false confidence in the applicability of
 652 the results. An underfit model, by contrast, would contain the population value more frequently than the
 653 stated interval, producing results that are too conservative. The single-assemblage simulation produces overfit
 654 results: 8 / 9 (89%) of the 80% credible intervals and 9 / 9 (100%) of the 95% credible intervals from the
 655 posterior distributions of the population parameters contain the true value of the Shetland sheep population.
 656 The multisite simulation produces relatively well-calibrated estimates for assemblage-level parameters, with
 657 97 / 135 (72%) of the 80% credible intervals and 128 / 135 (95%) of the 95% credible intervals from the
 658 posterior distributions of the (site-level) population parameters containing the true values of the Shetland
 659 sheep population with relevant modifications.

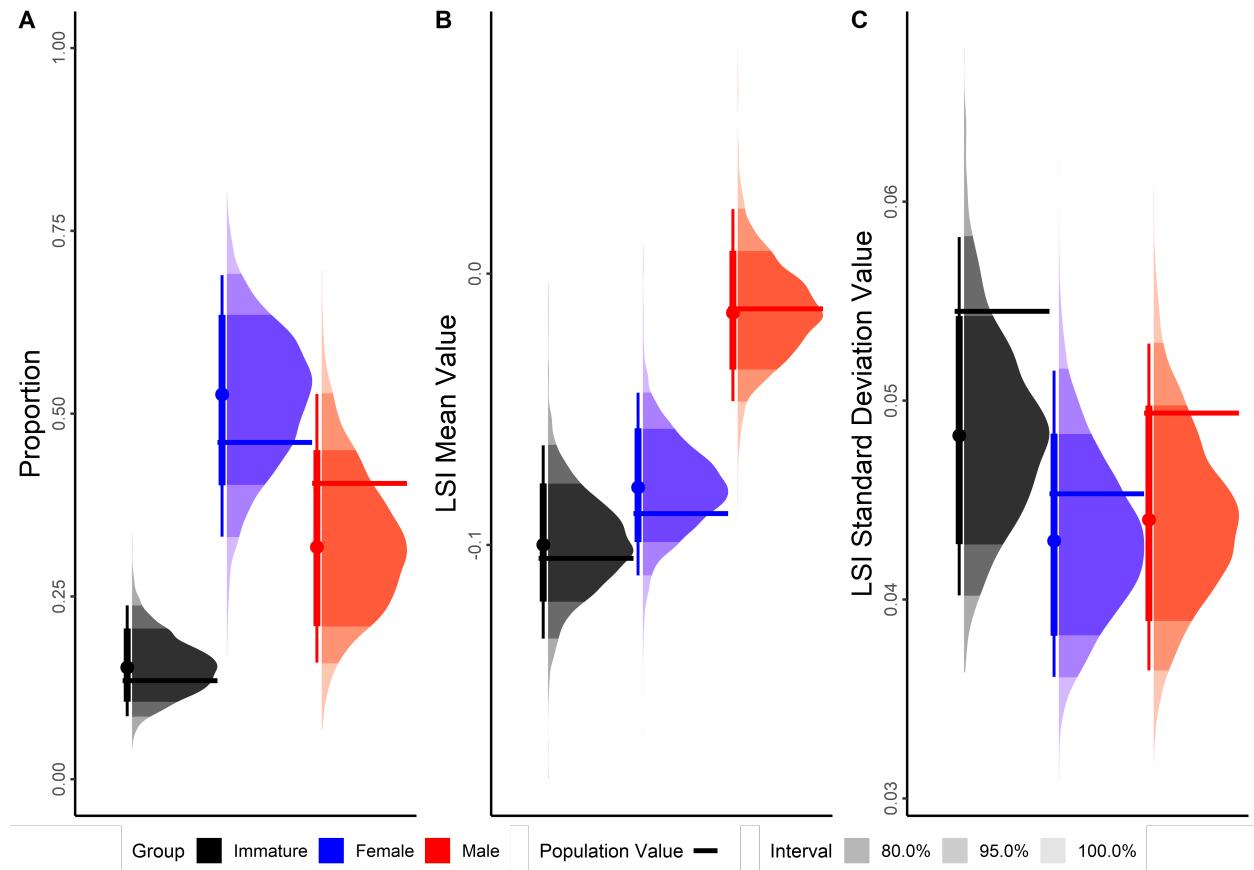


Figure 6: Posterior Distributions of Model Parameters for the Simulated Single Assemblage with Relevant Population Parameter Values

660 Figure 6 shows the posterior distributions of the population parameters for the single-assemblage model,
661 highlighting the 80% and 95% credible intervals and the corresponding value from the overall Shetland sheep
662 assemblage (“Population Value”). The single assemblage model estimates all of the model hyper-parameters
663 relatively accurately, though does tend to underestimate within-group size variability ($\sigma_{1,2,3}$). The multisite
664 model severely overfits when estimating the average body size for the three groups at each site, with 42 /
665 45 (93%) of the 80% credible intervals and 45 / 45 (100%) of the 95% credible intervals for the posterior
666 distributions of the site-specific model parameters containing the true population values (Figure 7). Size
667 variability ($\sigma_{1,2,3}$) is estimated relatively well, though again tends to underestimate size variability, while
668 proportions may not be accurate for extremely low populations values (e.g., immature animals in Sites 11-15).

669 Figure 8 shows posterior distributions of simulated group-specific compositions for both the single-
670 assemblage and total composition of the multisite models (i.e., all sites combined) alongside true counts
671 for each group. The models estimate the group-specific composition of the assemblages with great accuracy:
672 24 / 24 (100%) of the 80% and 24 / 24 (100%) of the 95% credible intervals contain the true group-specific
673 count for the element portion in the single-assemblage model and 21 / 24 (88%) of the 80% credible inter-
674 vals and 24 / 24 (100%) of the 95% credible intervals contain the true group-specific count for the element
675 portion in the combined multisite model. The multisite model also accurately estimates the group-specific
676 composition of each assemblage: 331 / 339 (98%) of the 80% credible intervals and 337 / 339 (99%) of the
677 95% credible intervals contain the true group-specific count.

678 This overfitting is to be expected since the measured assemblage is a subsample of the modeled assemblage.
679 The model does not overfit as much when estimating the composition of the modeled assemblages: for the
680 single-assemblage model, 20 / 24 (83%) of the 80% credible intervals and 23 / 24 (96%) of the 95% credible
681 intervals contain the true group-specific counts for the element portions. For the multisite model, the model
682 does not perform well for the combined assemblages (17 / 24 (71%) of the 80% credible intervals and 22 /
683 24 (92%) of the 95% credible intervals) but produces well-calibrated estimates for the site-specific estimates
684 of group composition, with 309 / 360 (86%) of the 80% credible intervals and 353 / 360 (98%) of the 95%
685 credible intervals containing the true group-specific counts. Note that the denominator for the site-specific
686 estimates differs between the measured and modeled assemblages. This is caused by the additional sampling
687 to create modeled assemblages and the multilevel structure of the model, which estimates element-specific
688 offsets and interaction terms (ν_{element} and $\nu_{\text{interaction}}$) for elements that are present in at least one site. Among
689 the 7 newly-observed elements in the modeled assemblage, 18 / 21 (86%) of the 80% credible intervals and
690 21 / 21 (100%) of the 95% credible intervals included the true group-specific abundance, despite having
691 no observed measurements. That these estimates are broadly as accurate as the estimates from observed
692 element portions suggests that the hyper-parameters can be used to create estimates of unobserved (i.e.,

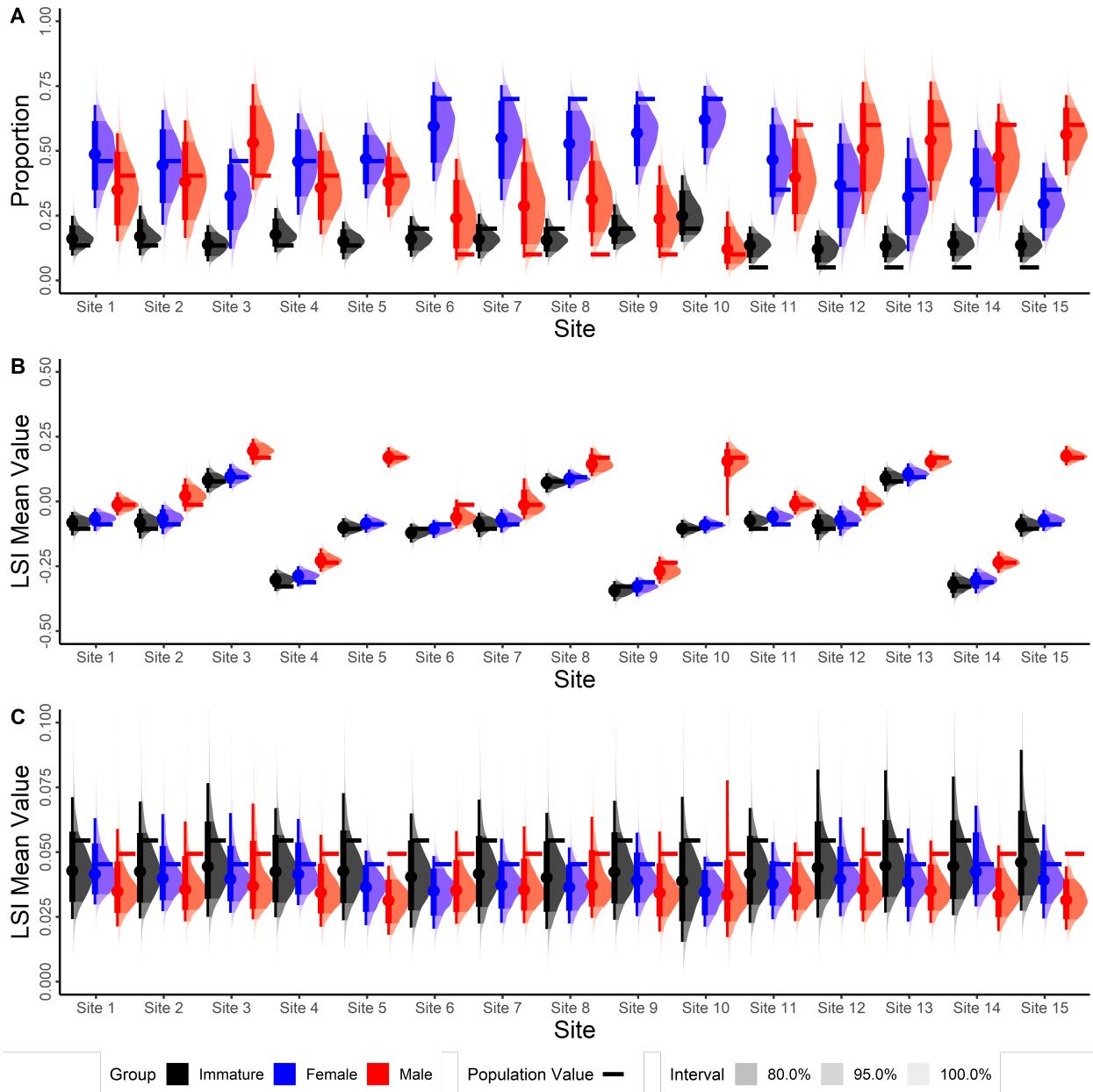


Figure 7: Posterior Distributions of Model Parameters for the Simulated Multisite Assemblage with Relevant Population Parameter Values

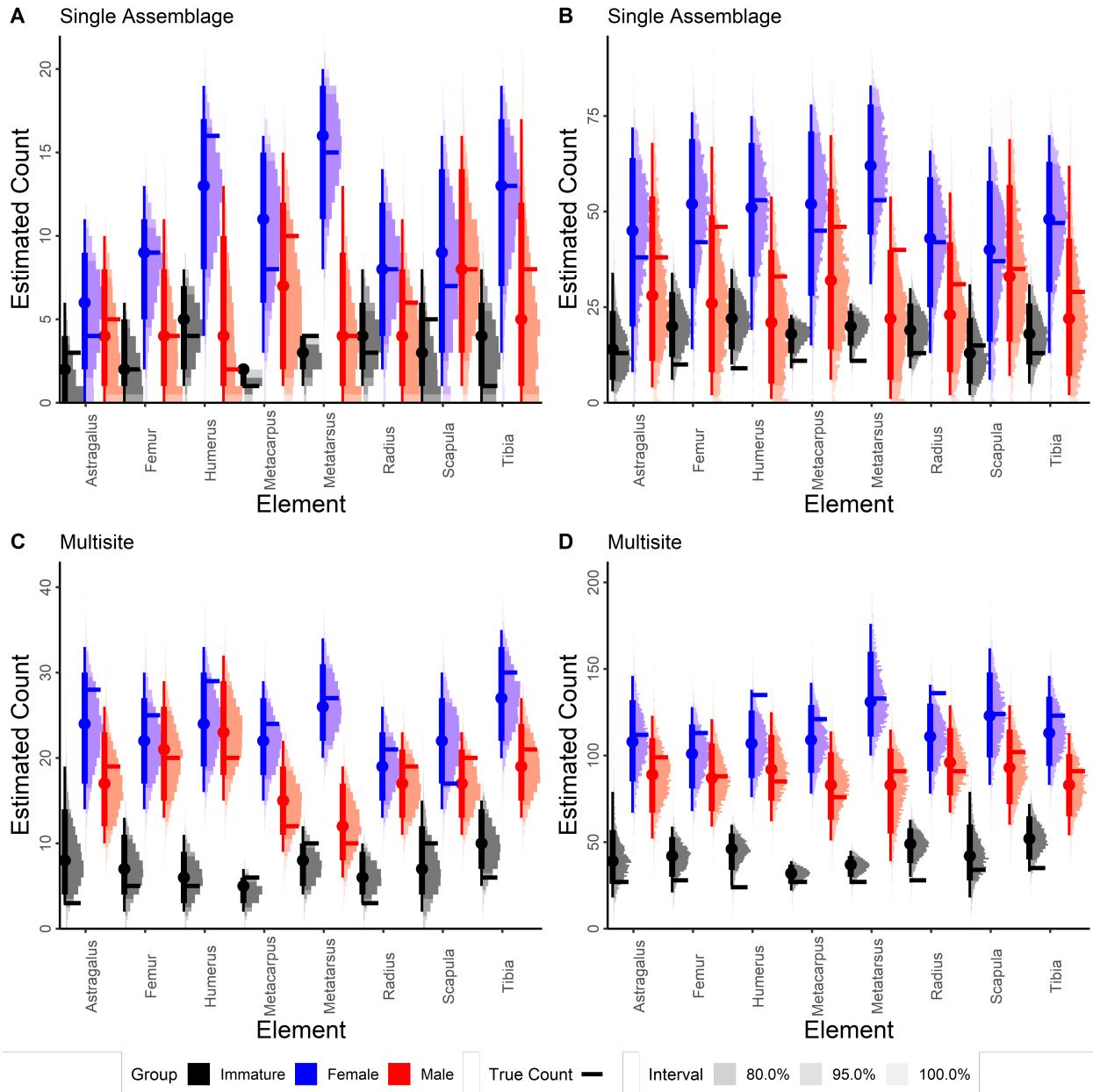


Figure 8: Posterior Distributions of Group-Specific Compositions for the Simulated Measured and Modeled Assemblages

693 unmeasured) element portions in an assemblage.

694 4.2 Archaeological Case Studies

695 4.2.1 Pinarbaşı B Sheep

696 Figure 9A shows the posterior distributions assemblage-level π mixture components. In general, the
697 Pinarbaşı B sheep assemblage is overwhelmingly composed of immature animals (posterior π_1 median =
698 89%; 95% posterior confidence interval for $\pi_1 = 80\text{--}95\%$), somewhat lower than the observed fusion rate of
699 proximal and middle phalanges ($59 / 62 = 95\%$). The posterior distribution of the overall adult sex ratio
700 (θ_{female}) suggests that females are more common than males (74% of the posterior samples are above 0.5),
701 but this estimate is uncertain, owing to the low overall proportions of female and male animals.

