

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 2022-07-12

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 morphology and behavior. As such, the accuracy of our estimates underlies our ability to ascertain details about site formation and gain insights into how people interacted with different animals in the past. Unfortunately, our estimates typically rely on only a small number of bones, limiting our ability to fruitfully use these estimates to make meaningful comparisons to theoretical expectations or even between multiple assemblages. 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 ensure that the parameter estimates are biologically meaningful. Simulated assemblages show that the model can accurately estimate the biometry and composition of zooarchaeological assemblages. Two archaeological case studies also show how the model can be applied to produce tangible insights. The first, focused on sheep from Neolithic Pinarbaşı B, highlights the model's ability to elucidate site formation and function. The second, focused on cattle remains from four assemblages from 7th-6th Millennium BCE northwestern Anatolia, showcases how to use the mixture modeling results to compare assemblages to one another and to specific hypotheses. 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. Introduction

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

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

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

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

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

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

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

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

2. A Bayesian Multilevel Mixture Model for Zooarchaeological Measurements

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

The multilevel structure of the model uses partial pooling to allow the mixture model parameters to vary between element portions while resisting overfitting. These element-specific parameters are related to each other through hyper-parameters, which describe the average value of the model parameters and the variability of model parameters across element portions (Wolfhagen 2020). Prior distributions of mixture model hyper-parameters that are related to biometric variability in animal populations are derived from the 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).

2.1 Mixture Model Likelihood

The central likelihood of the mixture model uses parameters that are specific to each element portion. These parameters include the mixture proportions for the different components of immature animals, females, and 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 component ($\sigma_1, \sigma_2, \sigma_3$). For each element portion, immature animals are described with the first set of 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 males with the third set of parameters (π_3, μ_3, σ_3). This results in both a set of parameters that describe the composition of the assemblage (of measurements from that element portion) and an equation to estimate the probability that a particular specimen comes from an immature, adult female, or adult male individual.

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

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

Second, the proportion of specimens from an element portion that could be immature \$operatorname{proportion} say, for specimens from an early-fusing element-then the mixture components do not need to be re-weighted

2.2 Measurement Error and Observations

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

Because specimens can have multiple measurements that are included in the mixture model on them (e.g., a distal humerus with both a Bd and BT observations or a complete radius with Bp and Bd observations), the mixture model uses specimen-specific LSI values ($LSI_{specimen}$) that are related to $LSI_{measurement}$ values

in the same way. $LSI_{\text{measurement}}$ values are the “observations” with a standard deviation of 0.01 (in LSI_e scale) based on intra-individual variation of LSI_e values for the Popkin et al. (2012) sheep using the *Ovis orientalis* female standard animal (FMC 57951) from Uerpmann and Uerpmann (1994, Table 12).

Observation Error Equations:

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

The model also uses observations of sex ratios and fusion rates to estimate different proportions of the different components in the assemblage. These estimates are not directly used for element portion-specific mixture model parameters but rather inform the hyper-parameters that describe the average value across all element portions. These observations are part of a binomial process centered on the associated hyper-parameters. As such, the measurement error in the observation is based on the total number of observed specimens. The observation of the average proportion of immature specimens (μ_{π_1}) is based on the fusion rate of proximal and middle phalanges, which fuse at around the same time as the estimated time that animals reach adult body size in the Shetland sheep population (Popkin et al. 2012). The observation of the average adult sex ratio—the proportion of females among mature animals ($\frac{\mu_{\pi_2}}{\mu_{\pi_2} + \mu_{\pi_3}}$)—is based on the sex ratio of fused pelvises. In each case, the number of observable specimens (proximal or middle phalanges with observed fusion status, fused pelvises with estimated sexes) determines the measurement error using the binomial distribution. While this paper uses these observations for these hyper-parameters, other observations are possible and can be incorporated into the model in the same fashion.

Demographic Observation Equations:

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

2.3 Prior Distributions

Prior distributions are central to Bayesian inference and describe one’s prior beliefs in potential values of a model parameter. Prior distributions can be likened to a ‘filter’ from which parameter values are drawn

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

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

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

2.3.1 Mixture Proportion Priors

Prior distributions for the mixture components reflect our prior beliefs about the relative proportions of immature, adult-sized females, and adult-sized males in the assemblage. The three mixture components (π_1 , π_2 , π_3) are a three-value unit simplex, meaning that the values are constrained as a group to sum up to one. This means that the simplex can be described by only two variables because the third value cannot

vary once those two values are known. The model uses two unconstrained variables (θ_1 and θ_2) to describe the unit simplex of π values. These θ values are related back to π values using the following ‘stick-breaking’ transformation by iteratively estimating the relative proportions of the simplex taken up by each θ value (Stan Development Team 2022, sec. 10.7).

Stick-Breaking Transformations:

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

The distribution of potential π values in an assemblage is relatively broad—there is no reason to think that any combination of the three mixture components cannot occur. There are, however, general expectations about the relative proportion of each group (immature, female, and male) in a zooarchaeological assemblage that we can use to inform prior distributions of θ values. The θ_1 value can be directly related to the π_1 value using the first line of the stick-breaking transformation, meaning that one can examine the associated π_1 estimate for a given θ_1 value (Figure 1). While it is theoretically possible for all bones in an assemblage to be from immature animals, ethnographically recorded culling strategies and taphonomic factors suggest that it is more likely that immature animals are a smaller component of the assemblage. A ‘standard’ reference prior distribution, $\theta_1 \sim \text{Normal}(0, 2.5)$, only makes it slightly more likely that π_1 is below 50% than not (61%: Figure 1A), which does not fit our intuition about the proportion of immature animals. Using a slightly off-center Normal distribution with a smaller standard deviation, $\theta_1 \sim \text{Normal}(-0.5, 1.5)$, provides a distribution of potential π_1 values that better fit the expectation for the proportion of immature animals in an assemblage (79%: Figure 1B).