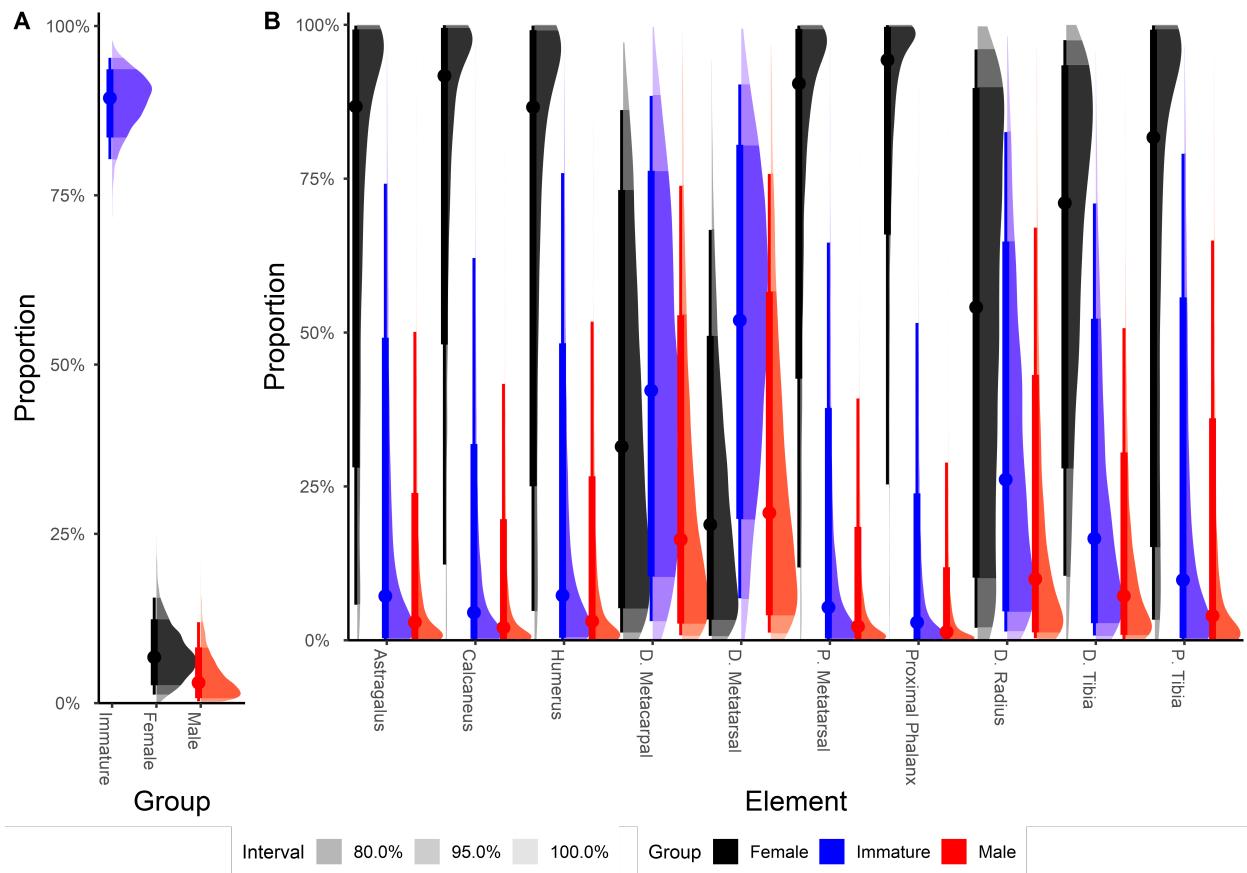


Figure 9: Posterior Distributions of (A) Overall Mixture Components (π) and (B) Element-Specific Mixture Components (π) for the Pinarbaşı B Sheep Assemblage

702 Figure 9B shows the element-specific distributions of π_1 for the Pinarbaşı B assemblage. Half of the
703 element-specific π_1 distributions are similarly concentrated in the upper range of possible π_1 values, with
704 posterior medians over 85%; all specimens from these element portions were considered potentially immature

705 by the model due to fusion status or being an element that does not fuse or exhibits post-fusion growth.
706 However, these element-specific distributions also have long tails extending into lower π_1 values, conveying
707 less certainty about element-specific π_1 estimates relative to the assemblage-wide estimate. This likely owes
708 to small element-specific sample sizes (the astragalus, calcaneus, and proximal phalanx have 9-10 specimens,
709 all other element-specific samples sizes are 1-4, see Table 4) and to the presence of some element portions with
710 lower modeled π_1 values. These element portions—especially the distal metacarpal and distal metatarsal—
711 have posterior π_1 median values below 50%, though again have long tails that extend into higher π_1 values.
712 Notably, all measured specimens from these two elements are not considered potentially immature because
713 their distal epiphyses are fused.

714 Figure 10 shows the distribution of simulated compositions for immature, female, and male specimens
715 in three Pınarbaşı B assemblages: the measured assemblage ($N = 44$), the assemblage of modeled element
716 portions ($N = 277$), and the full sheep assemblage including five element portions that were not modeled
717 due to lack of measurements (additional elements: proximal radius, ulna, proximal fused metacarpal 3 and
718 4, pelvis, and middle phalanx; total $N = 428$). While the model has much greater precision in estimating the
719 composition of the measured assemblage, including the non-measured specimens in the modeled assemblage
720 provides a clearer overall picture of the assemblage. Most element portions are composed predominantly of
721 immature specimens rather than adults (males or females), with the notable exceptions of distal metapodials
722 and, to a lesser extent, the distal radius. Inclusion of the unmodeled elements broadly reinforces this pattern,
723 while providing data to explore fusion rates or comparisons of larger animal portions.

724 **4.2.2 7th-6th Millennium BCE Northwest Anatolian Cattle**

725 Figure 11 shows the posterior distributions of average body sizes for female cattle (μ_2) from the four an-
726 alyzed assemblages. These distributions are produced from posterior samples; assemblage-specific estimates
727 from a single posterior sample share the same relevant hyper-parameters (μ_{μ_2} and $\sigma_{\text{site}[4]}$), meaning that
728 they covary with one another to an extent. To compare these distributions, then, a contrast is necessary
729 to account for this potential covariation. This is done by simply evaluating the difference between two
730 parameters (e.g., between the average female LSI_e value μ_2 for Barçın Höyük and μ_2 for Neolithic İlipinar)
731 in each posterior sample, shown in the right-hand panel of Figure 11. These contrasts show that the the
732 female cattle from Chalcolithic İlipinar are likely smaller, on average, than female cattle from the other sites.
733 These cattle measurements are smaller, on average, than those from the other northwestern Anatolian sites
734 relative to the standard animal's measurements.

735 Despite this size difference in female animals between the assemblages, the age and sex composition of the
736 four assemblages are broadly consistent with one another. Figure 12 shows the distributions of assemblage-

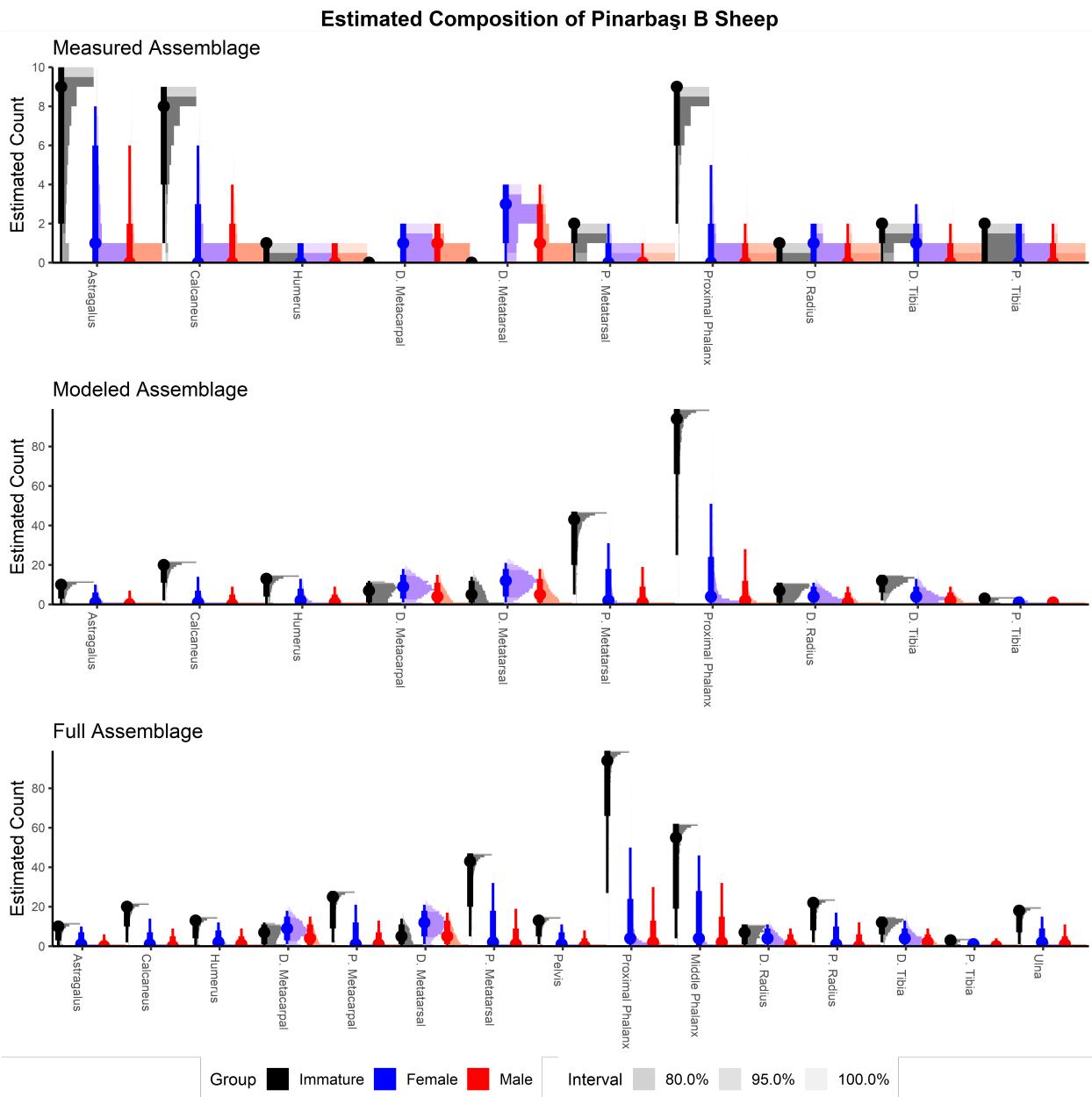


Figure 10: Posterior Distributions of Group-Specific Composition for the Pınarbaşı B Sheep Measured, Modeled, and Full Assemblages

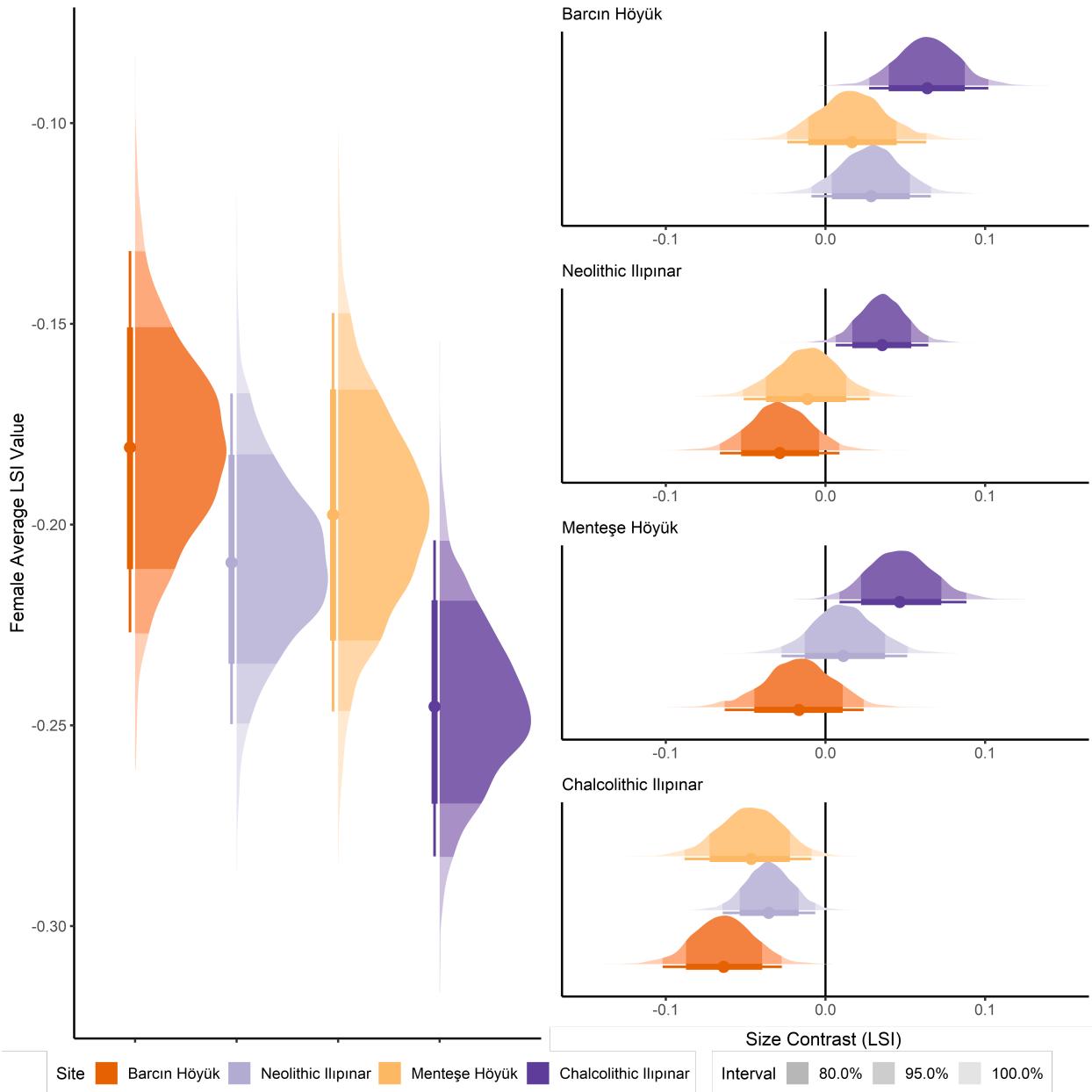


Figure 11: Posterior Distributions of Site-Specific Average Female LSI Values (μ_2) for the Northwest Anatolian Cattle Assemblages. The right panel shows site-specific contrasts for this parameter, indicating specific size differences between pairs of sites

737 level demographic variables—the proportion of immature animals and the adult sex ratio (the proportion of
738 adults that are female)—for the four northwest Anatolian assemblages. As expected from earlier syntheses
739 that indicated herds kept to produce milk, all of the assemblages have female-dominated adult sex ratios. The
740 relatively large proportions of immature cattle in the Menteşe Höyük and Barçın Höyük assemblages may
741 indicate that cattle were penned on the site, where infants killed by herders or natural causes were included in
742 the assemblage (Gillis et al. 2014, 2015). Contrasts indicate that both İlipinar Höyük assemblages have lower
743 the proportions of immature animals than either the Barçın Höyük or Menteşe Höyük assemblages. This
744 suggests that, despite evidence for penning deposits on-site for Chalcolithic İlipinar (Roodenberg 2012a),
745 cattle were not giving birth at the site or immature specimens did not preserve as well (Gillis et al. 2014).
746 The small proportion of very young mandibles in the assemblage corroborates the biometric modeling results
747 (Buitenhuis 2008: Figure 18).

748 Simulating sex-specific fusion rates for late-fusing elements (proximal femur, distal femur, proximal
749 humerus, distal radius, proximal tibia, proximal ulna: Grigson 1982) from the full northwest Anatolian
750 assemblages highlights the complexities of examining sex-specific fusion rates in zooarchaeological assem-
751 blages. In each assemblage, estimates of male fusion rates are extremely uncertain, owing to the small
752 number of estimated males in each iteration and thus large potential shifts in the denominator for fusion
753 rates (Figure 13). This uncertainty makes it difficult to clearly establish whether fusion rates differed between
754 males and females; regardless, in 67% of the posterior samples female fusion rates were higher than male
755 fusion rates for Chalcolithic İlipinar. This provides some support for the idea that males in the assemblage
756 were killed at younger ages than females, consistent with a milk-producing management strategy that kept
757 females alive longer than males (Zeder and Hesse 2000; Gillis et al. 2014).

758 5. DISCUSSION

759 The simulation analyses show that the Bayesian multilevel mixture model presented here can accurately re-
760 construct age- and sex-specific biometry of a faunal population represented in a measured assemblage, while
761 also producing relatively accurate estimates of the “demographic” (age and sex) composition of the assem-
762 blage. The performance of the Bayesian multilevel mixture models relies on the prior distributions, which
763 provide constraints against overfitting and ensure that the model produces biologically reasonable parameter
764 estimates. The prior distribution definitions in this paper were derived largely from the measurements of a
765 herd of known-age, known-sex population of Shetland sheep (Popkin et al. 2012), though for the multisite
766 cattle model some of the definitions were changed based on data on European aurochs (Degerbøl 1970). It
767 is important to note that prior distribution definitions can be derived from many different sources—including

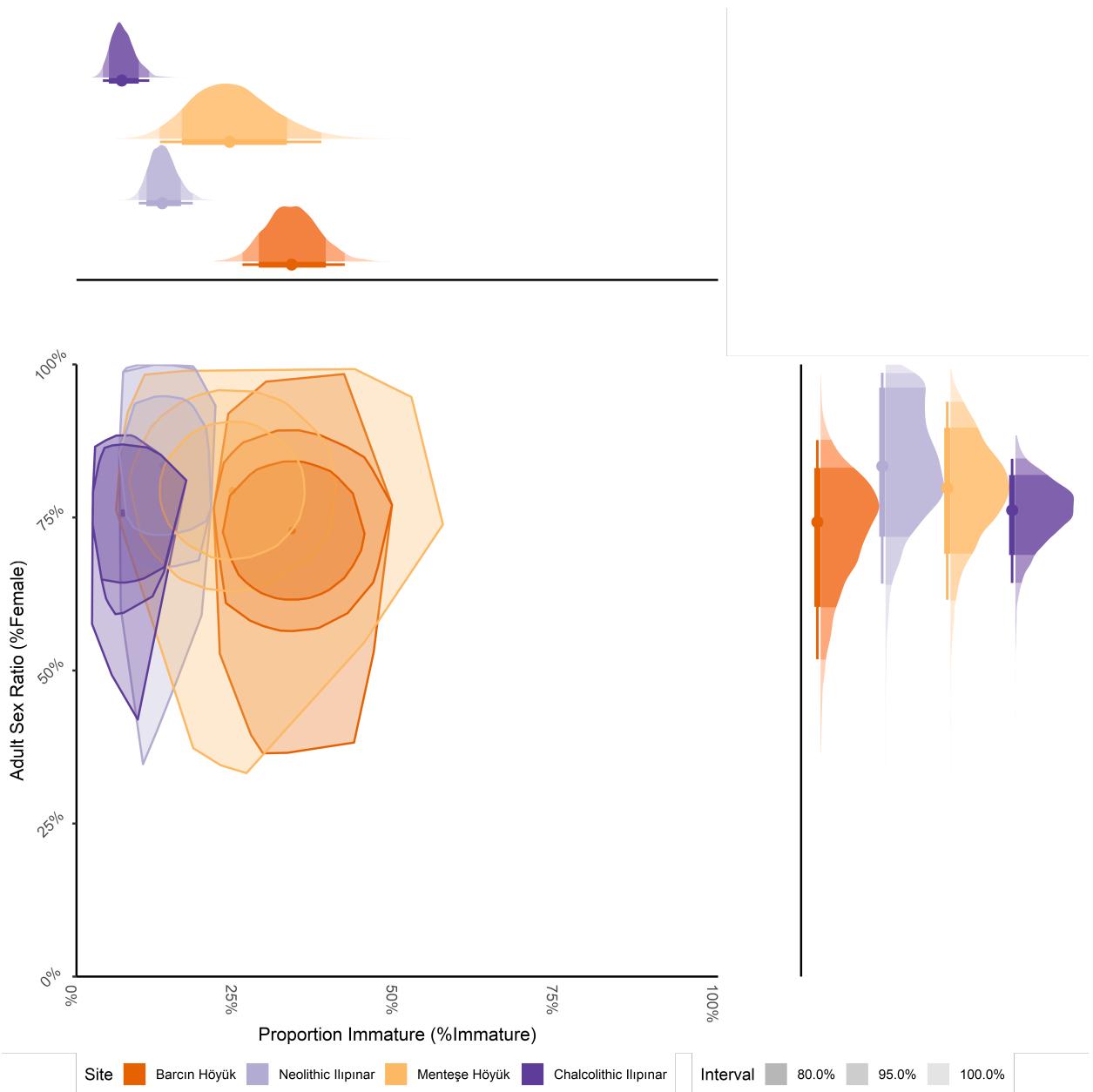


Figure 12: Comparison Posterior Distributions of Site-Specific Demographic Parameters (π_{immature} and θ_{female}) for the Northwest Anatolian Cattle Assemblages. Side panels show marginal plots to compare the distributions of each parameter individually across sites

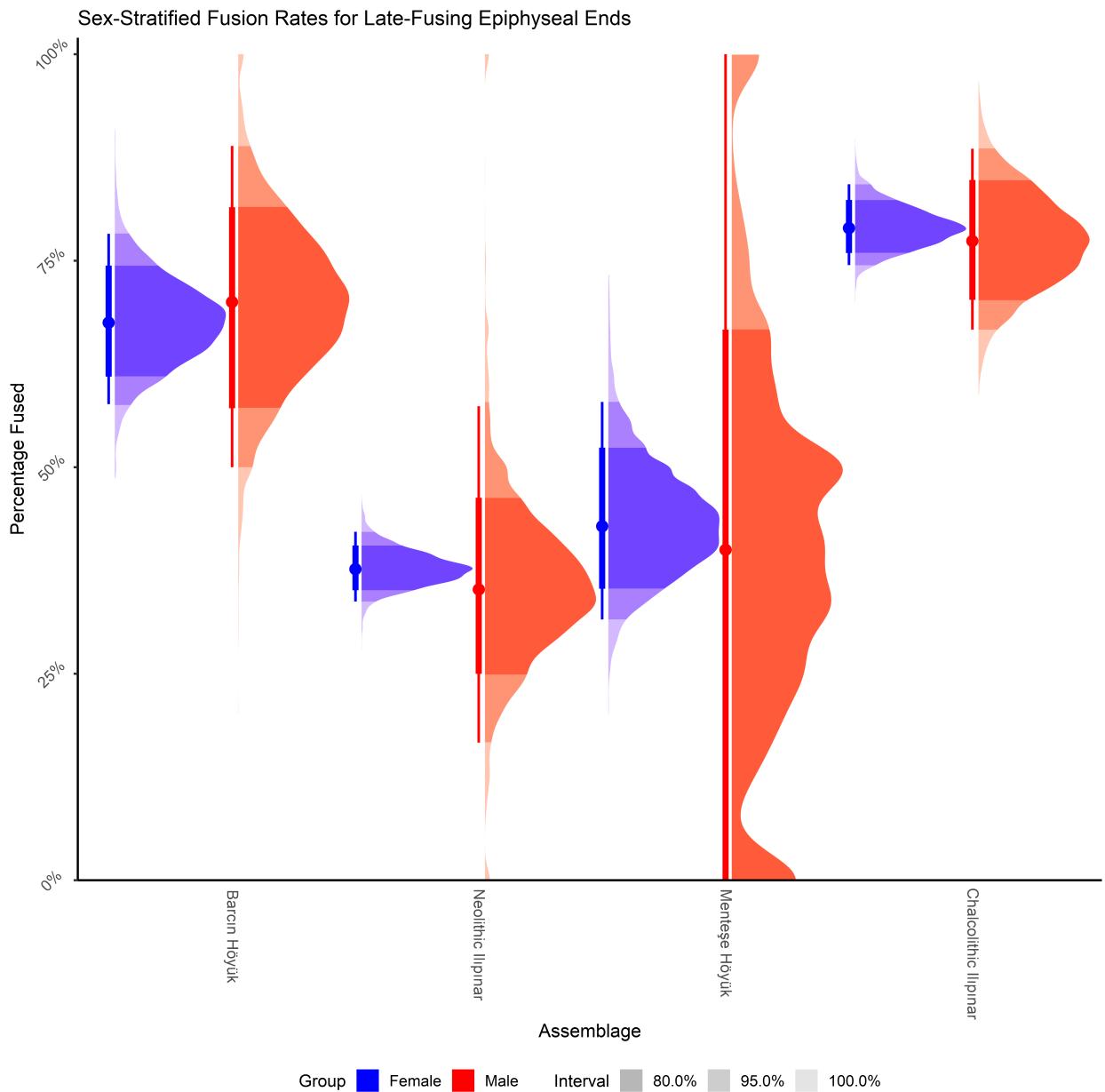


Figure 13: Posterior Distributions of Sex-Specific Fusion Rates for Late-Fusing Elements in Northwest Anatolian Cattle Assemblages

768 quantification based on one's judgment (e.g., Gelman et al. 2008; McCarthy and Masters 2005). More im-
769 portant than the source of one's prior distribution definitions, is investigating the expectations of those prior
770 distribution definitions by performing prior predictive checks as in Section 2.3 (Gabry et al. 2019; Gelman,
771 Vehtari, et al. 2020). Further, emphasis should be paid to increasing the diversity of known-age, known-sex
772 animal populations with individual measurement data (e.g., Lebennon and Munro 2022; Zeder and Lemoine
773 2020), which could help develop prior distributions relevant to different taxa and to understand how variable
774 different parameters, especially size variability (σ) parameters, are across populations.

775 One of the central tenets of the mixture model's extension to modeled assemblages is the idea that
776 "measurability" (adequate preservation to maintain a biological measurement) is unrelated to a specimen's
777 status as immature, female, or male. Variation in the mixture proportions π among elements, especially the
778 proportion of immature specimens (π_1) may highlight group-specific biases in the deposition of specimens
779 but could also indicate issues with the assumption that "measurability" is random. The Pinarbaşı B sheep
780 assemblage potentially demonstrates this issue, as the distal metapodials have much lower element-specific
781 π_1 estimates than other element portions. While it is plausible that metapodial bones from adult sheep
782 were selectively over-represented in the assemblage, say as a cache of raw materials, it is also likely that
783 distal metapodials from immature animals-particularly very young animals-are less likely to be measurable
784 compared to adult animals. Because the distal breadth measurement requires both distal condyles to be
785 present, distal metapodial specimens from neonatal or extremely juvenile individuals may be missed while
786 those from other element portions (e.g., proximal metapodial, distal humerus) would still be theoretically
787 measurable (Martín and García-González 2015). The inclusion of condyle-specific measurements could ad-
788 dress this issue, though would require identifying whether the isolated condyle is medial or lateral (e.g.,
789 width of condyle: Payne 1969). Still, noting this inter-element discrepancy in the Pinarbaşı sheep material
790 provides more nuance to the understanding of the assemblage's composition by further corroborating the
791 argument that many of the immature remains derive from neonatal individuals (Carruthers 2005).

792 The ability to create accurate simulated estimates of age and sex composition provides many opportu-
793 nities for further analyses. For instance, comparison of the composition of animals in different depositional
794 contexts could support contextual taphonomic analyses (e.g., Meier 2020). Access to certain kinds of animals
795 could highlight systems of provisioning or status-related restrictions (Arbuckle 2012; Twiss 2019: 73-97). Dif-
796 ferences in the age and sex composition of different body parts could also highlight ritual behaviors reflected
797 in the use of certain contexts or sites (e.g., Madgwick and Mulville 2015). On a more practical level, pro-
798 viding specimen-specific probabilities of being immature, female, or male can provide a useful baseline for
799 sampling strategies focused on ancient DNA or stable isotopes, allowing researchers to explore potential sex
800 differences in diets (e.g., Post et al. 2001) or more easily identify male specimens to isolate Y-chromosomal

801 DNA to explore sex-specific selection (e.g., McGrory et al. 2012; Daly et al. 2021).

802 The archaeological case studies highlight the importance of considering the presence of immature speci-
803 mens and elemental variation in body size when summarizing the biometry and composition of an assemblage,
804 problems addressed by the Bayesian multilevel mixture model. Variation in the proportion of immature
805 animals in the assemblage, as in the multisite case study for northwestern Anatolia, may also highlight dif-
806 ferences in culling strategies or even the seasonality of animal presence at the sites. The vast majority of the
807 Pinarbaşı B material derives from immature specimens, which could complicate inter-assemblage biometric
808 analyses that do not use sex-specific size estimates (e.g., Arbuckle et al. 2014). Restricting measurements
809 only to fused specimens removes useful information, particularly when fusion rates may differ between male
810 and female animals (Zeder and Hesse 2000); further, it does not resolve the problem of immature animals
811 in the measurement assemblage if early-fusing elements like the distal humerus are still included because of
812 post-fusion growth (Popkin et al. 2012). The ability to create sex-specific biometric estimates is important
813 to document large-scale spatial and temporal dynamics in animal body size (e.g., Arbuckle et al. 2016).

814 The increased ability to specify the age and sex composition of faunal assemblages with Bayesian multi-
815 level mixture models also highlights the limitations of our current language used to describe and interpret
816 these compositions. Many discussions that examine changes in the composition of faunal assemblages to
817 identify changes in exploitation patterns use terms like “prime-dominated age structure” (e.g., Stiner 1990),
818 “dominance of females,” (e.g., Peters, Driesch, and Helmer 2005) or “young male slaughter/kill-off” (e.g.,
819 Zeder and Hesse 2000; Arbuckle and Atici 2013). These terms are deceptive in their utility—they describe
820 some empirical pattern but it is up to the individual researcher to define the cut-off between a “dominant”
821 and “non-dominant” assemblage. In the case of the adult sex ratio for Barçın Höyük, 90% of the poste-
822 rior samples are above 60% (1.5 females:1 male), but only 47% of the posterior samples are over 75% (3
823 females:1 male). Meanwhile, 64% of the posterior samples for the adult sex ratio for Neolithic İlpınar are
824 over 80% (4 females:1 male). Are both assemblages “dominated by females”? More formalized language
825 in our hypotheses—or, rather, the adoption of statistical modeling frameworks (McElreath 2020: 4-17)—is
826 necessary to clarify what changes in assemblage-level estimates of biometry and composition mean for past
827 human-animal interactions.

828 6. CONCLUSIONS

829 This paper describes a new method for estimating the biometry and age/sex composition of faunal assem-
830 blages based on standard measurement data, Bayesian multilevel mixture modeling. The model produces
831 accurate estimates of sex-specific biometry, which can provide a more useful framework for inter-assemblage

832 analysis (e.g., Arbuckle et al. 2016; Helmer et al. 2005). Such a framework could better explore broad spatial
833 and chronological patterns in animal biometry while accounting for differences in assemblage composition
834 across the assemblages, ensuring reliable comparisons of animal body size in relation to other variables.
835 These analyses could investigate the processes behind size fluctuation in animals, particular in relation to
836 changing human-animal interactions and adaptation to new lifeways and anthropogenic environments.