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

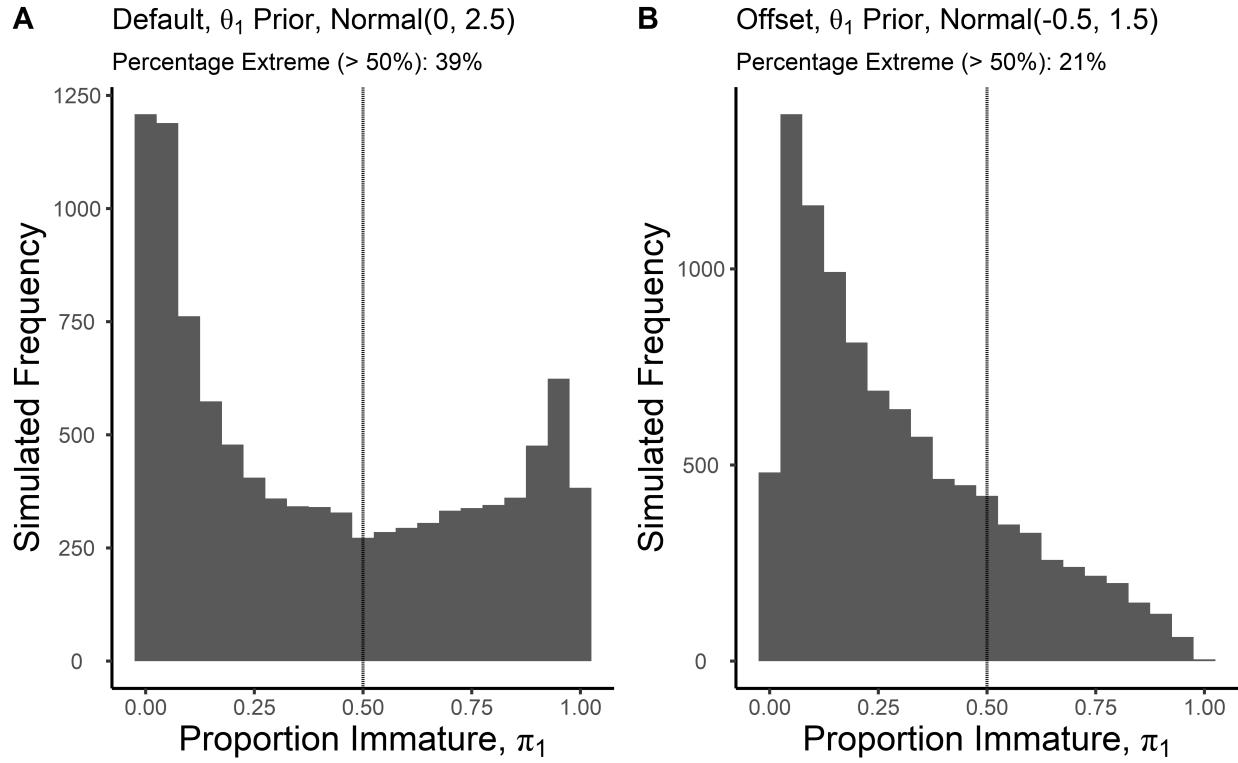


Figure 1: Comparison of π_1 distributions from different θ_1 distributions

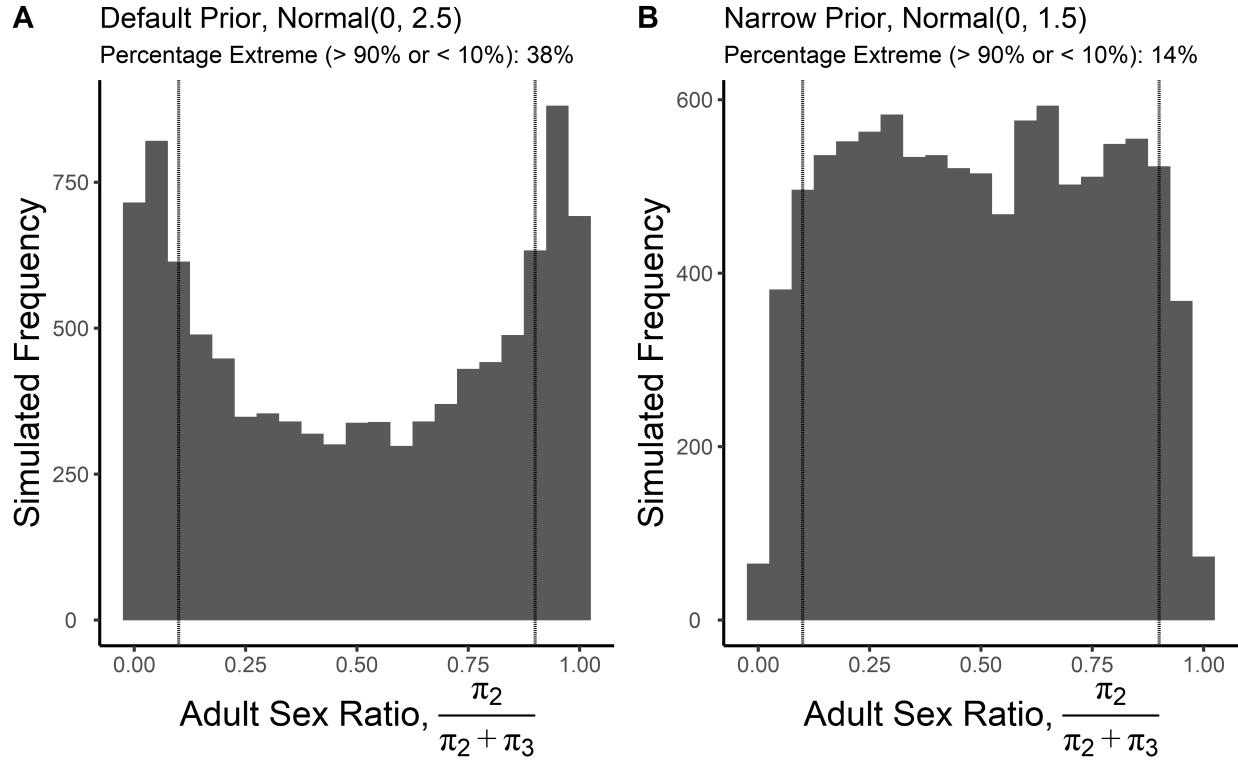


Figure 2: Comparison of adult sex ratios $\frac{\pi_2}{\pi_2 + \pi_3}$ distributions from different θ_2 distributions

Prior Distribution Definitions for θ Hyper-parameters:

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

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.

Offset Equations for μ Values

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

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

Marom and Bar-Oz 2013).

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

The prior distributions used in this Bayesian multilevel mixture model on the reference population are more straightforward. While the same transformations to create unconstrained parameters are necessary (e.g., modeling average size as μ_2 with offsets for immature and male animals), the definition of these prior

distributions can be broadly described as weakly-informative priors (Gelman et al. 2008). These weakly-informative prior distributions are reasonable in this case—and not in the archaeological case—because all the mixture model parameters have direct observations rather than relying on latent state estimations. That is, parameters like the size difference between males and females (δ_2) and the size variability in female animals (σ_2) can be directly observed because the group identities of every specimen are known. With these direct observations, the prior distributions have a more muted influence on the resulting posterior distributions. This does not mean that the prior distributions have no effect, however, which is why objective priors can have undesirable impacts on modeling results (Gabry et al. 2019).

Figure 3 shows the posterior distributions of the model hyper-parameters for both average body size (μ_2 , δ_1 , δ_2) and size variability (σ_1 , σ_2 , σ_3) based on the Shetland sheep sample. These distributions are associated with proposed prior distribution definitions for the same hyper-parameters for archaeological applications of the model when identity is unknown. Note that these hyper-parameters describe averaged estimates across element portions in the model. In general, the posterior distributions from the sample of known-identity Shetland sheep are narrower than the associated prior distribution proposals. This narrowness is due in part to the large sample size of the sampled assemblage but also to the fact that every element portion is represented by the same individuals, something that is extremely unlikely in archaeological scenarios. Thus, care should be taken before directly translating these results into prior distributions for an archaeological scenario.

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

For the average size difference between immature and female animals (δ_1), the narrowness of the posterior distribution likely reflects the fact that immature animals in the sample cover a narrow age range. Animals killed under one year of age span only 36 days and the youngest animals are nearly half a year old (178-214 days: Popkin et al. 2012). Thus, while the posterior results provide a useful starting point for estimating this offset, there is a good potential for larger δ_1 values (i.e., greater size differences between immature and adult-sized female animals) in other contexts that could include animals killed at a younger age (Figure 3B). To capture this possibility, the archaeological model uses a prior distribution with a larger standard deviation

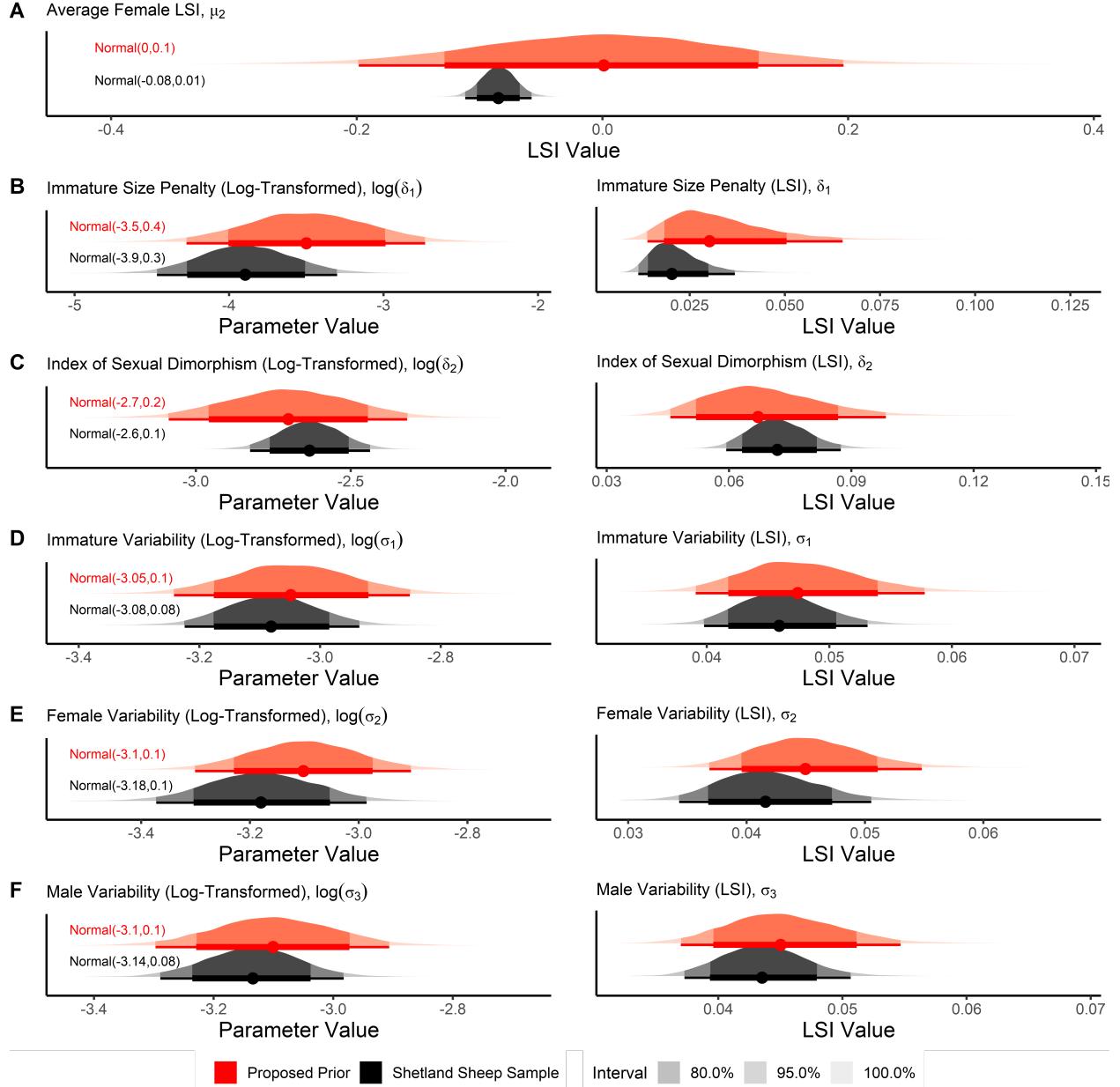


Figure 3: Posterior distributions of model hyper-parameters from sample of known-identity specimens and proposed prior distributions for mixture model applications

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 the LSI_e scale and a 95% probability that the size difference is between 0.01 and 0.07. This translates into expecting the average body size of immature animals in an assemblage being 3% smaller than the average body size of adult-sized female animals, but also plausibly believing that this size difference could range from 1-7% smaller.

The average size difference between adult males and females (δ_2), also known as the index of sexual dimorphism (Fernández and Monchot 2007), is likely to be under stricter biological control than the other “average body size” parameters in the model. This does not mean that this difference could not vary between contexts, however. Some models of animal domestication argue that initial domestication removed sexual selective pressures on male body size, reducing sexual dimorphism (e.g., Tchernov and Horwitz 1991). In a similar fashion, specialized hunting strategies could also reduce sexual dimorphism by targeting large-bodied males, for example (Zeder 2012; Proaktor, Coulson, and Milner-Gulland 2007; Milner, Nilsen, and Andreassen 2007). Again, the posterior distribution of the extent of sexual dimorphism in the Shetland sheep population provides a useful starting point to describe a prior distribution for the model (Figure 3C). Increasing the standard deviation of the distribution slightly, $\log \delta_2 \sim \text{Normal}(-2.7, 0.1)$, produces a distribution centered at 0.07 LSI_e units with a 95% probability that the value is between 0.06-0.08, translating to the average male being 6-9% larger than the average female relative to a standard measurement. The smaller standard deviation in the prior distribution of δ_2 than for δ_1 reflects our understanding that the extent of sexual dimorphism, as a biological phenomenon, is less likely to have extreme values than the average size difference between immature and female animals, since δ_2 is unaffected by the specific age structure of the assemblage.

Prior Distribution Definitions for μ and δ Hyper-parameters:

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

2.3.3 Size Variability Priors

The LSI_e size variability of animals within a group ($\sigma_1, \sigma_2, \sigma_3$) is a key variable in the Bayesian mixture model. The values of these standard deviation parameters play a major role in ensuring that the mixture components reflect biological entities rather than overfitting to specific sample noise. Previous research into size variability suggests that these σ values should be relatively stable across elements. Coefficients of variance (CVs) for raw mammal bone measurements from a single sex have been found to be relatively

consistent (Davis 1996; Popkin et al. 2012). When transforming these measurements using a logarithm, this produces consistent standard deviations of the transformed measurement values (Wolfhagen 2020: Figure 1).