837 Furthermore, the estimates of the age and sex composition of the assemblage can be used to simulate
838 assemblages of specimens with known group assignment (immature, female, and male). These simulations
839 are the baseline for comparing differences in the composition of sub-assemblages. Doing so allows researchers
840 to directly test hypotheses about differences in the age and sex composition of animal bones from different
841 parts of a site, different fusion groups, or other categories. The Bayesian structure of the model allows
842 researchers the flexibility to create bespoke hypotheses that can be tested directly, rather than relying on null
843 hypothesis testing for inference (Otárola-Castillo and Torquato 2018; Otárola-Castillo et al. 2022). Thus, the
844 mixture modeling framework described here provides a foundation for biometric and compositional analyses
845 that operate at multiple scales and present a new avenue for summarizing and comparing zooarchaeological
846 assemblages.

847 ACKNOWLEDGEMENTS

848 The initial modeling work was developed as part of my dissertation research, which was partially funded by a
849 Wenner-Gren Foundation Dissertation Fieldwork Grant (Grant #9192). I am grateful to the comments and
850 insights on earlier versions of this manuscript from Zenobie Garrett and Heather Lynch. Any errors are my
851 own.

852 References

- 853 Arbuckle, Benjamin S. 2012. "Animals and Inequality in Chalcolithic Central Anatolia." Journal Article.
854 *Journal of Anthropological Archaeology* 31 (3): 302–13. <https://doi.org/10.1016/j.jaa.2012.01.008>.
- 855 Arbuckle, Benjamin S., and Levent Atici. 2013. "Initial Diversity in Sheep and Goat Management in
856 Neolithic South-Western Asia." Journal Article. *Levant* 45 (2): 219–35.
- 857 Arbuckle, Benjamin S., Sarah Whitcher Kansa, Eric C. Kansa, David C. Orton, Canan Cakirlar, Lionel
858 Gourichon, A. Levent Atici, et al. 2014. "Data Sharing Reveals Complexity in the Westward Spread of
859 Domestic Animals Across Neolithic Turkey." Journal Article. *PLoS One* 9 (6): 1–11. <https://doi.org/10.1371/journal.pone.0099845>.
- 860 Arbuckle, Benjamin S., Max D. Price, Hitomi Hongo, and Banu Öksüz. 2016. "Documenting the Initial

- 862 Appearance of Domestic Cattle in the Eastern Fertile Crescent (Northern Iraq and Western Iran).”
863 Journal Article. *Journal of Archaeological Science* 72: 1–9. <https://doi.org/10.1016/j.jas.2016.05.008>.
- 864 Baird, Douglas, Eleni Asouti, Laurence Astruc, Adnan Baysal, Emma Baysal, Denise Carruthers, Andrew
865 Fairbairn, et al. 2013. “Juniper Smoke, Skulls and Wolves’ Tails. The Epipalaeolithic of the Anatolian
866 Plateau in Its South-West Asian Context; Insights from Pınarbaşı.” Journal Article. *Levant* 45 (2):
867 175–209.
- 868 Betancourt, Michael. 2017. “A Conceptual Introduction to Hamiltonian Monte Carlo.” Journal Article.
869 *arXiv* 1701.02434 [Preprint]: Available from <https://arxiv.org/abs/1701.02434v1>.
- 870 Breslawski, Ryan P. 2022. “Minimum Animal Units and the Standardized Count Problem.” Journal Article.
871 *Journal of Archaeological Method and Theory*. <https://doi.org/10.1007/s10816-022-09563-9>.
- 872 Breslawski, Ryan P., and David A. Byers. 2015. “Assessing Measurement Error in Paleozoological Osteo-
873 metrics with Bison Remains.” Journal Article. *Journal of Archaeological Science* 53: 235–42. <https://doi.org/10.1016/j.jas.2014.10.001>.
- 875 Buitenhuis, Hijlke. 2008. “Ilipinar: The Faunal Remains from the Late Neolithic and Early Chalcolithic
876 Levels.” Book Section. In *Archaeozoology of the Near East VIII*.
- 877 ———. 2013. “Ilipinar Zooarchaeology.” Dataset. <http://opencontext.org/projects/D297CD29-50CA-4B2C-4A07-498ADF3AF487>. <https://doi.org/https://doi.org/10.6078/M76H4FB>.
- 879 Çakırlar, Canan. 2013. “Rethinking Neolithic Subsistence at the Gateway to Europe With New Archaeo-
880 zoological Evidence from Istanbul.” Book Section. In *Barely Surviving or More Than Enough?: The*
881 *Environmental Archaeology of Subsistence, Specialisation and Surplus Food Production*, edited by Maaike
882 Groot, Daphne Lentjes, and Jørn Zeiler, 59–80. Leiden: Sidestone Press.
- 883 Carruthers, Denise. 2005. “Hunting and Herding in Central Anatolian Prehistory: The Sites at Pınarbasi.”
884 Book Section. In *Archaeozoology of the Near East VI*.
- 885 Corporation, Microsoft, and Steve Weston. 2022. *doParallel: Foreach Parallel Adaptor for the 'Parallel'*
886 *Package*. <https://CRAN.R-project.org/package=doParallel>.
- 887 Corti, Paulo, and David M. Shackleton. 2002. “Relationship Between Predation-Risk Factors and Sexual
888 Segregation in Dall’s Sheep (*Ovis Dalli Dalli*).” Journal Article. *Canadian Journal of Zoology* 80 (12):
889 2108–17. <https://doi.org/10.1139/z02-207>.
- 890 Crema, Enrico R. 2011. “Modelling Temporal Uncertainty in Archaeological Analysis.” Journal Article.
891 *Journal of Archaeological Method and Theory* 19 (3): 440–61. <https://doi.org/10.1007/s10816-011-9122-3>.
- 893 Dahl, Gudrun, and Anders Hjort. 1976. *Having Herds: Pastoral Herd Growth and Household Economy*.
894 Book. Stockholm Studies in Social Anthropology. Stockholm: University of Stokholm.

- 895 Daly, K. G., V. Mattiangeli, A. J. Hare, H. Davoudi, H. Fathi, S. B. Doost, S. Amiri, et al. 2021. "Hherded
896 and Hunted Goat Genomes from the Dawn of Domestication in the Zagros Mountains." Journal Article.
897 *Proc Natl Acad Sci U S A* 118 (25). <https://doi.org/10.1073/pnas.2100901118>.
- 898 Dataset. 2006.
- 899 ———. 2013.
- 900 ———. 2014. %3C<http://opencontext.org/subjects/AF6F6BD2-6387-4FBB-74D4-FD054C08EF79%3E>.
- 901 Davis, Simon J. M. 1981. "The Effects of Temperature Change and Domestication on the Body Size of Late
902 Pleistocene to Holocene Mammals of Israel." Journal Article. *Paleobiology* 7 (1): 101–14.
- 903 ———. 1996. "Measurements of a Group of Adult Female Shetland Sheep Skeletons from a Single Flock:
904 A Baseline for Zooarchaeologists." Journal Article. *Journal of Archaeological Science* 23: 593–612.
- 905 Degerbøl, Magnus. 1970. "The Urus (*Bos Primigenius Bojanus*) and Neolithic Domesticated Cattle (*Bos*
906 *Taurus Domesticus Linné*) in Denmark." Journal Article. *Det Kongelige Danske Videnskabernes Selskab*
907 *Biologiske Skrifter* 17 (1): 5–177.
- 908 Dong, Zhuan. 1997. "Mixture Analysis and Its Preliminary Application in Archaeology." Journal Article.
909 *Journal of Archaeological Science* 24: 141–61.
- 910 Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of 'Data.frame'*. <https://CRAN.R-project.org/package=data.table>.
- 911 Driesch, Angela von den. 1976. *A Guide to the Measurement of Animal Bones from Archaeological Sites*.
912 Book. Peabody Museum Bulletins. Cambridge, MA: Harvard University.
- 913 Evershed, Richard P., Sebastian Payne, Andrew G. Sherratt, Mark S. Copley, Jennifer Coolidge, Duska
914 Urem-Kotsu, Kostas Kotsakis, et al. 2008. "Earliest Date for Milk Use in the Near East and Southeastern
915 Europe Linked to Cattle Herding." Journal Article. *Nature* 455 (7212): 528–31. <https://doi.org/10.1038/nature07180>.
- 916 Fernández, Hélène, and Hervé Monchot. 2007. "Sexual Dimorphism in Limb Bones of Ibex (*Capra Ibex*
917 l.): Mixture Analysis Applied to Modern and Fossil Data." Journal Article. *International Journal of
918 Osteoarchaeology* 17: 479–91. <https://doi.org/10.1002/oa.876>.
- 919 Gabry, Jonah, and Rok Češnovar. 2022. *Cmdstanr: R Interface to 'CmdStan'*.
- 920 Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. "Visualization
921 in Bayesian Workflow." Journal Article. *Journal of the Royal Statistical Society, Series A* 182 (2): 389–
922 402.
- 923 Gelman, Andrew. 2006a. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do." Journal
924 Article. *Technometrics* 48 (3): 432–35. <https://doi.org/10.1198/004017005000000661>.
- 925 ———. 2006b. "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article

- 928 by Browne and Draper)." Journal Article. *Bayesian Analysis* 3: 515–34.
- 929 Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2020.
930 *Bayesian Data Analysis*. Book. Third Edition.
- 931 Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. "A Weakly Informative
932 Default Prior Distribution for Logistic and Other Regression Models." Journal Article. *The Annals of
933 Applied Statistics* 2 (4): 1360–83. <https://doi.org/10.1214/08-aoas191>.
- 934 Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren
935 Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. "Bayesian Workflow."
936 Journal Article. *ArXiv*.
- 937 Gerritsen, Fokke, and Rana Özbal. 2019. "Barcin Höyük, a Seventh Millennium Settlement in the Eastern
938 Marmara Region of Turkey." Journal Article. *Documenta Praehistorica* 46: 58–67. <https://doi.org/10.4312/dp.46.4>.
- 940 Gillis, R., R. M. Arbogast, J. F. Piningre, K. Debue, and J. D. Vigne. 2015. "Prediction Models for Age-
941 at-Death Estimates for Calves, Using Unfused Epiphyses and Diaphyses." Journal Article. *International
942 Journal of Osteoarchaeology* 25 (6): 912–22. <https://doi.org/10.1002/oa.2377>.
- 943 Gillis, R., I. Carrère, M. Saña Seguí, G. Radi, and J. D. Vigne. 2014. "Neonatal Mortality, Young Calf
944 Slaughter and Milk Production During the Early Neolithic of North Western Mediterranean." Journal
945 Article. *International Journal of Osteoarchaeology* 26 (2): 303–13. <https://doi.org/10.1002/oa.2422>.
- 946 Gourichon, Lionel, and Daniel Helmer. 2008. "Etude de La Faune Neolithique de Mentese." Book Section. In
947 *Life and Death in a Prehistoric Settlement in Northwest Anatolia. The Ilipinar Excavations, Volume III.*,
948 edited by Jacob Roodenberg and Songül Alpaslan Roodenberg, 435–48. Leiden: Nederlands Instituut
949 voor he Nabije Oosten.
- 950 Grigson, Caroline. 1982. "Sex and Age Determination of Some Bones and Teeth of Domestic Cattle: A
951 Review of the Literature." Book Section. In *Ageing and Sexing Animal Bones from Archaeological Sites*,
952 edited by Bob Wilson, Caroline Grigson, and Sebastian Payne, 7–23. BAR British Series. Oxford: British
953 Archaeological Reports.
- 954 ———. 1989. "Size and Sex: Evidence for the Domestication of Cattle in the Near East." Book Section.
955 In *People and Culture in Change: Preceedings of the 2nd Symposium on Upper Paleolithic, Mesolithic,
956 and Mesolithic Populations of Europe and the Mediterranean Basin*, edited by Israel Hershkovitz, 77–109.
957 BAR International Series. Oxford: Archaeopress.
- 958 Helmer, Daniel, Lionel Gourichon, Hervé Monchot, Joris Peters, and María Saña Segui. 2005. "Identifying
959 Early Domestic Cattle from Pre-Pottery Neolithic Sites on the Middle Euphrates Using Sexual Dimor-
960 phism." Book Section. In *First Steps of Animal Domestication: New Archaeozoological Approaches*,

- 961 edited by Jean-Denis Vigne, Joris Peters, and Daniel Helmer, 86–95. Proceedings of the 9th Conference
962 of the International Council of Archaeozoology, Durham, August 2002. Oxford: Oxbow Books.
- 963 Jasra, A., C. C. Holmes, and D. A. Stephens. 2005. “Markov Chain Monte Carlo Methods and the Label
964 Switching Problem in Bayesian Mixture Modeling.” Journal Article. *Statistical Science* 20 (1): 50–67.
965 <https://doi.org/10.1214/088342305000000016>.
- 966 Kabukcu, Ceren. 2017. “Woodland Vegetation History and Human Impacts in South-Central Anatolia
967 16,000–6500 Cal BP: Anthracological Results from Five Prehistoric Sites in the Konya Plain.” Journal
968 Article. *Quaternary Science Reviews* 176: 85–100. <https://doi.org/10.1016/j.quascirev.2017.10.001>.
- 969 Karul, Necmi. 2019. “Early Farmers in Northwestern Anatolia in the Seventh Millennium.” Book Section.
970 In *Concluding the Neolithic: The Near East in the Second Half of the Seventh Millennium BC*, edited by
971 Arkadiusz Marciniak, 269–86. Atlanta, Georgia: Lockwood Press.
- 972 Kassambara, Alboukadel. 2020. *Ggpubr: ‘Ggplot2’ Based Publication Ready Plots.* <https://CRAN.R-project.org/package=ggpubr>.
- 974 Kay, Matthew. 2022. *ggdist: Visualizations of Distributions and Uncertainty.* <https://doi.org/10.5281/zenodo.3879620>.
- 976 Lebenzon, Roxanne, and Natalie D. Munro. 2022. “Body Size Variation in a Modern Collection of Mountain
977 Gazelle (*Gazella Gazella*) Skeletons.” Journal Article. *Journal of Archaeological Science: Reports* 41.
978 <https://doi.org/10.1016/j.jasrep.2021.103285>.
- 979 Lyman, R. Lee. 2008. *Quantitative Paleozoology.* Book. Cambridge Manuals in Archaeology. Cambridge:
980 Cambridge University Press.
- 981 Madgwick, Richard, and Jacqui Mulville. 2015. “Reconstructing Depositional Histories Through Bone
982 Taphonomy: Extending the Potential of Faunal Data.” Journal Article. *Journal of Archaeological Science*
983 53: 255–63. <https://doi.org/10.1016/j.jas.2014.10.015>.
- 984 Manning, Katie, Adrian Timpson, Stephen Shennan, and Enrico Crema. 2015. “Size Reduction in Early
985 European Domestic Cattle Relates to Intensification of Neolithic Herding Strategies.” Journal Article.
986 *PLoS One* 10 (12): e0141873. <https://doi.org/10.1371/journal.pone.0141873>.
- 987 Marom, Nimrod, and Guy Bar-Oz. 2013. “The Prey Pathway: A Regional History of Cattle (*Bos Taurus*)
988 and Pig (*Sus Scrofa*) Domestication in the Northern Jordan Valley, Israel.” Journal Article. *PLoS One*
989 8 (2): e55958. <https://doi.org/10.1371/journal.pone.0055958.g001>.
- 990 Martín, Patricia, and Ricardo García-González. 2015. “Identifying Sheep (*Ovis Aries*) Fetal Remains in
991 Archaeological Contexts.” Journal Article. *Journal of Archaeological Science* 64: 77–87. <https://doi.org/10.1016/j.jas.2015.10.003>.
- 993 Marwick, Ben. 2017. “Computational Reproducibility in Archaeological Research: Basic Principles and a