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

As in the average body size parameters, prior distributions for the size variability model parameters are developed from the Bayesian model of known-identity Shetland sheep measurements. The resulting σ hyper-parameters provide a baseline for establishing hyper-parameter prior distributions in archaeological cases. Figure 3D-F shows the posterior distributions of these σ hyper-parameters in both the log-transformed values and associated LSI_e values. Average size variability within an element portion for immature animals (σ_1) is higher, on average, than for females (σ_2) and males (σ_3). The immature category includes both male and female animals, so larger size variability makes sense; again, it is possible that σ_1 is relatively low in this population relative to other contexts given the narrow age range of immature animals in the Shetland sheep population. Unlike the average body size parameters, there are not compelling reasons to believe that size variability parameters for females and males (σ_2 and σ_3) should vary widely in different contexts given the consistency of coefficients of variation in mammals broadly (Davis 1996). Thus, the results of this analysis are used for the prior distributions of $\log_e \sigma_2$ and $\log_e \sigma_3$, while the prior distribution of $\log \sigma_1$ is given an increased standard deviation and slightly increased average value. Overall, however, these prior distributions suggest that the average size variability within an element portion is between 0.04-0.06 for females and males and is between 0.04-0.05 for immature animals. Note that even though σ_2 and σ_3 have the same prior distributions, these values can still vary from each other in different contexts.

Prior Distribution Definitions for σ Hyper-parameters:

$$\begin{aligned} \log \sigma_1 &\sim \text{Normal}(-3.05, 0.1) \\ \log \sigma_2 &\sim \text{Normal}(-3.1, 0.1) \\ \log \sigma_3 &\sim \text{Normal}(-3.1, 0.1) \end{aligned} \tag{8}$$

2.4 Multilevel Structure for Element Portions

The previous section described prior distributions that describe the *average* value for different mixture model parameters across all element portions. To create parameter estimates that are specific to different element

portions, it is necessary to estimate the *variation* around these average values that different parameters can have among different element portions. The model uses a Multivariate Normal definition of the model parameters to allow for correlations between different parameters; effectively, the possibility that multiple model parameters will covary from element portion to element portion. To do this, each hyper-parameter has an associated σ_{element} parameter that describes inter-element variation in parameter values. The model uses a non-centered parameterization, wherein the Multivariate Normal distribution is centered at zero to calculate offsets, ν_{element} , that are added to the average hyper-parameters to calculate model parameters for each element portion. This definition provides computational stability and makes it more straightforward to incorporate other levels of multilevel structure (see Section 2.5).

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

$$\begin{aligned} \nu_{\text{element}} &\sim \text{MultivariateNormal} \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma_{\text{element}} \right) \\ \Sigma_{\text{element}} &= \begin{pmatrix} \sigma_{\text{element}}[1] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\text{element}}[8] \end{pmatrix} \rho_{\text{element}} \begin{pmatrix} \sigma_{\text{element}}[1] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\text{element}}[8] \end{pmatrix} \\ \rho_{\text{element}} &= LKJcorr(2) \\ \theta_1[\text{element}] &= \theta_1 + \nu_{\text{element}}[1] \\ \theta_2[\text{element}] &= \theta_2 + \nu_{\text{element}}[2] \\ \mu_2[\text{element}] &= \mu_2 + \nu_{\text{element}}[3] \\ \log \delta_1[\text{element}] &= \log \delta_1 + \nu_{\text{element}}[4] \\ \log \delta_2[\text{element}] &= \log \delta_2 + \nu_{\text{element}}[5] \\ \log \sigma_1[\text{element}] &= \log \sigma_1 + \nu_{\text{element}}[6] \\ \log \sigma_2[\text{element}] &= \log \sigma_2 + \nu_{\text{element}}[7] \\ \log \sigma_3[\text{element}] &= \log \sigma_3 + \nu_{\text{element}}[8] \end{aligned} \tag{9}$$

Prior distributions for the σ_{element} values are weakly-informative priors based on the scale of the parameter and the expectation for variation in the parameter values among element portions. For example, there is likely more variation in θ parameters—that govern the relative composition of immature, female, and male animals—among element portions than variation in σ parameters that govern size variability within each group. The scale of different parameters also affects the expected spread of values; for instance, μ_2 values

are on the direct LSI_e scale while δ and σ parameters are on the log-scale of LSI_e differences. The prior distributions for σ_{element} parameters associated with μ , δ , and σ parameters are based on results of the known-specimen model, erring on the side of more variability for most parameters. Increased variability in the value of $\log_e \delta_1$ allows for greater size variation across elements, which makes intuitive sense because δ_1 is affected by age composition and biology rather than strictly biology.

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

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

2.4.1 Extending the Multilevel Analysis to Multiple Sites

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

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

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

The inclusion of multiple sites changes the definition of the ‘grand mean’ variable (θ_1 in the example equation) from a site-level estimate to an overall mean across the sites and elements. These parameter estimates thus describe the average composition of the entire set of assemblages. The details of the assemblages included in the analyses would affect how useful these estimates are for interpretation. Assemblage-specific estimates can be calculated for each model parameter by adding the relevant ν_{site} estimate to the ‘grand

mean' parameter, which would again act to describe the average composition of the assemblage regardless of its elemental composition.

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

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

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

The second extension focuses on the multilevel structure of the model, specifically the assumption that the overall mixture model hyper-parameters for *modeled* element portions are equally valid for *unmodeled* element portions. Because the model produces estimates of the "average" hyper-parameters and the vari-

ability of these parameters across element portions (σ_{element}), one could use these data to estimate the element portion-specific mixture model parameter values of a completely unmodeled element portion (McElreath 2020; Gelman, Carlin, et al. 2020). These parameter estimates, then, could be used to estimate π_{specimen} membership probabilities for the unmodeled (and unmeasured) specimens to serve as the baseline for simulated assemblages.

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

2.6 Computational Details of the Bayesian Model

The Bayesian multilevel mixture model is written in Stan, version 2.21.0 (Stan Development Team 2022). All analyses in this paper use R version 4.1.3 (2022-03-10), in Rstudio 2022.2.3.492 (Prairie Trillium) (R Core Team 2022; RStudio Team 2022). The analytical scripts use the following packages: “data.table” version 1.14.2, “parallel” version 4.1.3, and “doParallel” version 1.0.17 for data aggregation and multi-core processing (Dowle and Srinivasan 2021; Microsoft Corporation and Weston 2022), “cmdstanr” version 0.5.0 and “rstan” version 2.21.3 for running the Bayesian models in the R environment and summarizing the posterior distributions of the models (Gabry and Cešnovar 2022; Stan Development Team 2021), “zoolog” version 0.4.1 for standard animal measurements (Pozo et al. 2021), and “Cairo” version 1.5.15, “ggplot2” version 3.3.5, “ggdist” version 3.1.1, “ggpubr” version 0.4.0, “rnatural-earth” version 0.1.0, “rnatural-earthdata” version 0.1.0, and “sf” version 1.0.7 to create visualizations (Urbanek and Horner 2022; Wickham 2016; Kay 2022; Kassambara 2020; South 2017a, 2017b; Pebesma 2018).

The model Stan code and analytical R code necessary to replicate and apply the analyses in this paper are freely available in a GitHub page (<https://github.com/wolfhagenj/ZooarchMixMod>) and Open Science Framework page (<https://osf.io/4h9w6/>). The project includes a copy of the Shetland sheep data file from the supplemental files published in Popkin et al. (2012) and archaeological datasets for the case studies downloaded from OpenContext (Carruthers 2006; Buitenhuis 2013; Galik 2013; Gourichon and Helmer 2014). The analytical code includes two script files—a script for replication and one for application. The

R markdown file (“ZooarchMixMod.Rmd”) file replicates the entire analytical workflow of the paper, with a specific seed set to ensure exact replicability of the submitted manuscript. Another set of scripts to standardize the analytical workflow for faunal datasets structured like the OpenContext faunal datasets used in these case studies, see the GitHub for more details. All scripts (R and Stan) are released under the MIT license and figures are released as CC-BY to encourage reuse and reproducibility (Marwick 2017; Marwick and Pilaar Birch 2018).

3. Testing the Bayesian Multilevel Mixture Model

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

3.1 Simulated Assemblages

A series of simulated assemblages of known-age and sex composition are created from the Shetland sheep population by randomly drawing element portions (and all associated measurements) from the total assemblage without replacement. The first test, using a single-assemblage model, uses 150 element portions from the Shetland sheep population where every element portion has an equal probability of being selected. There is no guarantee, however, that the element portions have equal representation or even that all element portions are present in the simulated assemblage, which better approximates archaeological assemblages. The result of this first simulation produces an assemblage of 231 measurements from 125 individual animals. The second test creates 15 simulated assemblages using the same procedure that are analyzed in a single multi-assemblage model. To further test the model’s flexibility, these assemblages vary in sample size and some are manipulated to vary in average body size and expected composition from the original Shetland sheep

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

population. Demographic observations for phalanx fusion rates and pelvis sex ratios were also simulated from the Shetland sheep population using the same underlying probabilities as the measurement assemblages. Table 3 describes the sample sizes of the measurement assemblages, including any manipulations to the measurement values. The specific elemental composition and measurements of the assemblages, along with the simulated demographic observations, used in both simulations can be recovered from the replication script with the recorded random seed (see also Supplemental Tables 1-3); using another random seed would provide a conceptual replication of new assemblages drawn from the same underlying populations.

For each model test, accuracy is evaluated in two ways. First, the posterior distributions of the mixture model hyper-parameters are compared to the known values for the population (“parametric accuracy”). The model hyper-parameters are first transformed into estimates of composition (π), body size (μ) and size variability (σ) rather than the unconstrained variables that were modeled directly. These known values are calculated from the entire Shetland sheep population from which the assemblages were sampled (Popkin et al. 2012), including any manipulations of composition or size as described in Table 3. Second, the mixture model results are used to estimate the age and sex composition of the relevant measured and modeled assemblages—estimates of the number of immature, female, and male specimens for each element portion—that are compared to the actual composition of the assemblages (“compositional accuracy”). For the multi-assemblage model, compositional accuracy is evaluated both overall (all the assemblages combined)

and for each assemblage individually. As described in Section 2.5, modeled assemblages are the measured and unmeasured specimens from the relevant element portions that are modeled with the mixture model. The measurement assemblage is considered a sample from this modeled assemblage, under the assumption that measurability—sufficient preservation of anatomical structures to be measured—is unrelated to the age and sex category of the specimen. Thus, just as the parametric accuracy is compared against the “population” parameters from which the measured assemblages were sampled, so to the compositional accuracy can be compared against the “modeled” assemblage that includes non-measured specimens sampled from the same underlying population.

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

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

$$\begin{aligned}
 \theta_1 &\sim \text{Normal}(-0.5, 1.5) \\
 \theta_2 &\sim \text{Normal}(0.0, 1.5) \\
 \mu_2 &\sim \text{Normal}(0.0, 0.2) \\
 \log \delta_1 &\sim \text{Normal}(-3.5, 0.5) \\
 \log \delta_2 &\sim \text{Normal}(-2.7, 0.5) \\
 \log \sigma_1 &\sim \text{Normal}(-3.05, 0.25) \\
 \log \sigma_2 &\sim \text{Normal}(-3.10, 0.2) \\
 \log \sigma_3 &\sim \text{Normal}(-3.10, 0.2)
 \end{aligned} \tag{11}$$

3.2 Archaeological Case Studies

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

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

The site of Pinarbaşı, located in the Konya Plain of central Turkey, consists of a series of rock shelter and open-air sites at the foothills of the Karadağ volcanic region and Lake Hotamış and associated wetlands (Baird et al. 2011; Kabukcu 2017). This case study examines the Pinarbaşı B late Neolithic occupation, which is dated to the second half of the 7th millennium BCE and includes a large number of domesticated

sheep and goat remains (Baird et al. 2011; Carruthers 2005). Carruthers (2005) analyzed fauna from the 1994-1995 excavations by Trevor Watkins (Watkins 1996), interpreting the presence of fetal sheep remains and other juvenile remains in the assemblage as evidence for herders penning sheep on-site. The Neolithic assemblage was thus described as the result of seasonal occupation by sheep and goat herders during the lambing season and the fall, with culling in the spring possibly focused on young males (Carruthers 2005). This analysis makes several claims that can be evaluated with the Bayesian multilevel mixture model: the dominance of immature remains, a female-dominated adult sex ratio, and sex-specific differences in fusion rates for later-fusing elements.

The Bayesian multilevel mixture model for the late Neolithic Pinarbaşı B assemblage uses 44 sheep measurements from 44 specimens (see Table 4). In addition to these measurements, the observed proportion of immature animals from unfused first and second phalanges is 59 / 62 (95%), including specimens identified to sheep and to sheep/goat. There are 0 observed sheep (or sheep/goat) pelvis bones with sex identifications; this is entered into the model by having an observed adult sex ratio for the assemblage of 0 / 0 (females / females + males). All data come from the Pinarbaşı faunal assemblage uploaded to OpenContext, focusing only on specimens in the Site B Neolithic contexts (Carruthers 2006). The Pinarbaşı B sheep model uses the same prior distribution definitions for the model hyper-parameters as the single assemblage simulation since both models, even though the sheep body sizes likely differ between the two populations, showcasing the flexibility of the standard prior distribution definitions.

Table 4: Elemental composition of the Pinarbaşı B sheep measurement 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

3.2.2 Multiple Assemblages: Biometric Analysis of Cattle from 7th-6th Millennium BCE Northwest Anatolia (Barçın Höyük, İlipinar Höyük, Menteşe Höyük)

Understanding the development of Neolithic communities in northwestern Anatolia has long been of interest for researchers interested in studying the spread of agricultural lifeways from southwest Asia into Europe (e.g., Çakırlar 2013; Karul 2019; Özdoğan 2011, 2019). Agricultural communities first appear in the

Marmara region in the mid-seventh millennium BCE in sites like Barcın Höyük (Gerritsen and Özbal 2019; Karul 2019). The domestic animal economies of these Late Neolithic and Early Chalcolithic communities appears to be focused on cattle and caprine (sheep and goat) herding, rather than pig husbandry (Buitenhuis 2008; Çakırlar 2013; Gourichon and Helmer 2008). Milk residues on pottery recovered from these sites suggest that these communities regularly consumed milk, potentially orienting herd management strategies of sheep, goats, and particularly cattle to specialize in milk production (Evershed et al. 2008; Thissen et al. 2010).

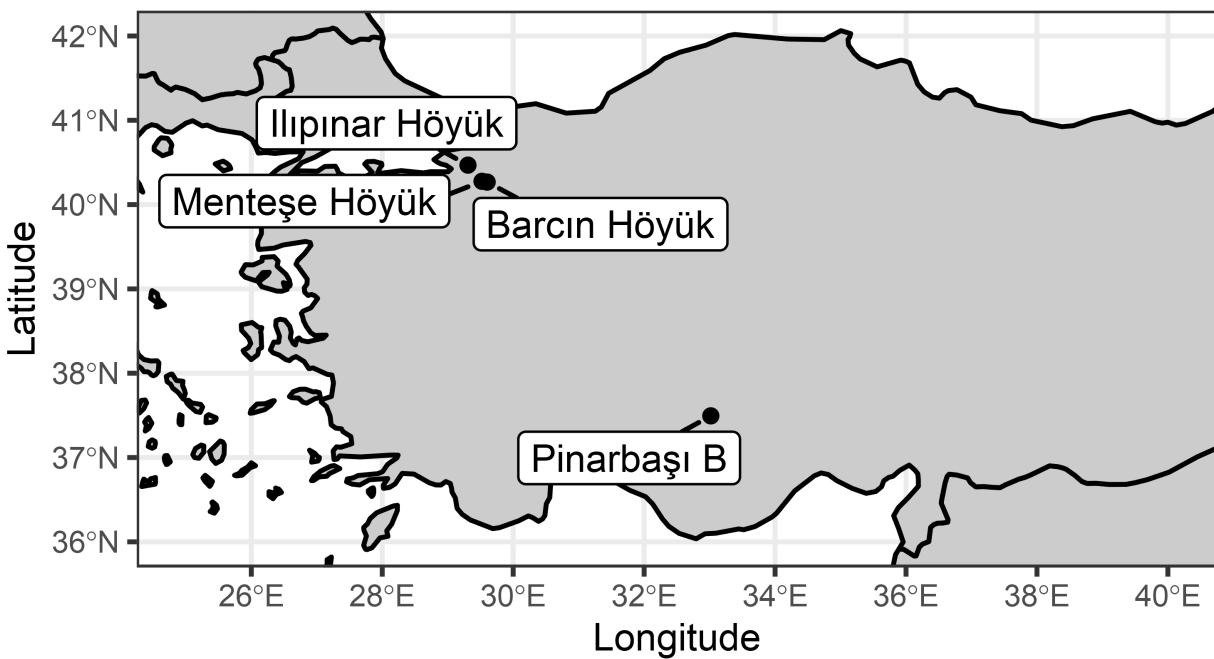


Figure 4: Map of archaeological sites included in this analysis

Four archaeological components from three sites are used in this case study, located near Lake İznik and on the Yenişehir Plain in the Bursa province of Turkey (Figure 4). The Neolithic layers from Barçın Höyük (Phase VI) is the earliest of these assemblages, with occupation roughly from 6500-6000 cal BCE; excavations revealed a subsistence economy focused on cereal agriculture and the herding of cattle, sheep, and goat (Galik 2013; Gerritsen and Özbal 2019). Menteşe Höyük is located approximately five km west of Barçın Höyük on the Yenişehir Plain; the three Neolithic layers at the site date to 5800-5600 cal BCE (Gourichon and Helmer 2014; Roodenberg et al. 2003). Previous faunal analysis of the Neolithic assemblage identified animal economies that shifted from predominantly cattle to sheep herding over the course of the occupation (Gourichon and Helmer 2008). İlipinar Höyük is located near Lake İznik, separated from the Yenişehir Plain by a mountain ridge (Roodenberg 2012a). The Neolithic/Early Chalcolithic occupation of the site spanned 6200-5400 cal BCE (Buitenhuis 2013); I divided the assemblage into two sub-assemblages (Neolithic İlipinar = Phases X-VII, 6000-5700 cal BCE; Chalcolithic İlipinar = Phases VI-V, 5600-5400 cal BCE), marked by the introduction of mudbrick architecture and expanded storage (Roodenberg 2012a, 2012b). Sheep and goat are common in the earlier assemblages of the site, with cattle becoming predominate in later phases of the site (Buitenhuis 2008; Roodenberg 2012a). Notably for this biometric analysis, Buitenhuis (2008) notes that cattle body sizes are stable throughout the Neolithic layers.

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

Previous syntheses of the Late Neolithic and Early Chalcolithic animal economies in northwest Anatolia

Table 5: Elemental composition of the Northwest Anatolian cattle measurement 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

provide several assumptions about the age and sex structure of cattle bone assemblages that can be evaluated with the results of the Bayesian multilevel mixture model. First, the general cultural continuity of the assemblages suggests that the biometry and composition of cattle bone assemblages may be similar at the sites, having been produced by similar processes (e.g., Çakırlar 2013; Özdoğan 2019). Second, the widespread evidence of milk consumption from pottery residue analyses from these sites and others in the region (Evershed et al. 2008; Thissen et al. 2010) suggest that cattle were managed for milk production (Gourichon and Helmer 2008; Roodenberg 2012a); thus, one may expect that each assemblage has a female-dominated sex ratio and potentially higher fusion rates for later-fusing elements among females than males. The multilevel modeling results can be used to evaluate the feasibility of these assumptions by examining posterior distributions of relevant parameters and simulations of sex-specific fusion rates.

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

wild counterparts (e.g., Tchernov and Horwitz 1991).