- 994 Case Study of Their Implementation.” Journal Article. *Journal of Archaeological Method and Theory* 24
995 (2): 424–50. <https://doi.org/10.1007/s10816-015-9272-9>.
- 996 Marwick, Ben, and Suzanne E. Pilaar Birch. 2018. “A Standard for the Scholarly Citation of Archaeological
997 Data as an Incentive to Data Sharing.” Journal Article. *Advances in Archaeological Practice* 6 (02):
998 125–43. <https://doi.org/10.1017/aap.2018.3>.
- 999 McCarthy, Michael A., and P. I. P. Masters. 2005. “Profiting from Prior Information in Bayesian Analyses
1000 of Ecological Data.” Journal Article. *Journal of Applied Ecology* 42 (6): 1012–19. <https://doi.org/10.1111/j.1365-2664.2005.01101.x>.
- 1002 McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. Book.
1003 Texts in Statistical Science. Boca Raton, FL: CRC Press.
- 1004 McGrory, S., E. M. Svensson, A. Götherström, J. Mulville, A. J. Powell, M. J. Collins, and T. P. O’Connor.
1005 2012. “A Novel Method for Integrated Age and Sex Determination from Archaeological Cattle Mandibles.”
1006 Journal Article. *Journal of Archaeological Science* 39 (10): 3324–30. <https://doi.org/10.1016/j.jas.2012.05.021>.
- 1008 Meadow, Richard. 1999. “The Use of Size Index Scaling Techniques for Research on Archaeozoological
1009 Collections from the Middle East.” Book Section. In *Historia Animalium Ex Ossibus: Beiträge Zur
1010 Paläoanatomie, Archäologie, Ägyptologie, Ethnologie Und Geschichte Der Tiermedizin*, edited by Cornelia
1011 Becker, Henriette Manhart, Joris Peters, and Jörg Schibler, 285–300. Leidorf: Verlag Marie.
- 1012 Meier, Jacqueline. 2020. “The Contextual Taphonomy of Middens at Neolithic Kfar HaHoresh.” Journal
1013 Article. *Journal of Archaeological Science: Reports* 33. <https://doi.org/10.1016/j.jasrep.2020.102531>.
- 1014 Milner, J. M., E. B. Nilsen, and H. P. Andreassen. 2007. “Demographic Side Effects of Selective Hunting in
1015 Ungulates and Carnivores.” Journal Article. *Conserv Biol* 21 (1): 36–47. <https://doi.org/10.1111/j.1523-1739.2006.00591.x>.
- 1017 Monchot, Hervé, and Jacques Léchelle. 2002. “Statistical Nonparametric Methods for the Study of Fossil
1018 Populations.” Journal Article. *Paleobiology* 28 (1): 55–69. [https://doi.org/http://dx.doi.org/10.1666/0094-8373\(2002\)028%3C0055:SNMFTS%3E2.0.CO;2](https://doi.org/http://dx.doi.org/10.1666/0094-8373(2002)028%3C0055:SNMFTS%3E2.0.CO;2).
- 1020 Otárola-Castillo, Erik, and Melissa G. Torquato. 2018. “Bayesian Statistics in Archaeology.” Journal Article.
1021 *Annual Review of Anthropology* 47 (1). <https://doi.org/10.1146/annurev-anthro-102317-045834>.
- 1022 Otárola-Castillo, Erik, Melissa G. Torquato, Jesse Wolfhagen, Matthew E. Hill, and Caitlin E. Buck. 2022.
1023 “Beyond Chronology, Using Bayesian Inference to Evaluate Hypotheses in Archaeology.” Journal Article.
1024 *Advances in Archaeological Practice*.
- 1025 Özdogan, Mehmet. 2011. “Archaeological Evidence on the Westward Expansion of Farming Communities
1026 from Eastern Anatolia to the Aegean and the Balkans.” Journal Article. *Current Anthropology* 52 (S4):

- 1027 S415–30. <https://doi.org/10.1086/658895>.
- 1028 ———. 2019. “Early Farmers in Northwestern Turkey: What Is New?” Book Section. In *Concluding*
1029 *the Neolithic: The Near East in the Second Half of the Seventh Millennium BC*, edited by Arkadiusz
1030 Marciniak, 307–28. Atlanta, Georgia: Lockwood Press.
- 1031 Payne, Sebastian. 1969. “A Metrical Distinction Between Sheep and Goat Metacarpals.” Book Section. In
1032 *The Domestication and Exploitation of Plants and Animals*, edited by Peter J. Ucko and Geoffrey W.
1033 Dimbleby, 295–305. London: Duckworth.
- 1034 ———. 1973. “Kill-Off Patterns in Sheep and Goats: The Mandibles from Aşvan Kale.” Journal Article.
1035 *Anatolian Studies* 23: 281–303.
- 1036 Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R*
1037 *Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- 1038 Pérez-Barbería, F. Javier, Iain J. Gordon, and M. Pagel. 2002. “The Origins of Sexual Dimorphism in Body
1039 Size in Ungulates.” Journal Article. *Evolution* 56 (6): 1276–85.
- 1040 Peters, Joris, Angela von den Driesch, and Daniel Helmer. 2005. “The Upper Euphrates-Tigris Basin:
1041 Cradle of Agro-Pastoralism?” Book Section. In *The First Steps of Animal Domestication*, edited by
1042 Jean-Denis Vigne, Joris Peters, and Daniel Helmer, 96–124. Proceedings of the 9th Conference of the
1043 International Council of Archaeozoology, Durham, August 2002. Oxford: Oxbow Books.
- 1044 Popkin, Peter R. W., Polydora Baker, Fay Worley, Sebastian Payne, and Andy Hammon. 2012. “The Sheep
1045 Project (1): Determining Skeletal Growth, Timing of Epiphyseal Fusion and Morphometric Variation
1046 in Unimproved Shetland Sheep of Known Age, Sex, Castration Status and Nutrition.” Journal Article.
1047 *Journal of Archaeological Science* 39 (6): 1775–92. <https://doi.org/10.1016/j.jas.2012.01.018>.
- 1048 Post, Diane M., Trent S. Armbrust, Eva A. Horne, and Jacob R. Goheen. 2001. “Sexual Segregation
1049 Results in Differences in Content and Quality of Bison (*Bos Bison*) Diets.” Journal Article. *Journal of*
1050 *Mammalogy* 82 (2): 407–13.
- 1051 Pozo, Jose M, Valenzuela-Lamas, Silvia, Trentacoste, Angela, Nieto-Espinet, Ariadna, Guimarães Chiarelli,
1052 and Silvia. 2021. “Zoolog: Zooarchaeological Analysis with Log-Ratios.” *Journal of Statistical Software*
1053 In preparation. <https://josempozo.github.io/zooolog/>.
- 1054 Proaktor, G., T. Coulson, and E. J. Milner-Gulland. 2007. “Evolutionary Responses to Harvesting in
1055 Ungulates.” Journal Article. *J Anim Ecol* 76 (4): 669–78. <https://doi.org/10.1111/j.1365-2656.2007.01244.x>.
- 1057 R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R
1058 Foundation for Statistical Computing. <https://www.R-project.org/>.
- 1059 Redding, Richard W. 1984. “Theoretical Determinants of a Herder’s Decisions: Modeling Variation in the

- 1060 Sheep/Goat Ratio.” Book Section. In *Animals and Archaeology*, edited by Juliet Clutton-Brock and
1061 Caroline Grigson, 3. Early Herders and their Flocks:223–41. BAR International Series. Oxford: British
1062 Archaeological Reports.
- 1063 Roodenberg, Jacob. 2012a. “Change in Food Production and Its Impact on an Early 6th Millennium
1064 Community in Northwest Anatolia. The Example of Ilipinar.” Journal Article. *Praehistorische Zeitschrift*
1065 87 (2). <https://doi.org/10.1515/pz-2012-0015>.
- 1066 ———. 2012b. “Ilipinar: A Neolithic Settlement in the Eastern Marmara Region.” Book Section. In
1067 *The Oxford Handbook of Ancient Anatolia (10,000-323 BCE)*, edited by Gregory McMahon and Sharon
1068 Steadman. <https://doi.org/10.1093/oxfordhb/9780195376142.013.0044>.
- 1069 Roodenberg, Jacob, A. van As, L. Jacobs, and M. H. Wijnen. 2003. “Early Settlement in the Plain of
1070 Yenisehir (NW Anatolia): The Basal Occupation Layers at Mentese.” Journal Article. *Anatolica* XXIX
1071 (1): 18–58.
- 1072 RStudio Team. 2022. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC.
1073 <http://www.rstudio.com/>.
- 1074 Ruckstuhl, K. E. 2007. “Sexual Segregation in Vertebrates: Proximate and Ultimate Causes.” Journal
1075 Article. *Integr Comp Biol* 47 (2): 245–57. <https://doi.org/10.1093/icb/icm030>.
- 1076 Ruckstuhl, K. E., and P. Neuhaus. 2002. “Sexual Segregation in Ungulates: A Comparative Test of
1077 Three Hypotheses.” Journal Article. *Biological Reviews* 77 (1): 77–96. <https://doi.org/10.1017/s1464793101005814>.
- 1079 Said, Sonia, Vincent Tolon, Serge Brandt, and Eric Baubet. 2011. “Sex Effect on Habitat Selection in
1080 Response to Hunting Disturbance: The Study of Wild Boar.” Journal Article. *European Journal of
1081 Wildlife Research* 58 (1): 107–15. <https://doi.org/10.1007/s10344-011-0548-4>.
- 1082 Smith, A. F. M., and Alan E. Gelfand. 1992. “Bayesian Statistics Without Tears: A Sampling-Resampling
1083 Perspective.” Journal Article. *The American Statistician* 46 (2): 84–88.
- 1084 South, Andy. 2017a. *Rnaturalearth: World Map Data from Natural Earth*. <https://CRAN.R-project.org/package=rnaturalearth>.
- 1086 ———. 2017b. *Rnaturalearthdata: World Vector Map Data from Natural Earth Used in 'Rnaturalearth'*.
1087 <https://CRAN.R-project.org/package=rnaturalearthdata>.
- 1088 Speth, John D. 2013. “Thoughts about Hunting: Some Things We Know and Some Things We Don’t Know.”
1089 Journal Article. *Quaternary International* 297: 176–85. <https://doi.org/10.1016/j.quaint.2012.12.005>.
- 1090 Stan Development Team. 2021. “RStan: The R Interface to Stan.” <https://mc-stan.org/>.
- 1091 Stiner, Mary C. 1990. “The Use of Mortality Patterns in Archaeological Studies of Hominid Predatory
1092 Adaptations.” Journal Article. *Journal of Anthropological Archaeology* 9: 305–51.

- 1093 Tchernov, Eitan, and Liora Kolska Horwitz. 1991. "Body Size Diminution Under Domestication: Uncon-
1094 scious Selection in Primeval Domesticates." Journal Article. *Journal of Anthropological Archaeology* 10:
1095 54–75.
- 1096 Team, Stan Development. 2022. "Stan Modeling Language Users Guide and Reference Manual, Version
1097 2.29." Generic. <https://mc-stan.org>.
- 1098 Thissen, Laurens, Hadi Özbal, Ayla Türkekul Büyik, Fokke Gerritsen, and Rana Özbal. 2010. "The Land
1099 of Milk? Approaching Dietary Preferences of Late Neolithic Communities in NW Anatolia." Journal
1100 Article. *Leiden Journal of Pottery Studies* 26: 157–72.
- 1101 Twiss, Katheryn C. 2019. *The Archaeology of Food*. Book. <https://doi.org/10.1017/9781108670159>.
- 1102 Uerpmann, Margarethe, and Hans-Peter Uerpmann. 1994. "Animal Bones." Book Section. In *Qala'at Al-*
1103 *Bahrain*, edited by Flemming Hojlund and H. Hellmuth Andersen, 1: The Northern City Wall and the
1104 Islamic Fortress:417–54. Jutland Archaeological Society Publications. Aarhus: Aarhus University Press.
- 1105 Urbanek, Simon, and Jeffrey Horner. 2022. *Cairo: R Graphics Device Using Cairo Graphics Library for*
1106 *Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11*
1107 *and Win32) Output*. <https://CRAN.R-project.org/package=Cairo>.
- 1108 Watkins, Trevor. 1996. "Excavations at Pinarbaşı: The Early Stages." Book Section. In *On the Surface:*
1109 *Çatalhöyük 1993-1995*, edited by Ian Hodder, 47–57. Cambridge: McDonald Institute for Archaeological
1110 Research.
- 1111 Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- 1112 Wolfhagen, Jesse. 2020. "Re-Examining the Use of the LSI Technique in Zooarchaeology." Journal Article.
1113 *Journal of Archaeological Science* 123: 105254. <https://doi.org/10.1016/j.jas.2020.105254>.
- 1114 Wright, Elizabeth, and Sarah Viner-Daniels. 2015. "Geographical Variation in the Size and Shape of the
1115 European Aurochs (*Bos Primigenius*).". Journal Article. *Journal of Archaeological Science* 54: 8–22.
1116 <https://doi.org/10.1016/j.jas.2014.11.021>.
- 1117 Zeder, Melinda A. 2012. "Pathways to Animal Domestication." Book Section. In *Biodiversity in Agriculture:*
1118 *Domestication, Evolution, and Sustainability*, edited by P. Gepts, T. R. Famula, and R. L. Bettinger,
1119 227–59. Cambridge: Cambridge University Press.
- 1120 Zeder, Melinda A., and Brian Hesse. 2000. "The Initial Domestication of Goats (*Capra Hircus*) in the Zagros
1121 Mountains 10,000 Years Ago." Journal Article. *Science* 287 (5461): 2254–57.
- 1122 Zeder, Melinda A., and Ximena Lemoine. 2020. "A Method for Constructing Demographic Profiles in Sus
1123 *Scrofa* Using Logarithm Size Index Scaling." Journal Article. *Journal of Archaeological Science* 116.
1124 <https://doi.org/10.1016/j.jas.2020.105115>.