Prior Distribution Definitions for the Northwest Anatolian Cattle Model Hyper-Parameters

$$\begin{aligned}
\theta_1 &\sim \text{Normal}(-0.5, 1.5) \\
\theta_2 &\sim \text{Normal}(0, 1.5) \\
\mu_2 &\sim \text{Normal}(-0.1, 0.1) \\
\log \delta_1 &\sim \text{Normal}(-3.5, 0.5) \\
\log \delta_2 &\sim \text{Normal}(-2, 0.5) \\
\log \sigma_1 &\sim \text{Normal}(-3.05, 0.25) \\
\log \sigma_2 &\sim \text{Normal}(-3.10, 0.2) \\
\log \sigma_3 &\sim \text{Normal}(-3.10, 0.2)
\end{aligned} \tag{12}$$

4. Results

4.1 Simulated Assemblages

Bayesian models work by updating prior information with new data to produce posterior distributions of parameters of interest (Otárola-Castillo et al. 2022). Thus, the difference between a model parameter's prior and posterior distribution shows the amount that the model "learns" from the data-if the data do not provide relevant information on a parameter's potential values, then the posterior distribution will resemble the prior distribution. Figure 5 compares the prior and posterior distributions of the main model hyper-parameters for the single assemblage simulation. The results show that the data provides much more information about the likely values of the two demographic parameters (the proportion of immature animals, π_1 , and the 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 prior distribution definitions were weakly-informative priors (Gelman et al. 2008), but also shows how these choices did not appear to severely influence the resulting posterior distributions.

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

that was used to develop the prior distributions, so it is possible that this overlap reflects that fact.

Parametric accuracy for the simulations relies on relating the posterior distributions for “assemblage-level” parameters to the known values for the overall Shetland sheep population (including any relevant modifications). Ideally, the 80% and 95% credible intervals from these posterior distributions should contain the true population value 80% and 95% of the time, respectively. An overfit model would contain the population value less frequently than the stated interval, producing false confidence in the applicability of the results. An underfit model, by contrast, would contain the population value more frequently than the stated interval, producing results that are too conservative. The single-assemblage simulation produces overfit results: 8 / 9 (89%) of the 80% credible intervals and 9 / 9 (100%) of the 95% credible intervals from the posterior distributions of the population parameters contain the true value of the Shetland sheep population. The multisite simulation produces relatively well-calibrated estimates for assemblage-level parameters, with 98 / 135 (73%) of the 80% credible intervals and 128 / 135 (95%) of the 95% credible intervals from the posterior distributions of the (site-level) population parameters containing the true values of the Shetland sheep population with relevant modifications.

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

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

This overfitting is to be expected since the measured assemblage is a subsample of the modeled assemblage.

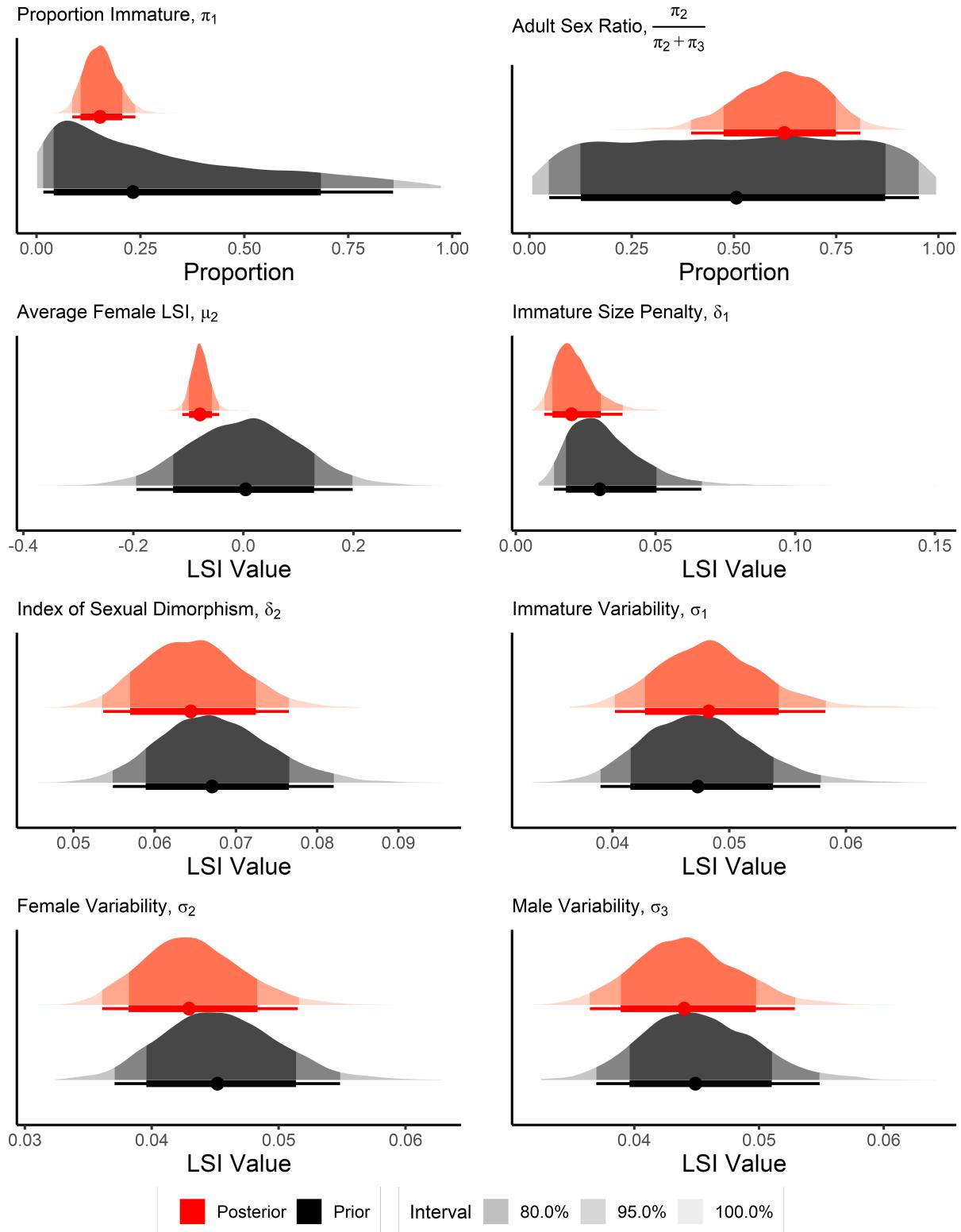


Figure 5: Prior-posterior comparison of single assemblage simulation model hyper-parameters

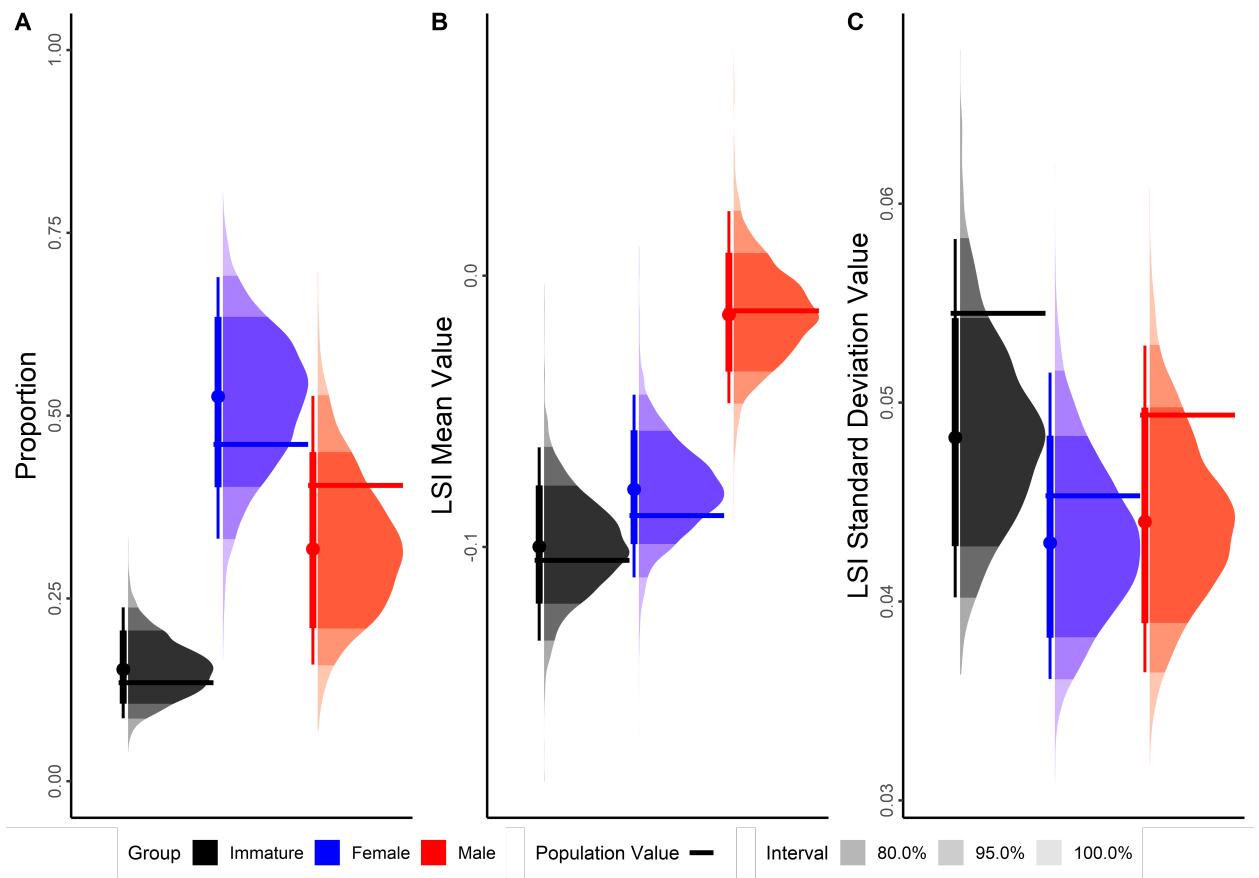


Figure 6: Posterior distributions of model parameters for the simulated single assemblage with relevant population parameter values

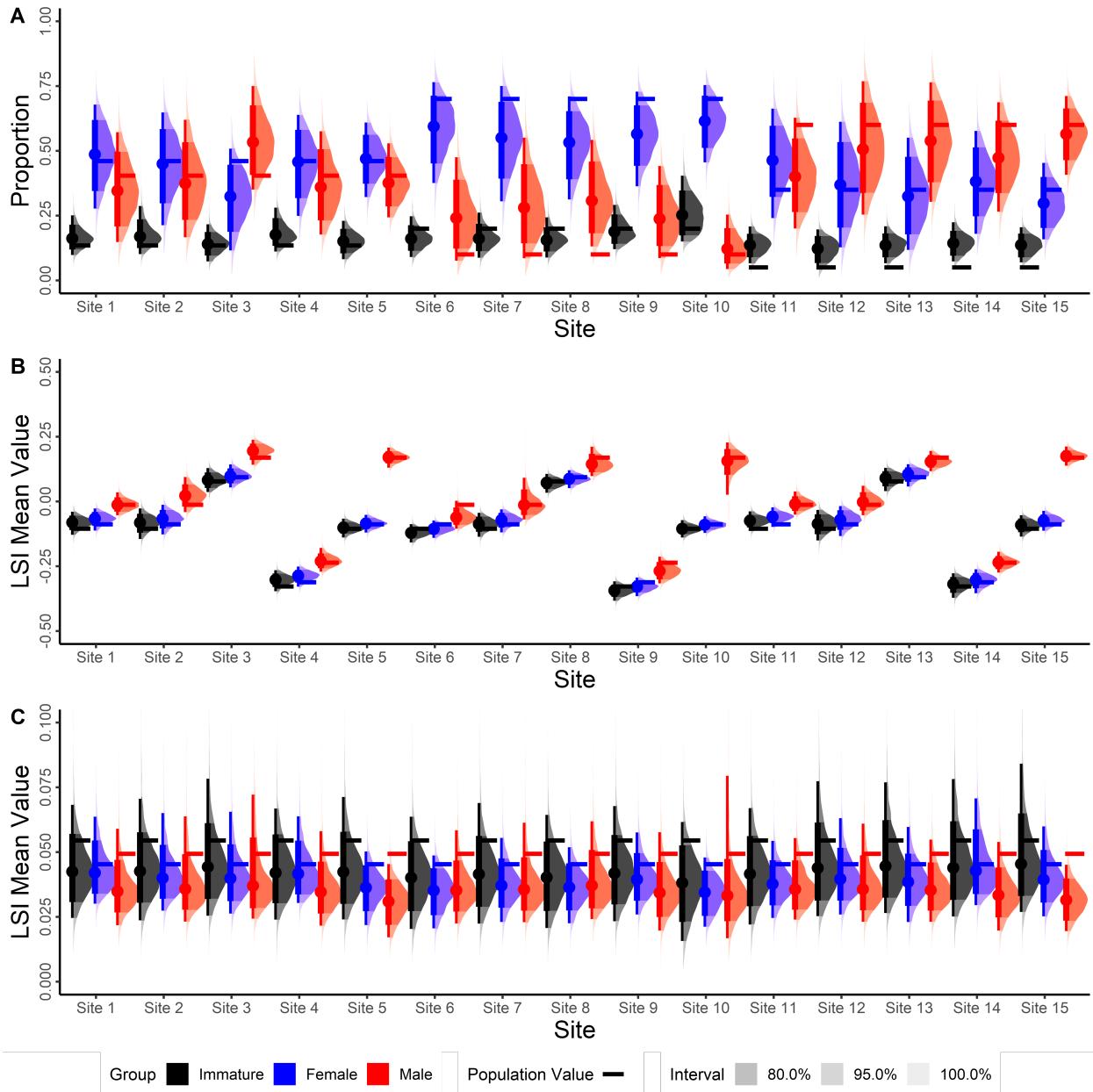


Figure 7: Posterior distributions of model parameters for the simulated multisite assemblages with relevant population parameter values

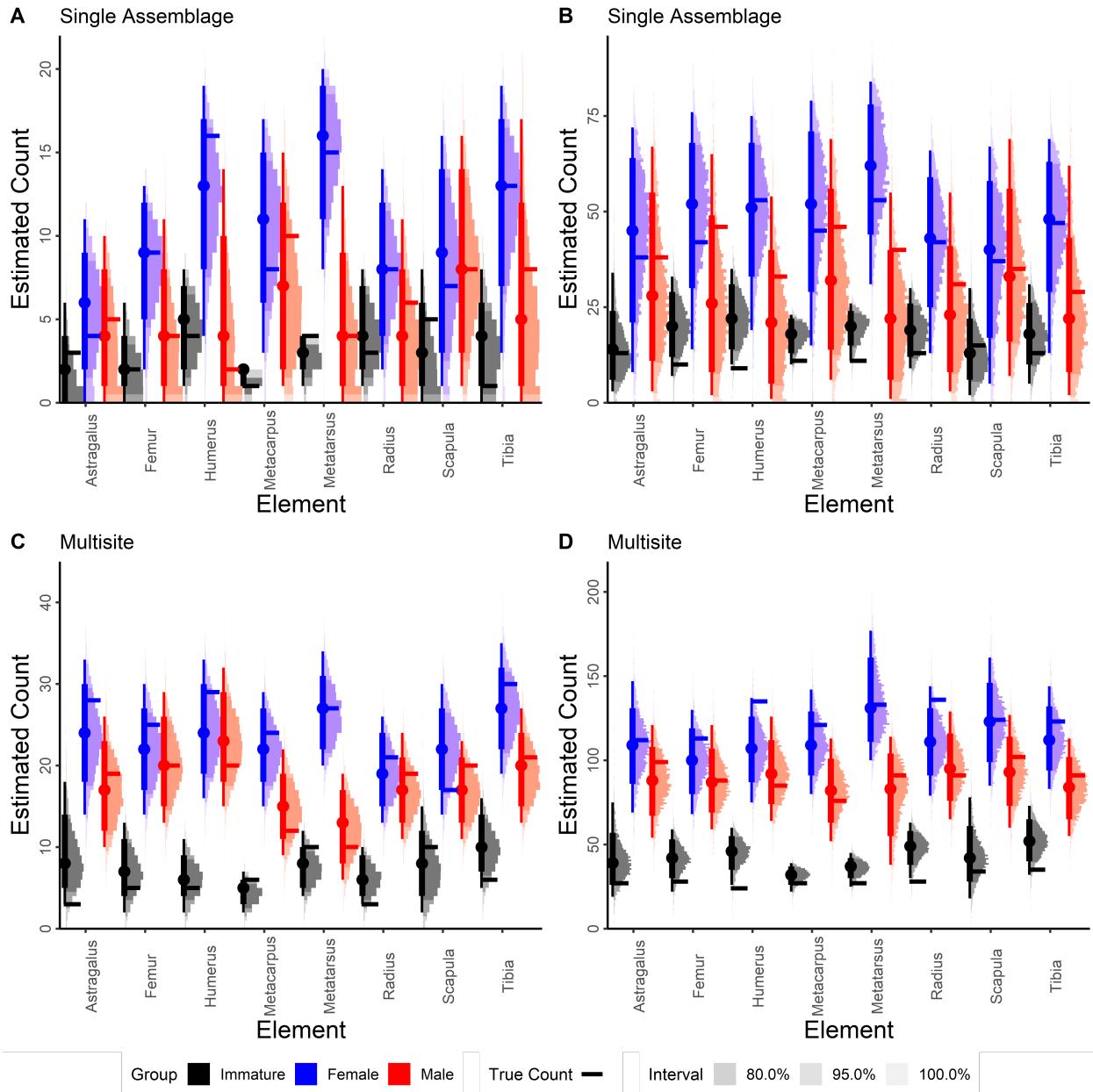


Figure 8: Posterior distributions of group-specific compositions for the simulated measured and modeled assemblages

The model does not overfit as much when estimating the composition of the modeled assemblages: for the single-assemblage model, 20 / 24 (83%) of the 80% credible intervals and 23 / 24 (96%) of the 95% credible intervals contain the true group-specific counts for the element portions. For the multisite model, the model does not perform well for the combined assemblages (17 / 24 (71%) of the 80% credible intervals and 22 / 24 (92%) of the 95% credible intervals) but produces well-calibrated estimates for the site-specific estimates of group composition, with 305 / 360 (85%) of the 80% credible intervals and 353 / 360 (98%) of the 95% credible intervals containing the true group-specific counts. Note that the denominator for the site-specific estimates differs between the measured and modeled assemblages. This is caused by the additional sampling to create modeled assemblages and the multilevel structure of the model, which estimates element-specific offsets and interaction terms (ν_{element} and $\nu_{\text{interaction}}$) for elements that are present in at least one site. Among the 7 newly-observed elements in the modeled assemblage, 18 / 21 (86%) of the 80% credible intervals and 21 / 21 (100%) of the 95% credible intervals included the true group-specific abundance, despite having no observed measurements. That these estimates are broadly as accurate as the estimates from observed element portions suggests that the hyper-parameters can be used to create estimates of unobserved (i.e., unmeasured) element portions in an assemblage.

4.2 Pinarbaşı B Sheep: The impact of immature specimens

Figure 9A shows the posterior distributions assemblage-level π mixture components. In general, the Pinarbaşı B sheep assemblage is overwhelmingly composed of immature animals (posterior μ_{π_1} median = 89%; 95% posterior confidence interval for $\mu_{\pi_1} = 80\text{-}95\%$), somewhat lower than the observed fusion rate of proximal and middle phalanges (59 / 62 = 95%). The posterior distribution of the overall adult sex ratio (θ_{female}) suggests that females are more common than males (76% of the posterior samples are above 0.5), but this estimate is uncertain, owing to the low overall proportions of female and male animals.

Figure 9B shows the element-specific distributions of π_1 for the Pinarbaşı B assemblage. Half of the element-specific π_1 distributions are similarly concentrated in the upper range of possible π_1 values, with posterior medians over 85%; all specimens from these element portions were considered potentially immature by the model due to fusion status or being an element that does not fuse or exhibits post-fusion growth. However, these element-specific distributions also have long tails extending into lower π_1 values, conveying less certainty about element-specific π_1 estimates relative to the assemblage-wide estimate. This likely owes to small element-specific sample sizes (the astragalus, calcaneus, and proximal phalanx have 9-10 specimens, all other element-specific samples sizes are 1-4, see Table 4) and to the presence of some element portions with lower modeled π_1 values. These element portions—especially the distal metacarpal and distal metatarsal—have posterior π_1 median values below 50%, though again have long tails that extend into higher π_1 values.