1126 **Appendix 1 (Supplemental Table 6)**

1127 ***Posterior Summary Tables for Overall and Site-Level Model Parameters: Simulated Assem-***
 1128 ***blages***

Table 6: Posterior Fit Summaries for Model Parameters (Single Assemblage Simulation)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
Single Assemblage Model θ_1	0.15	0.15	0.04	0.10	0.22	1	3396	2582
Single Assemblage Model θ_2	0.52	0.53	0.09	0.37	0.66	1	2806	3088
Single Assemblage Model θ_3	0.32	0.32	0.09	0.18	0.49	1	2419	2169
Single Assemblage Model μ_1	-0.10	-0.10	0.02	-0.13	-0.07	1	1260	1630
Single Assemblage Model μ_2	-0.08	-0.08	0.02	-0.10	-0.05	1	1118	1421
Single Assemblage Model μ_3	-0.01	-0.01	0.02	-0.04	0.02	1	1299	1596
Single Assemblage Model σ_1	0.05	0.05	0.00	0.04	0.06	1	3289	1902
Single Assemblage Model σ_2	0.04	0.04	0.00	0.04	0.05	1	3076	1396
Single Assemblage Model σ_3	0.04	0.04	0.00	0.04	0.05	1	3719	2828

Table 7: Posterior Fit Summaries for Model Parameters (Multisite Simulation)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
Multisite Model θ_1	0.15	0.15	0.02	0.12	0.19	1	3114	2854
Multisite Model θ_2	0.47	0.47	0.06	0.37	0.56	1	2214	2365
Multisite Model θ_3	0.38	0.38	0.06	0.28	0.48	1	2181	2157
Multisite Model μ_1	-0.10	-0.10	0.04	-0.17	-0.03	1	1255	1880
Multisite Model μ_2	-0.08	-0.08	0.04	-0.16	-0.01	1	1241	1873
Multisite Model μ_3	-0.01	-0.01	0.05	-0.08	0.07	1	1269	1967
Multisite Model σ_1	0.04	0.04	0.01	0.03	0.06	1	3001	3049
Multisite Model σ_2	0.04	0.04	0.00	0.03	0.05	1	2825	2754
Multisite Model σ_3	0.04	0.04	0.01	0.03	0.04	1	2301	2360

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
							(Bulk)	(Tail)
Site 1 θ_1	0.16	0.16	0.04	0.11	0.23	1.00	4225	3370
Site 2 θ_1	0.48	0.49	0.10	0.31	0.65	1.00	3867	3364
Site 3 θ_1	0.35	0.35	0.11	0.18	0.53	1.00	3557	3049
Site 4 θ_1	0.18	0.17	0.05	0.11	0.26	1.00	4911	3396
Site 5 θ_1	0.44	0.45	0.11	0.25	0.62	1.00	4126	3516

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
						(Bulk)	(Tail)	
Site 6 θ_1	0.38	0.38	0.12	0.19	0.58	1.00	4418	3526
Site 7 θ_1	0.14	0.14	0.04	0.08	0.20	1.00	3094	3412
Site 8 θ_1	0.32	0.33	0.10	0.15	0.48	1.00	2366	3056
Site 9 θ_1	0.54	0.53	0.10	0.37	0.72	1.00	2768	2908
Site 10 θ_1	0.18	0.18	0.04	0.12	0.26	1.00	3444	3389
Site 11 θ_1	0.45	0.46	0.10	0.29	0.61	1.00	3329	3523
Site 12 θ_1	0.36	0.36	0.10	0.20	0.54	1.00	3952	3434
Site 13 θ_1	0.15	0.15	0.04	0.09	0.21	1.00	2339	3318
Site 14 θ_1	0.47	0.47	0.08	0.34	0.59	1.00	3478	3639
Site 15 θ_1	0.38	0.38	0.07	0.27	0.50	1.00	3327	3511
Site 1 θ_2	0.16	0.16	0.04	0.10	0.23	1.00	4332	3135
Site 2 θ_2	0.59	0.60	0.10	0.42	0.74	1.00	2952	2806
Site 3 θ_2	0.25	0.24	0.10	0.10	0.43	1.00	3191	3035
Site 4 θ_2	0.16	0.16	0.04	0.10	0.24	1.00	3937	3141
Site 5 θ_2	0.54	0.55	0.12	0.35	0.73	1.00	4021	2986
Site 6 θ_2	0.29	0.29	0.12	0.11	0.50	1.00	3902	2734
Site 7 θ_2	0.16	0.16	0.04	0.10	0.22	1.00	4011	3387
Site 8 θ_2	0.52	0.53	0.10	0.35	0.69	1.00	3043	3149
Site 9 θ_2	0.32	0.31	0.11	0.15	0.50	1.00	3437	3345
Site 10 θ_2	0.19	0.19	0.04	0.13	0.28	1.00	3052	2587
Site 11 θ_2	0.56	0.57	0.09	0.40	0.71	1.00	3471	2734
Site 12 θ_2	0.24	0.24	0.09	0.11	0.41	1.00	4423	3006
Site 13 θ_2	0.26	0.25	0.07	0.16	0.38	1.00	1914	2047
Site 14 θ_2	0.61	0.62	0.08	0.47	0.73	1.00	1837	2924
Site 15 θ_2	0.13	0.12	0.06	0.05	0.24	1.00	2704	1611

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
						(Bulk)	(Tail)	
Site 1 θ_3	0.14	0.14	0.04	0.08	0.20	1.00	3193	3332
Site 2 θ_3	0.46	0.47	0.11	0.29	0.63	1.00	4162	3441
Site 3 θ_3	0.40	0.40	0.11	0.22	0.59	1.00	4382	3052
Site 4 θ_3	0.12	0.12	0.04	0.05	0.18	1.00	2456	2814
Site 5 θ_3	0.37	0.37	0.12	0.16	0.57	1.00	4555	3309
Site 6 θ_3	0.51	0.51	0.13	0.29	0.73	1.00	4201	3042
Site 7 θ_3	0.14	0.13	0.04	0.08	0.20	1.00	2461	3307
Site 8 θ_3	0.32	0.32	0.11	0.14	0.52	1.00	3808	3243
Site 9 θ_3	0.54	0.54	0.12	0.34	0.74	1.00	3998	3024
Site 10 θ_3	0.14	0.14	0.04	0.08	0.21	1.00	2446	2624
Site 11 θ_3	0.38	0.38	0.10	0.21	0.55	1.00	3113	3304
Site 12 θ_3	0.48	0.48	0.11	0.30	0.65	1.00	3346	3389
Site 13 θ_3	0.14	0.14	0.04	0.08	0.20	1.00	2455	3303
Site 14 θ_3	0.30	0.30	0.08	0.18	0.43	1.00	3738	3207
Site 15 θ_3	0.56	0.56	0.08	0.43	0.69	1.00	3779	3483
Site 1 μ_1	-0.08	-0.08	0.02	-0.12	-0.05	1.00	2155	2604
Site 2 μ_1	-0.08	-0.08	0.03	-0.13	-0.04	1.00	2757	3373
Site 3 μ_1	0.08	0.08	0.02	0.04	0.12	1.00	2101	2473
Site 4 μ_1	-0.30	-0.30	0.02	-0.34	-0.27	1.00	2214	2293
Site 5 μ_1	-0.10	-0.10	0.02	-0.13	-0.07	1.00	1935	3160
Site 6 μ_1	-0.12	-0.12	0.02	-0.15	-0.09	1.00	2085	2822
Site 7 μ_1	-0.09	-0.09	0.02	-0.13	-0.05	1.00	3029	3448
Site 8 μ_1	0.07	0.07	0.02	0.04	0.10	1.00	2411	3197
Site 9 μ_1	-0.34	-0.34	0.02	-0.38	-0.31	1.00	2112	3039
Site 10 μ_1	-0.11	-0.10	0.02	-0.13	-0.08	1.00	2149	2296

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
						(Bulk)	(Tail)	
Site 11 μ_1	-0.08	-0.07	0.02	-0.11	-0.04	1.00	2378	2817
Site 12 μ_1	-0.09	-0.09	0.03	-0.14	-0.04	1.00	2905	3399
Site 13 μ_1	0.09	0.09	0.02	0.05	0.13	1.00	2297	2715
Site 14 μ_1	-0.32	-0.32	0.03	-0.36	-0.28	1.00	2288	2743
Site 15 μ_1	-0.09	-0.09	0.02	-0.13	-0.06	1.00	2402	3192
Site 1 μ_2	-0.07	-0.07	0.02	-0.11	-0.03	1.00	2059	2621
Site 2 μ_2	-0.07	-0.07	0.03	-0.12	-0.02	1.00	2731	3004
Site 3 μ_2	0.10	0.10	0.02	0.06	0.14	1.00	2096	2395
Site 4 μ_2	-0.29	-0.29	0.02	-0.32	-0.26	1.00	2165	2566
Site 5 μ_2	-0.09	-0.09	0.02	-0.11	-0.06	1.00	1950	2783
Site 6 μ_2	-0.11	-0.11	0.02	-0.13	-0.08	1.00	2198	2729
Site 7 μ_2	-0.07	-0.07	0.02	-0.11	-0.04	1.00	2992	2851
Site 8 μ_2	0.09	0.09	0.02	0.06	0.12	1.00	2293	2710
Site 9 μ_2	-0.33	-0.33	0.02	-0.36	-0.30	1.00	2043	2612
Site 10 μ_2	-0.09	-0.09	0.02	-0.12	-0.06	1.00	2127	2479
Site 11 μ_2	-0.06	-0.06	0.02	-0.09	-0.03	1.00	2406	2689
Site 12 μ_2	-0.07	-0.07	0.03	-0.12	-0.02	1.00	2895	3436
Site 13 μ_2	0.10	0.11	0.02	0.07	0.14	1.00	2132	2339
Site 14 μ_2	-0.30	-0.30	0.02	-0.35	-0.27	1.00	2321	2692
Site 15 μ_2	-0.07	-0.07	0.02	-0.11	-0.04	1.00	2337	2665
Site 1 μ_3	-0.01	-0.01	0.02	-0.05	0.03	1.00	2631	3343
Site 2 μ_3	0.02	0.02	0.03	-0.03	0.08	1.00	2652	3410
Site 3 μ_3	0.19	0.20	0.03	0.15	0.23	1.00	1525	2015
Site 4 μ_3	-0.23	-0.23	0.02	-0.26	-0.19	1.00	2833	3218
Site 5 μ_3	0.17	0.17	0.02	0.14	0.20	1.00	2220	2845

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size (Bulk)	Size (Tail)
Site 6 μ_3	-0.06	-0.06	0.03	-0.10	-0.01	1.00	2601	3043
Site 7 μ_3	-0.01	-0.01	0.04	-0.06	0.07	1.00	3023	2799
Site 8 μ_3	0.15	0.14	0.03	0.10	0.20	1.00	2208	2737
Site 9 μ_3	-0.27	-0.27	0.03	-0.31	-0.22	1.00	2726	3165
Site 10 μ_3	0.15	0.16	0.06	0.01	0.22	1.02	271	64
Site 11 μ_3	-0.01	-0.01	0.03	-0.04	0.03	1.00	2642	2850
Site 12 μ_3	0.00	0.00	0.03	-0.05	0.05	1.00	2833	3192
Site 13 μ_3	0.15	0.15	0.02	0.12	0.19	1.00	2362	2974
Site 14 μ_3	-0.24	-0.24	0.02	-0.27	-0.20	1.00	2158	2820
Site 15 μ_3	0.18	0.17	0.02	0.14	0.21	1.00	2284	2918
Site 1 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3607	3548
Site 2 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3357	3162
Site 3 σ_1	0.05	0.04	0.01	0.03	0.07	1.00	3591	3113
Site 4 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3588	3047
Site 5 σ_1	0.04	0.04	0.01	0.03	0.07	1.00	3804	3201
Site 6 σ_1	0.04	0.04	0.01	0.02	0.06	1.00	2916	2810
Site 7 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3030	1612
Site 8 σ_1	0.04	0.04	0.01	0.02	0.06	1.00	3200	3509
Site 9 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3196	2984
Site 10 σ_1	0.04	0.04	0.02	0.02	0.06	1.01	662	230
Site 11 σ_1	0.04	0.04	0.01	0.03	0.06	1.00	3645	3221
Site 12 σ_1	0.05	0.04	0.01	0.03	0.07	1.00	3706	3412
Site 13 σ_1	0.05	0.04	0.01	0.03	0.07	1.00	3649	3251
Site 14 σ_1	0.05	0.04	0.01	0.03	0.07	1.00	3613	2926
Site 15 σ_1	0.05	0.05	0.02	0.03	0.08	1.00	3705	3483

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
						(Bulk)	(Tail)	
Site 1 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	2727	3366
Site 2 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	3008	2851
Site 3 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	2816	2582
Site 4 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	3176	3171
Site 5 σ_2	0.04	0.04	0.01	0.02	0.05	1.00	2482	3184
Site 6 σ_2	0.03	0.03	0.01	0.02	0.05	1.00	2088	3042
Site 7 σ_2	0.04	0.04	0.01	0.02	0.05	1.00	3556	3397
Site 8 σ_2	0.04	0.04	0.01	0.02	0.05	1.00	3066	3520
Site 9 σ_2	0.04	0.04	0.01	0.03	0.05	1.00	3211	3438
Site 10 σ_2	0.03	0.03	0.01	0.02	0.05	1.00	2573	3231
Site 11 σ_2	0.04	0.04	0.01	0.03	0.05	1.00	3496	3505
Site 12 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	3377	3163
Site 13 σ_2	0.04	0.04	0.01	0.03	0.05	1.00	3392	2948
Site 14 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	2560	3361
Site 15 σ_2	0.04	0.04	0.01	0.03	0.06	1.00	2938	3014
Site 1 σ_3	0.04	0.03	0.01	0.02	0.05	1.00	2740	2432
Site 2 σ_3	0.04	0.04	0.01	0.03	0.06	1.00	2436	2489
Site 3 σ_3	0.04	0.04	0.01	0.03	0.06	1.00	1655	1877
Site 4 σ_3	0.04	0.03	0.01	0.02	0.05	1.00	3031	3228
Site 5 σ_3	0.03	0.03	0.01	0.02	0.04	1.00	2184	2835
Site 6 σ_3	0.04	0.04	0.01	0.02	0.05	1.00	3174	2907
Site 7 σ_3	0.04	0.04	0.01	0.02	0.05	1.00	2940	2412
Site 8 σ_3	0.04	0.04	0.01	0.03	0.06	1.00	2468	2946
Site 9 σ_3	0.04	0.03	0.01	0.02	0.05	1.00	2767	3028
Site 10 σ_3	0.04	0.03	0.02	0.02	0.06	1.00	1476	655

Table 8: Posterior Fit Summaries for Site-Level Model Parameters (Multisite Simulation) (*continued*)

Parameter	Posterior	Posterior	Posterior	5%	95%	\hat{R}	Effective	Effective
	Mean	Median	SD	Quantile	Quantile		Sample	Sample
							Size	Size
Site 11 σ_3	0.04	0.04	0.01	0.03	0.05	1.00	2986	2739
Site 12 σ_3	0.04	0.04	0.01	0.03	0.05	1.00	3129	3187
Site 13 σ_3	0.04	0.04	0.01	0.02	0.05	1.00	3308	3409
Site 14 σ_3	0.03	0.03	0.01	0.02	0.05	1.00	2740	3035
Site 15 σ_3	0.03	0.03	0.01	0.02	0.04	1.00	2047	2826

¹¹²⁹ **Appendix 2 (Supplemental Table 7)**

¹¹³⁰ ***Posterior Summary Tables for Overall and Site-Level Model Parameters: Archaeological Case***
¹¹³¹ ***Studies***

Table 9: Posterior Fit Summaries for Model Parameters (Pinarbaşı B Sheep)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
Pinarbaşı B Sheep θ_1	0.89	0.89	0.04	0.82	0.95	1	3681	2888
Pinarbaşı B Sheep θ_2	0.07	0.07	0.04	0.02	0.14	1	2767	2389
Pinarbaşı B Sheep θ_3	0.04	0.03	0.03	0.00	0.10	1	2677	2439
Pinarbaşı B Sheep μ_1	-0.13	-0.13	0.02	-0.16	-0.09	1	1807	2232
Pinarbaşı B Sheep μ_2	-0.10	-0.09	0.02	-0.13	-0.06	1	1688	1904
Pinarbaşı B Sheep μ_3	-0.03	-0.03	0.02	-0.07	0.01	1	1796	2098
Pinarbaşı B Sheep σ_1	0.05	0.05	0.00	0.04	0.06	1	4425	2547
Pinarbaşı B Sheep σ_2	0.04	0.04	0.00	0.04	0.05	1	4487	2791
Pinarbaşı B Sheep σ_3	0.05	0.05	0.00	0.04	0.05	1	4665	2598

Table 10: Posterior Fit Summaries for Model Parameters (Northwest Anatolian Cattle)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
NW Anatolian Cattle θ_1	0.18	0.17	0.07	0.09	0.32	1	1173	1907
NW Anatolian Cattle θ_2	0.63	0.64	0.09	0.48	0.76	1	1340	2010
NW Anatolian Cattle θ_3	0.18	0.18	0.07	0.08	0.29	1	1324	1864
NW Anatolian Cattle μ_1	-0.24	-0.24	0.04	-0.29	-0.18	1	1584	1766
NW Anatolian Cattle μ_2	-0.20	-0.20	0.03	-0.25	-0.14	1	1475	1676
NW Anatolian Cattle μ_3	-0.07	-0.07	0.04	-0.13	0.00	1	1666	1714
NW Anatolian Cattle σ_1	0.07	0.06	0.01	0.05	0.09	1	2851	2949
NW Anatolian Cattle σ_2	0.04	0.04	0.01	0.03	0.06	1	2338	3419
NW Anatolian Cattle σ_3	0.04	0.04	0.01	0.03	0.06	1	1503	2263