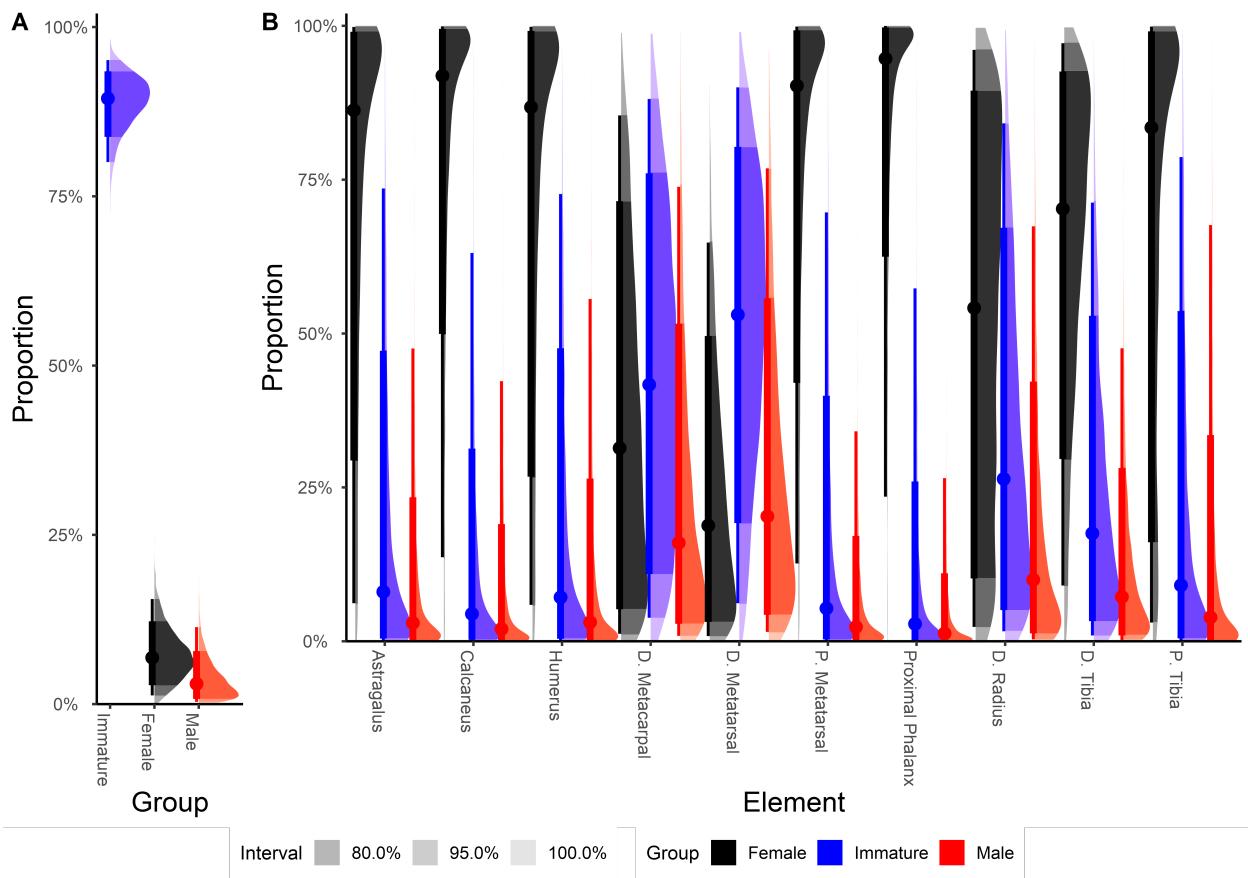


Figure 9: Posterior distributions of (A) overall mixture components (π) and (B) element-specific mixture components (π) for the Pinarbaşı B sheep assemblage

Notably, all measured specimens from these two elements are not considered potentially immature because their distal epiphyses are fused.

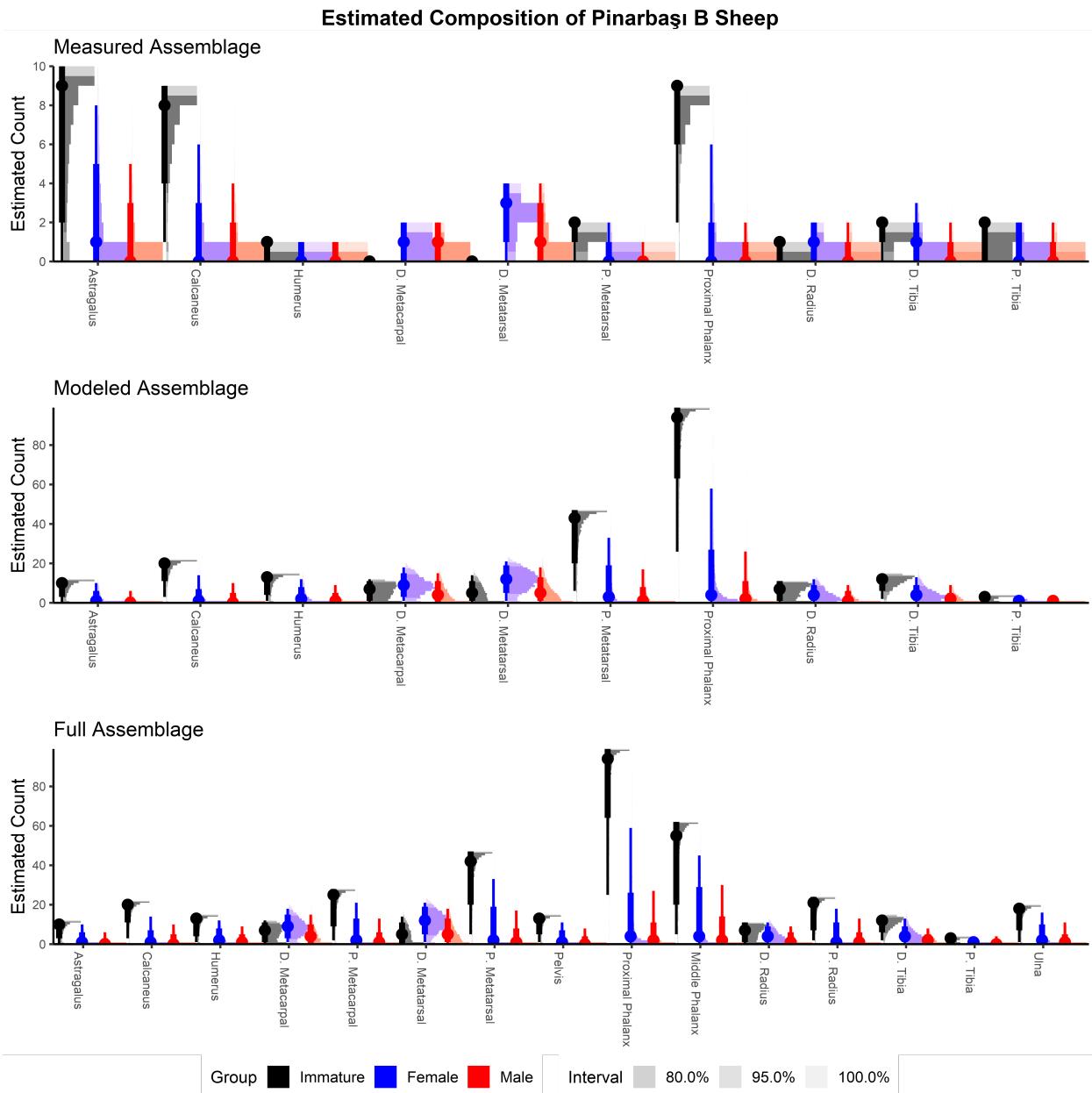


Figure 10: Posterior distributions of group-specific composition for the Pinarbaşı B sheep measured, modeled, and full assemblages

Figure 10 shows the distribution of simulated compositions for immature, female, and male specimens in three Pinarbaşı B assemblages: the measured assemblage ($N = 44$), the assemblage of modeled element portions ($N = 277$), and the full sheep assemblage including five element portions that were not modeled due to lack of measurements (additional elements: proximal radius, ulna, proximal fused metacarpal 3 and

4, pelvis, and middle phalanx; total N = 428). While the model has much greater precision in estimating the composition of the measured assemblage, including the non-measured specimens in the modeled assemblage provides a clearer overall picture of the assemblage. Most element portions appear to be predominantly from immature animals rather than adults (males or females), with the notable exceptions of distal metapodials and, to a lesser extent, the distal radius. Inclusion of the unmodeled elements broadly reinforces this pattern, while providing data to explore fusion rates or comparisons of larger animal portions. These results strongly reinforce the interpretation that Pinarbaşı B was used by herders as a camp where sheep gave birth, with on-site culling largely reflecting either the first seasonal cull of animals before winter or animals that died naturally in their first year of life (Carruthers 2005; Martín et al. 2015). The model also provides a testable estimate of the adult sex ratio, despite the assemblage being dominated by immature animals, allowing researchers to infer the overall biometry and herding strategy from a seasonal signal.

4.3 7th-6th Millennium BCE Northwest Anatolian Cattle: Examining differences between assemblages

Figure 11 shows the posterior distributions of average body sizes for female cattle (μ_2) from the four analyzed assemblages. These distributions are produced from posterior samples; assemblage-specific estimates from a single posterior sample share the same relevant hyper-parameters (μ_{μ_2} and $\sigma_{\text{site}[4]}$), meaning that they covary with one another to an extent. To compare these distributions, then, a contrast is necessary to account for this potential covariation. This is done by simply evaluating the difference between two parameters (e.g., between the average female LSI_e value μ_2 for Barçın Höyük and μ_2 for Neolithic İlipinar) in each posterior sample, shown in the right-hand panel of Figure 11. These contrasts show that the female cattle from Chalcolithic İlipinar are likely smaller, on average, than female cattle from the other sites. These cattle measurements are 4-6% smaller, on average, than those from the other northwestern Anatolian sites relative to the standard animal's measurements.