Table 11: Posterior Fit Summaries for Site-Level Model Parameters (Northwest Anatolian Cattle)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
Barçın θ_1	0.34	0.34	0.04	0.27	0.41	1.00	3928	3135
Neolithic İlpınar θ_1	0.48	0.49	0.07	0.37	0.59	1.00	1566	2807
Menteşe θ_1	0.18	0.17	0.06	0.09	0.30	1.00	1457	2836
Chalcolithic İlpınar θ_1	0.13	0.13	0.02	0.10	0.17	1.00	3868	3520
Barçın θ_2	0.72	0.72	0.08	0.58	0.85	1.01	830	1861
Neolithic İlpınar θ_2	0.14	0.14	0.08	0.02	0.28	1.01	767	1604
Menteşe θ_2	0.24	0.24	0.06	0.15	0.35	1.00	5014	3080
Chalcolithic İlpınar θ_2	0.60	0.60	0.08	0.47	0.73	1.00	2866	3255
Barçın θ_3	0.16	0.15	0.06	0.06	0.27	1.00	1683	2551
Neolithic İlpınar θ_3	0.07	0.07	0.02	0.05	0.11	1.00	3012	2548
Menteşe θ_3	0.70	0.71	0.05	0.62	0.78	1.00	1518	2591
Chalcolithic İlpınar θ_3	0.23	0.22	0.05	0.15	0.31	1.00	1411	2527
Barçın μ_1	-0.39	-0.39	0.03	-0.44	-0.33	1.00	2723	3325
Neolithic İlpınar μ_1	-0.23	-0.23	0.02	-0.27	-0.19	1.00	2282	3368
Menteşe μ_1	-0.24	-0.23	0.03	-0.29	-0.19	1.00	2333	2590
Chalcolithic İlpınar μ_1	-0.27	-0.27	0.02	-0.31	-0.23	1.00	2033	2681
Barçın μ_2	-0.18	-0.18	0.02	-0.22	-0.14	1.00	1549	2733
Neolithic İlpınar μ_2	-0.21	-0.21	0.02	-0.24	-0.17	1.00	1966	2882
Menteşe μ_2	-0.20	-0.20	0.03	-0.24	-0.16	1.00	1932	2544
Chalcolithic İlpınar μ_2	-0.24	-0.25	0.02	-0.28	-0.21	1.00	1835	2489
Barçın μ_3	-0.04	-0.04	0.03	-0.09	0.01	1.00	1595	2728
Neolithic İlpınar μ_3	-0.10	-0.11	0.03	-0.16	-0.04	1.00	1083	1517
Menteşe μ_3	-0.06	-0.06	0.04	-0.11	0.00	1.00	1893	2940
Chalcolithic İlpınar μ_3	-0.10	-0.10	0.02	-0.14	-0.06	1.00	1701	2663
Barçın σ_1	0.06	0.06	0.01	0.04	0.09	1.00	2478	2761
Neolithic İlpınar σ_1	0.08	0.07	0.02	0.05	0.10	1.00	2690	3250
Menteşe σ_1	0.07	0.06	0.02	0.04	0.09	1.00	3212	3365
Chalcolithic İlpınar σ_1	0.08	0.07	0.02	0.05	0.11	1.00	2517	3155
Barçın σ_2	0.04	0.04	0.01	0.02	0.06	1.00	1684	2163
Neolithic İlpınar σ_2	0.07	0.07	0.01	0.05	0.08	1.00	2513	2926
Menteşe σ_2	0.04	0.04	0.01	0.03	0.06	1.00	2003	2447
Chalcolithic İlpınar σ_2	0.04	0.04	0.01	0.03	0.05	1.00	1148	2305
Barçın σ_3	0.04	0.04	0.01	0.03	0.06	1.00	1395	1141
Neolithic İlpınar σ_3	0.04	0.04	0.01	0.03	0.06	1.00	2274	2770
Menteşe σ_3	0.05	0.04	0.01	0.03	0.07	1.00	1870	2729
Chalcolithic İlpınar σ_3	0.04	0.04	0.01	0.03	0.06	1.00	1236	2031

1132 Appendix 3 (Supplemental Figures 1-4)

1133 *Traceplots of Posterior Distributions of Overall Model Parameters*

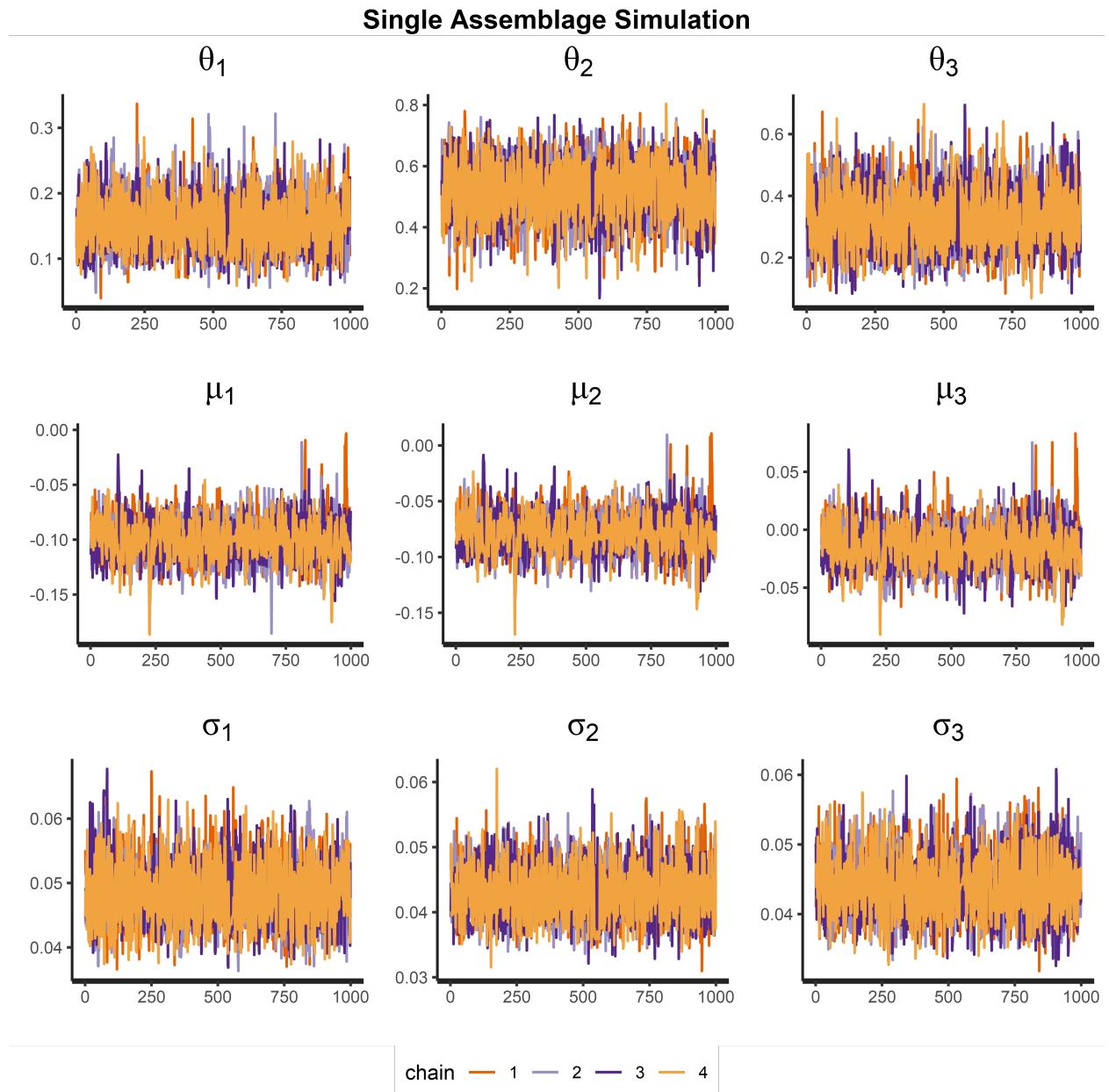


Figure 14: Traceplots of Model Parameters (Single Assemblage Simulation)

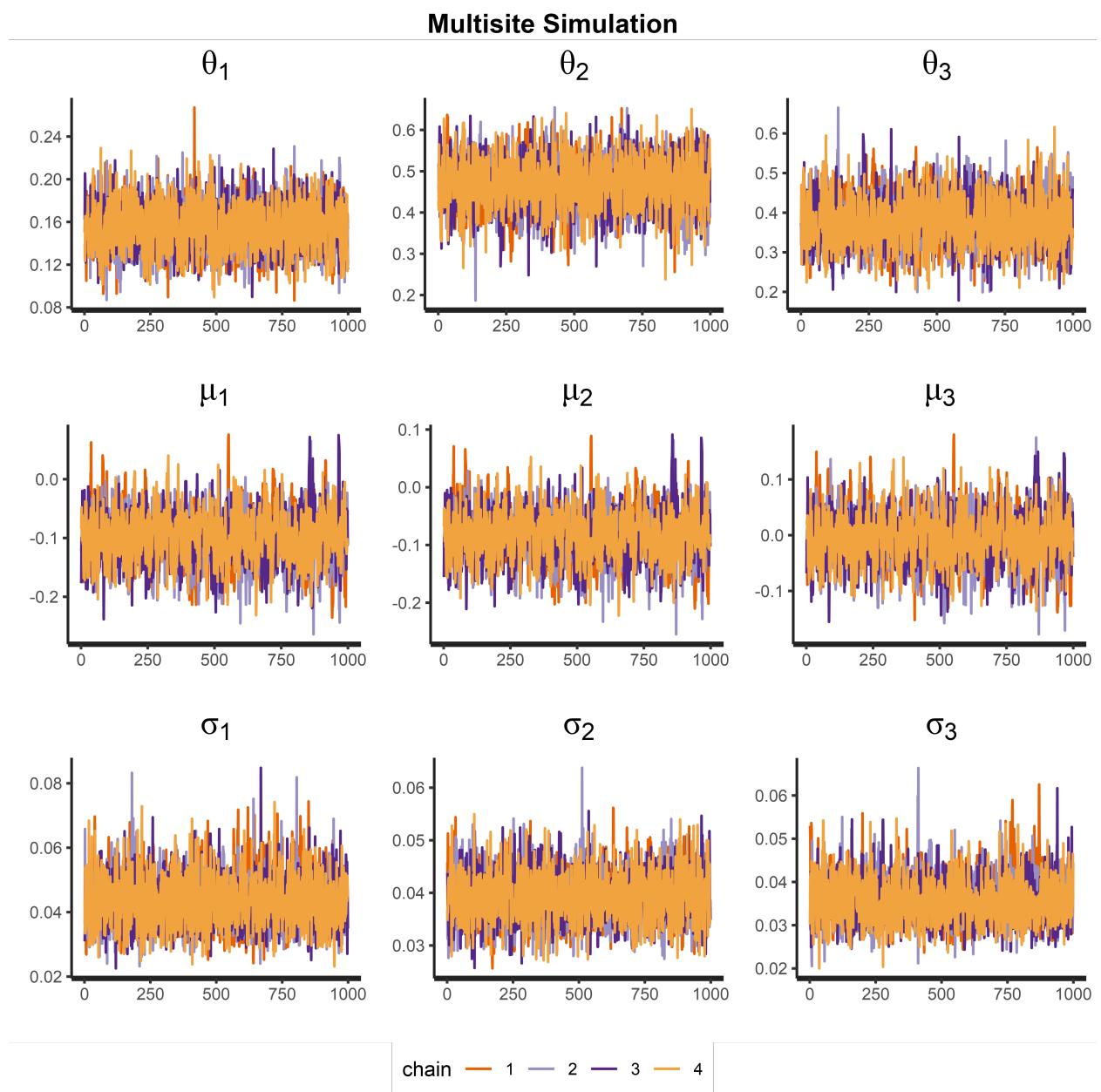


Figure 15: Traceplots of Model Parameters (Multisite Simulation)

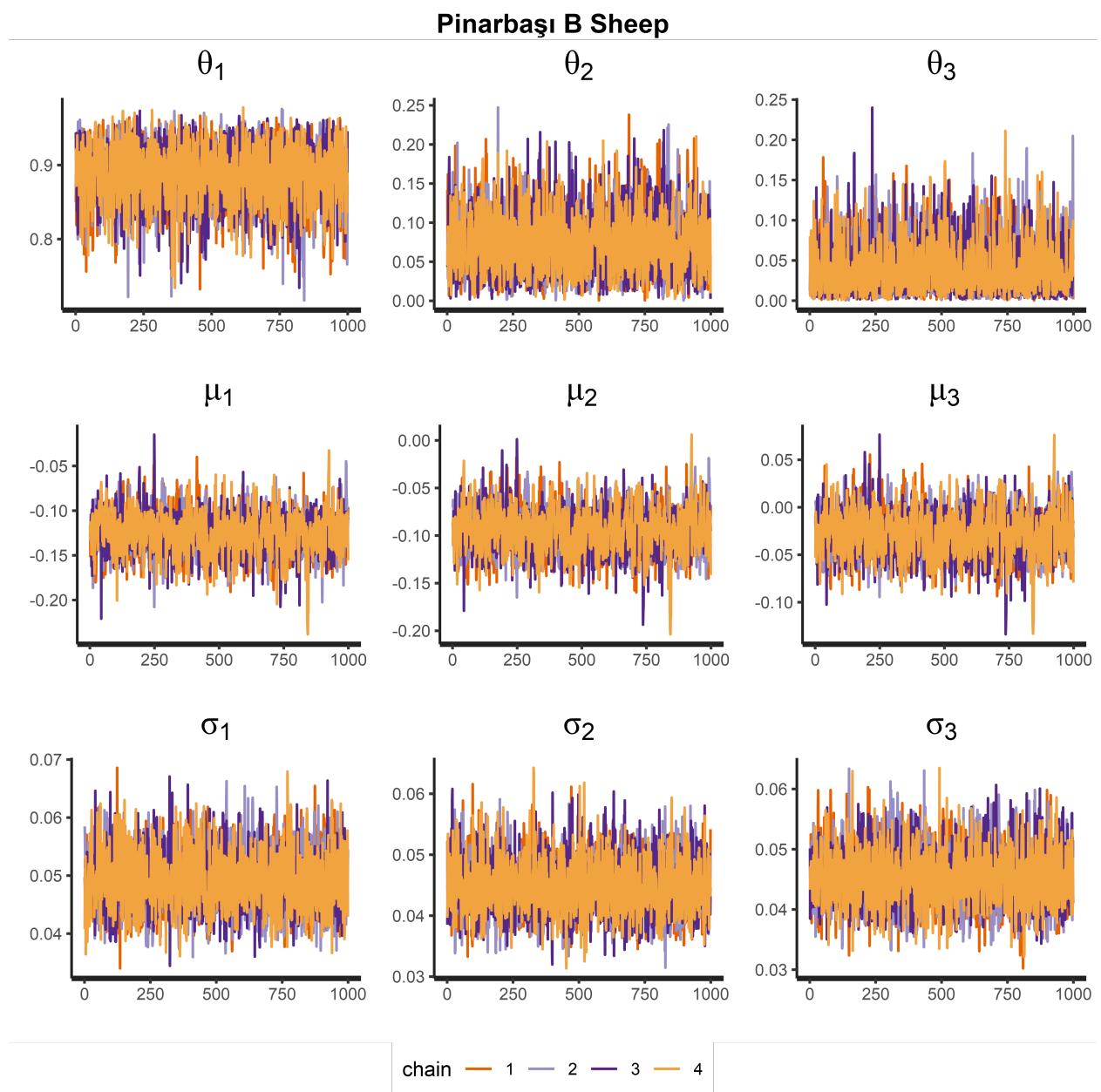


Figure 16: Traceplots of Model Parameters (Pınarbaşı B Sheep)

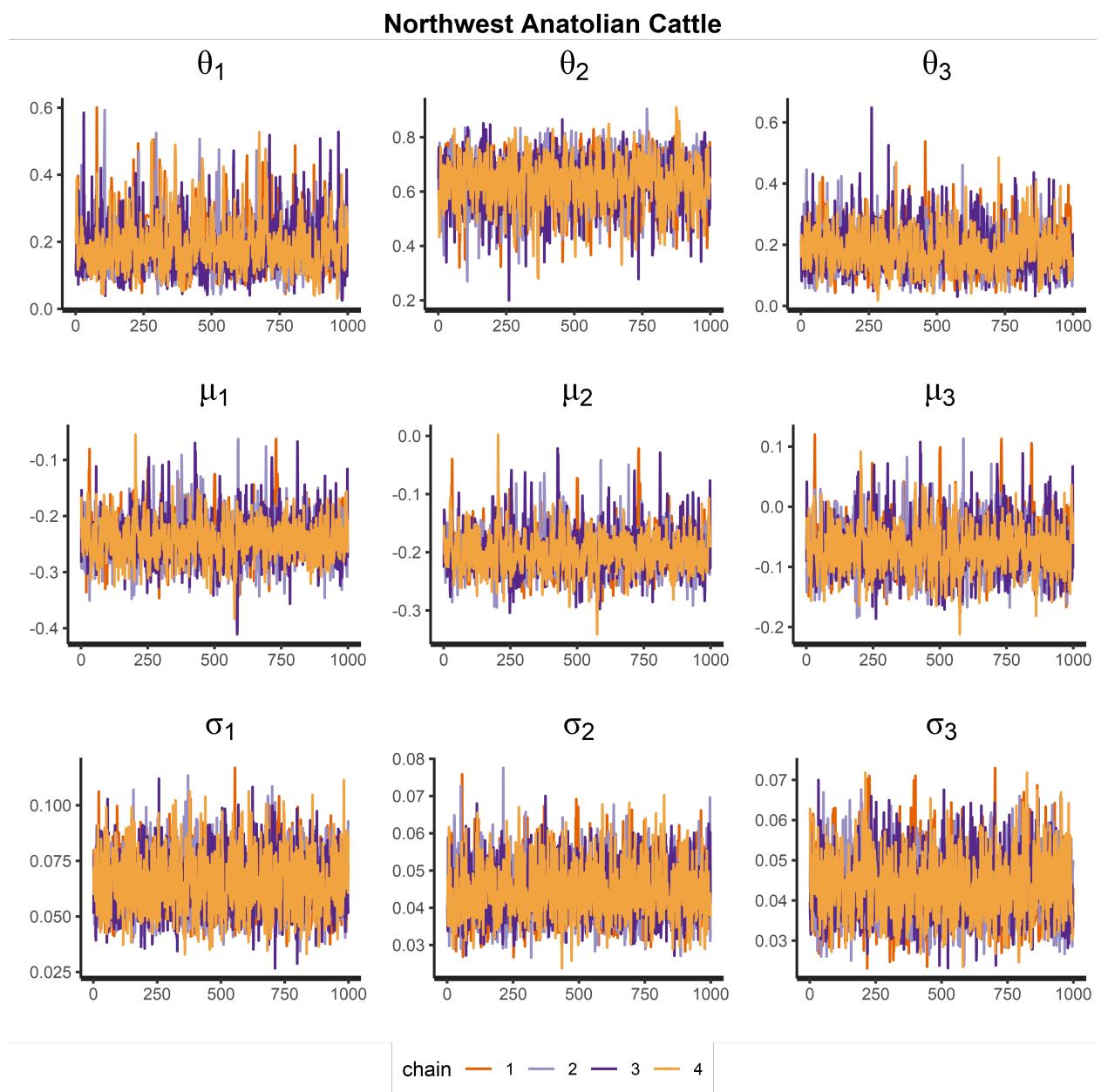


Figure 17: Traceplots of Model Parameters (Northwest Anatolian Cattle)

¹¹³⁴ Appendix 4 (Supplemental Figures 5-7)

¹¹³⁵ *Prior-Posterior Comparisons for Model Hyper-Parameters*

¹¹³⁶

Multisite Simulation

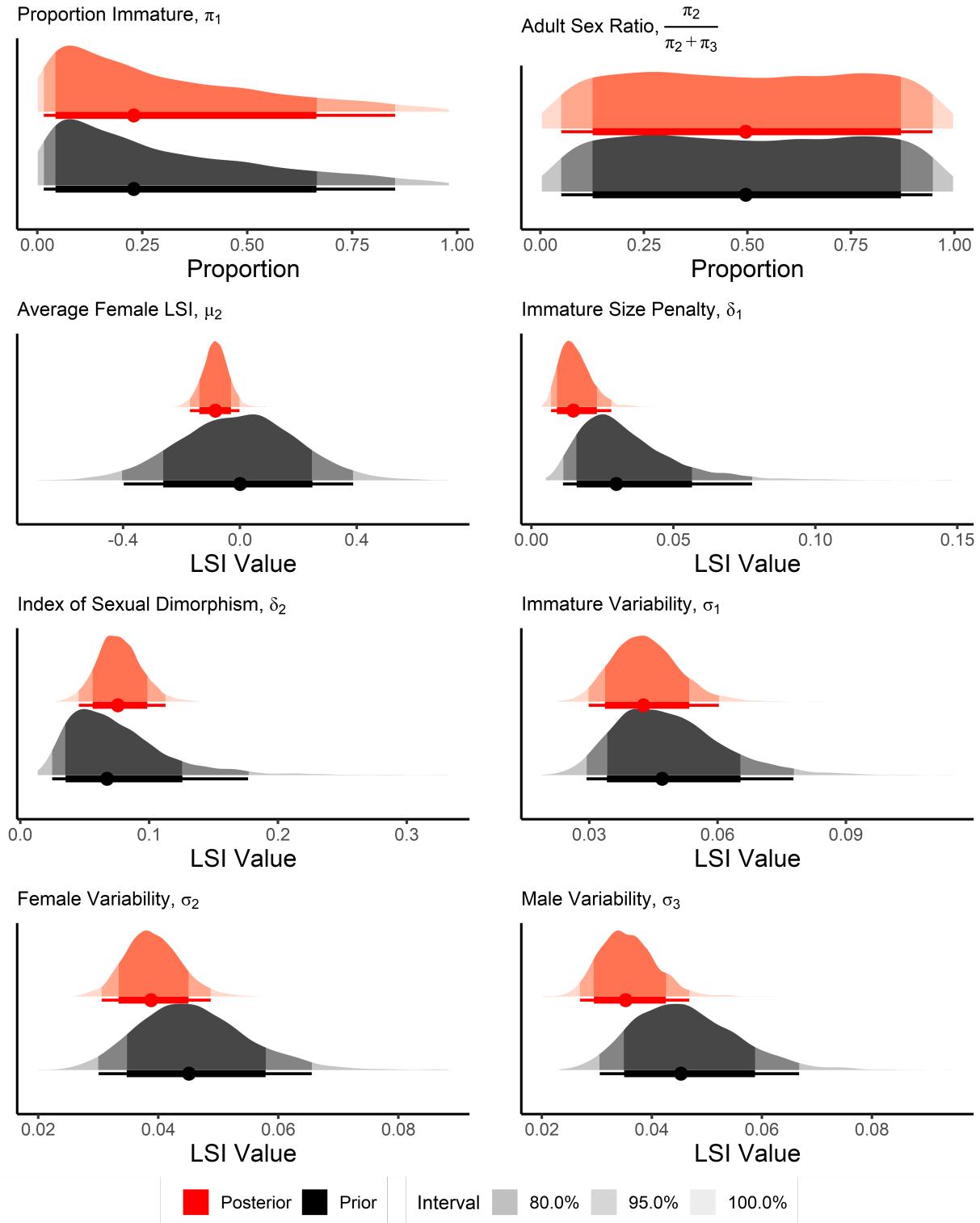


Figure 18: Prior-Posterior Comparison of Multisite Simulation Model Hyper-Parameters

Pinarbaşı B Sheep

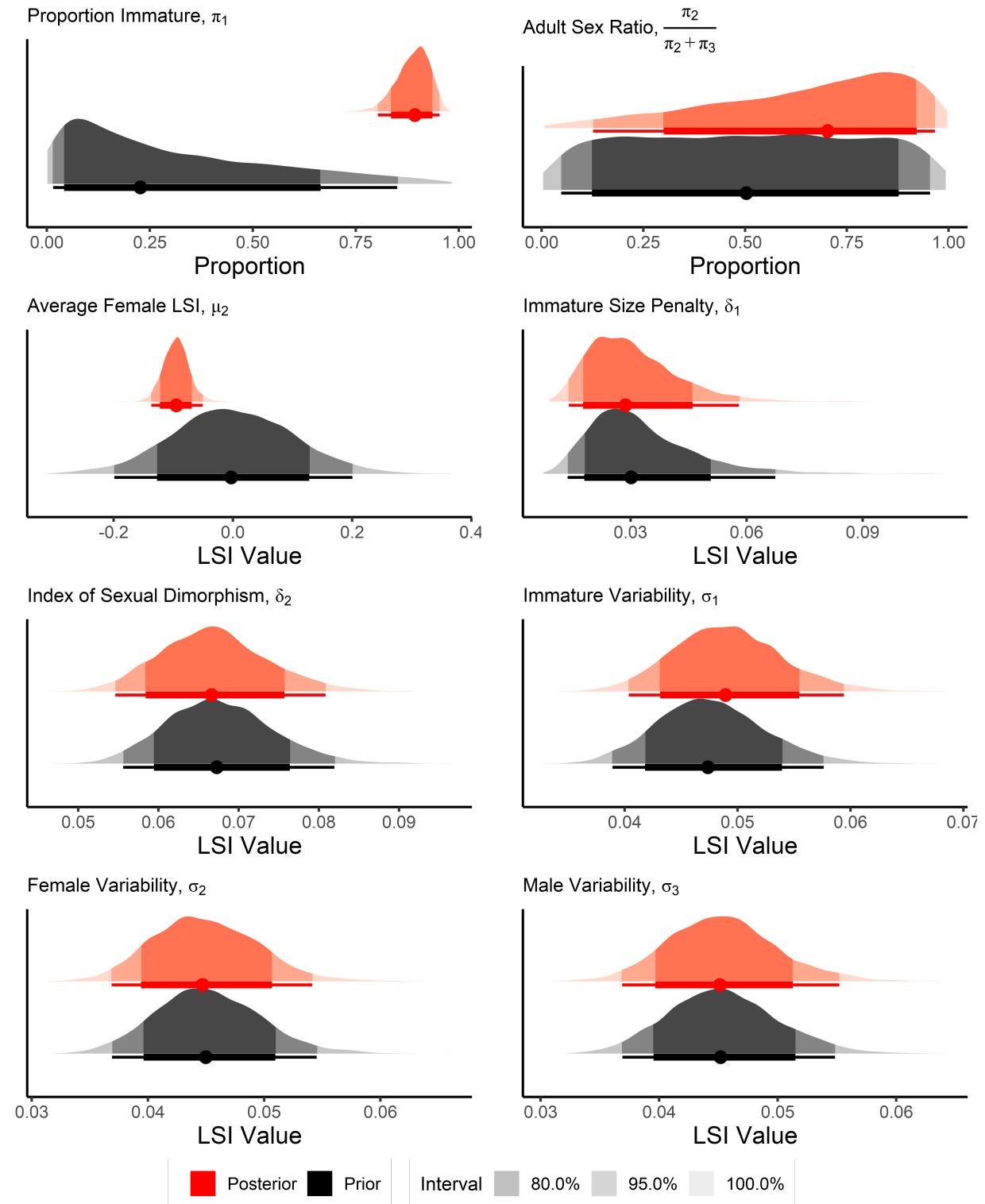


Figure 19: Prior-Posterior Comparison of Pinarbaşı B Sheep Model Hyper-Parameters

Northwest Anatolian Cattle

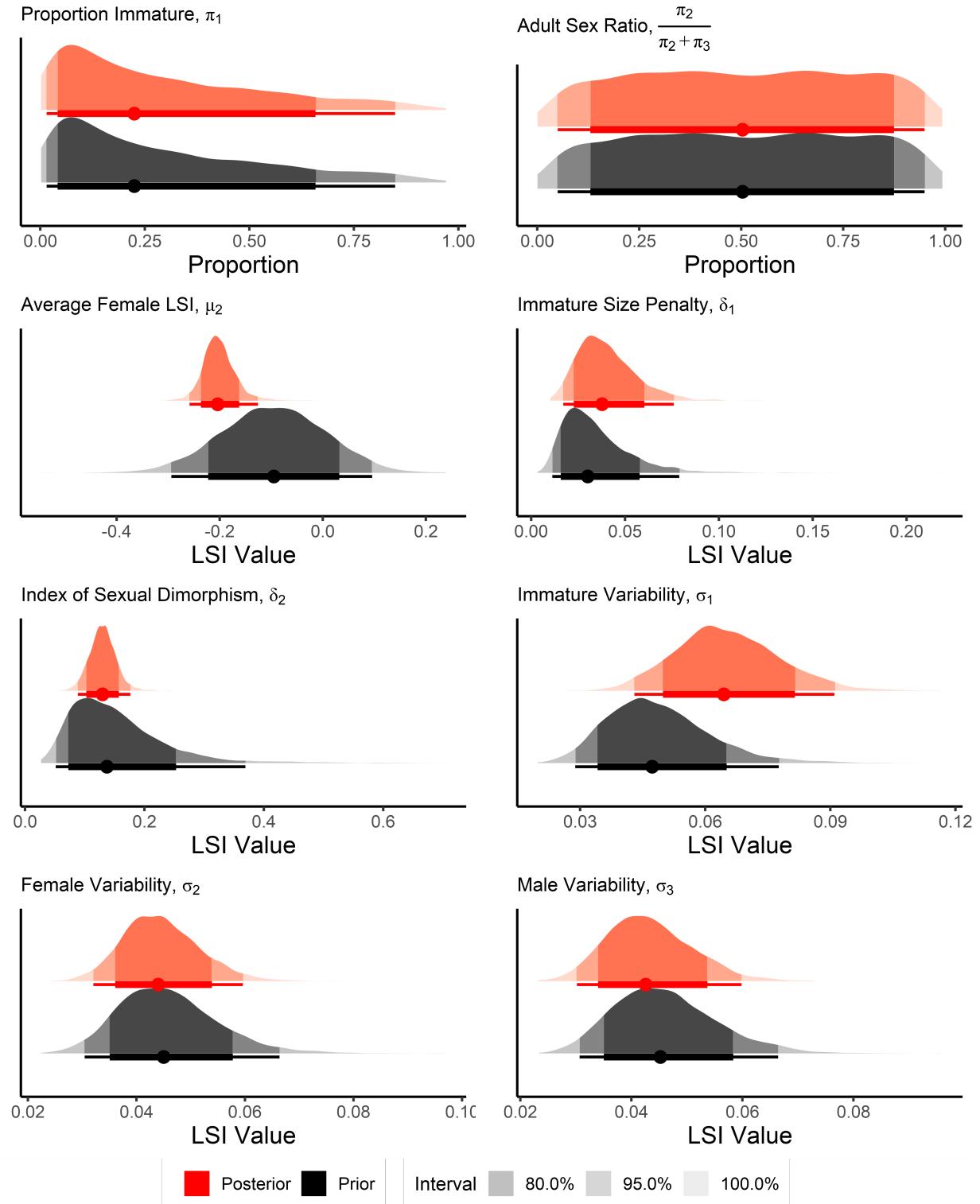


Figure 20: Prior-Posterior Comparison of NW Anatolian Cattle Model Hyper-Parameters