Despite this size difference in female animals between the assemblages, the age and sex composition of the four assemblages are broadly consistent with one another. Figure 12 shows the distributions of assemblage-level demographic variables—the proportion of immature animals and the adult sex ratio (the proportion of adults that are female)—for the four northwest Anatolian assemblages. As expected from earlier syntheses that indicated herds kept to produce milk, all of the assemblages have female-dominated adult sex ratios. The relatively large proportions of immature cattle in the Menteşe Höyük and Barçın Höyük assemblages may indicate that cattle were penned on the site, where infants killed by herders or natural causes were included in the assemblage (Gillis et al. 2014, 2015). Contrasts indicate that both İlipinar Höyük assemblages have lower the proportions of immature animals than either the Barçın Höyük or Menteşe Höyük assemblages. This

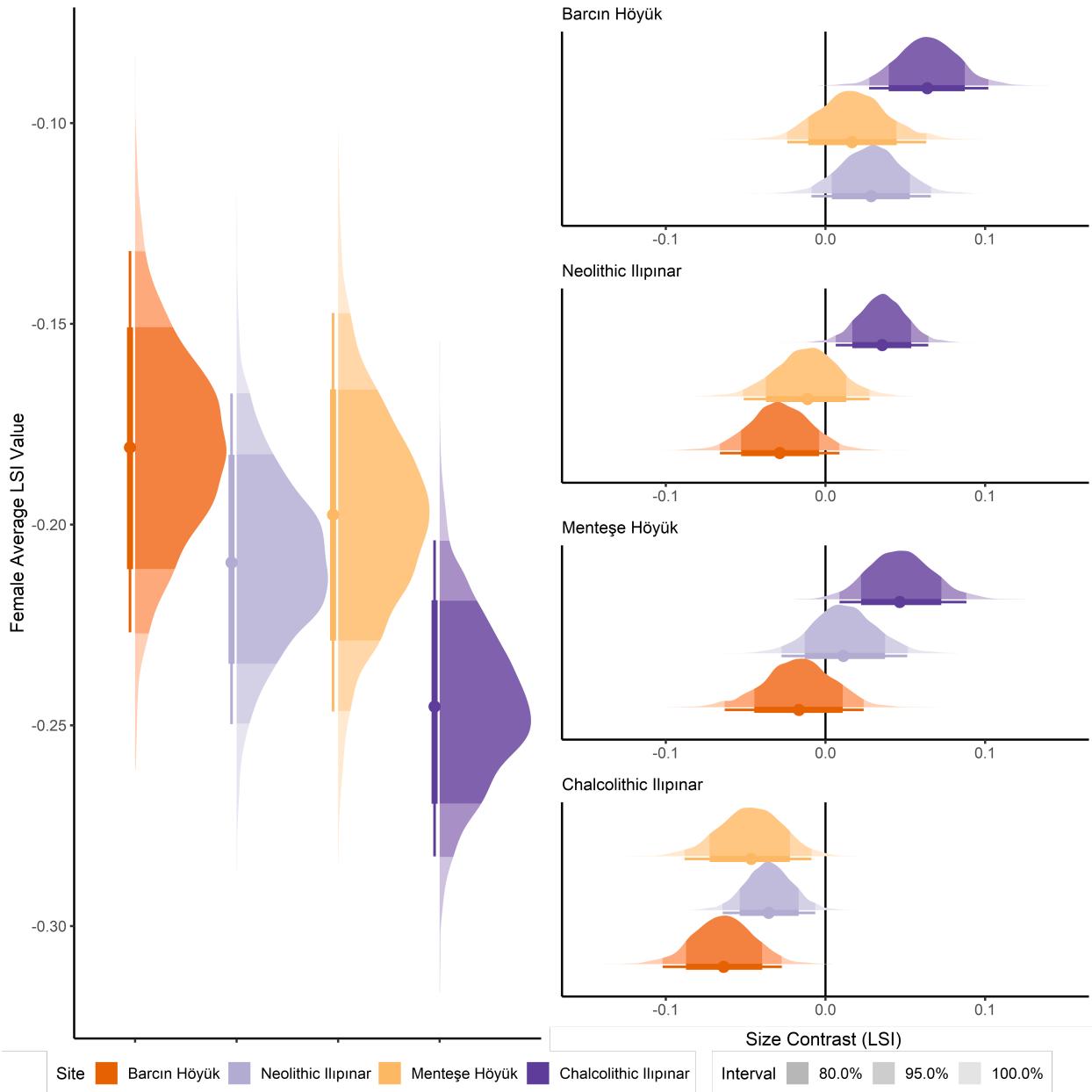


Figure 11: Posterior distributions of site-specific average female LSI values (μ_2) for the Northwest Anatolian cattle assemblages. The right panel shows site-specific contrasts for this parameter, indicating specific size differences between pairs of sites

suggests that, despite evidence for penning deposits on-site for Chalcolithic Iipinar (Roodenberg 2012a), cattle were not giving birth at the site or immature specimens did not preserve as well (Gillis et al. 2014). The small proportion of very young mandibles in the assemblage corroborates the biometric modeling results (Buitenhuis 2008: Figure 18).

Simulating sex-specific fusion rates for late-fusing elements (proximal femur, distal femur, proximal humerus, distal radius, proximal tibia, proximal ulna: Grigson 1982) from the full northwest Anatolian assemblages highlights the complexities of examining sex-specific fusion rates in zooarchaeological assemblages. In each assemblage, estimates of male fusion rates are extremely uncertain, owing to the small number of estimated males in each iteration and thus large potential shifts in the denominator for fusion rates (Figure 13). This uncertainty makes it difficult to clearly establish whether fusion rates differed between males and females; regardless, in 66% of the posterior samples female fusion rates were higher than male fusion rates for Chalcolithic Iipinar. This provides some support for the idea that males in the assemblage were killed at younger ages than females, consistent with a milk-producing management strategy that kept females alive longer than males (Zeder and Hesse 2000; Gillis et al. 2014). The small proportion of immature specimens in the assemblage, though, may further indicate that cattle herding involved some seasonal movement away from the site, rather than cattle living (and being culled) at or near the site year-round.

5. Discussion

The simulation analyses show that the Bayesian multilevel mixture model presented here can accurately reconstruct age- and sex-specific biometry of a faunal population represented in a measured assemblage, while also producing relatively accurate estimates of the “demographic” (age and sex) composition of the assemblage. The performance of the Bayesian multilevel mixture models relies on the prior distributions, which provide constraints against overfitting and ensure that the model produces biologically reasonable parameter estimates. The prior distribution definitions in this paper were derived largely from the measurements of a herd of known-age, known-sex population of Shetland sheep (Popkin et al. 2012), though for the multisite cattle model some of the definitions were changed based on data on European aurochs (Degerbøl 1970). It is important to note that prior distribution definitions can be derived from many different sources—including quantification based on one’s judgment (e.g., Gelman et al. 2008; McCarthy and Masters 2005). More important than the source of one’s prior distribution definitions, is investigating the expectations of those prior distribution definitions by performing prior predictive checks as in Section 2.3 (Gabry et al. 2019; Gelman, Vehtari, et al. 2020). Further, emphasis should be paid to increasing the diversity of known-age, known-sex animal populations with individual measurement data (e.g., Lebennon and Munro 2022; Zeder and Lemoine

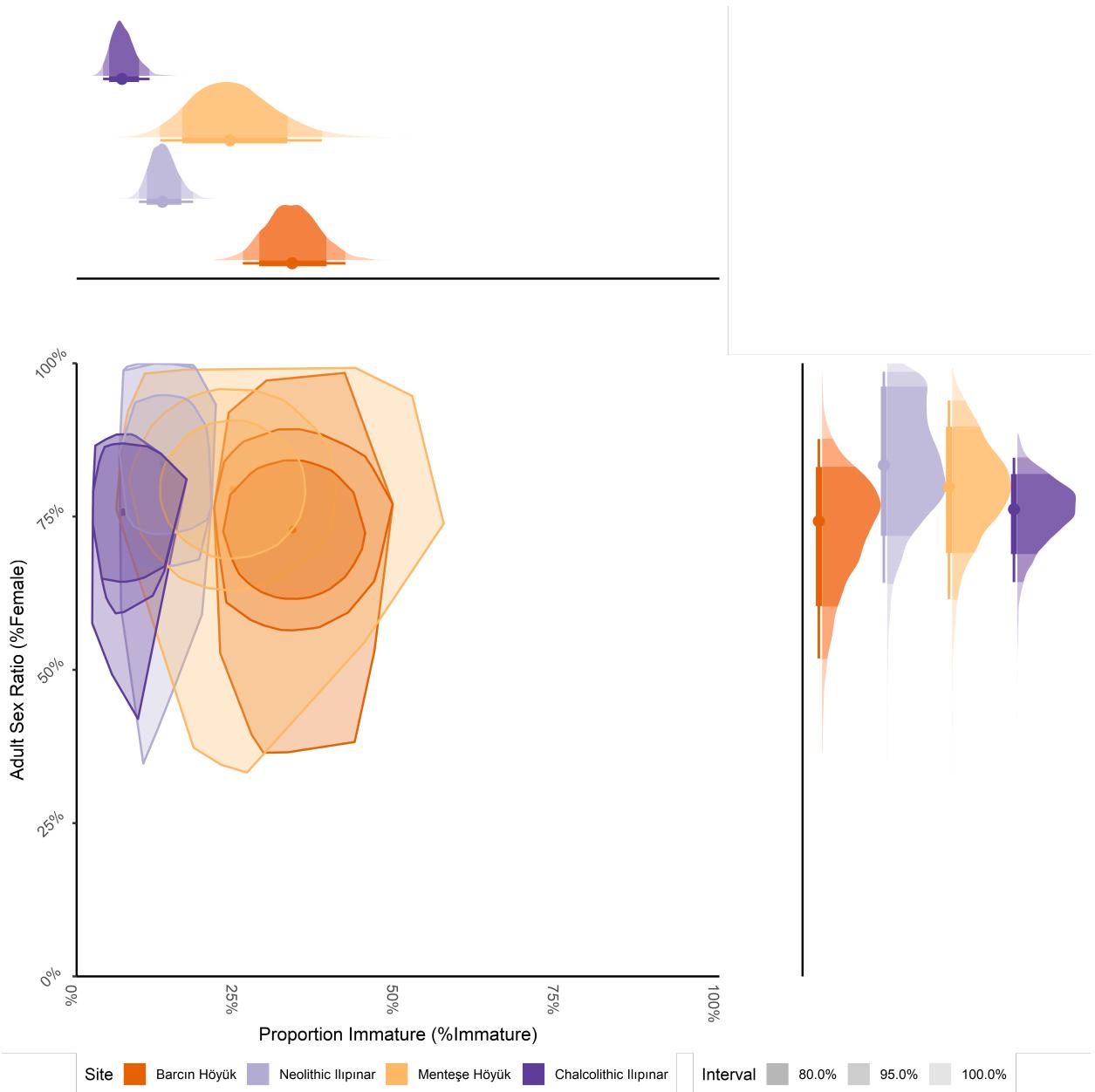


Figure 12: Comparison of posterior distributions of site-specific demographic parameters (π_{immature} and θ_{female}) for the Northwest Anatolian cattle assemblages. Side panels show marginal plots to compare the distributions of each parameter individually across sites

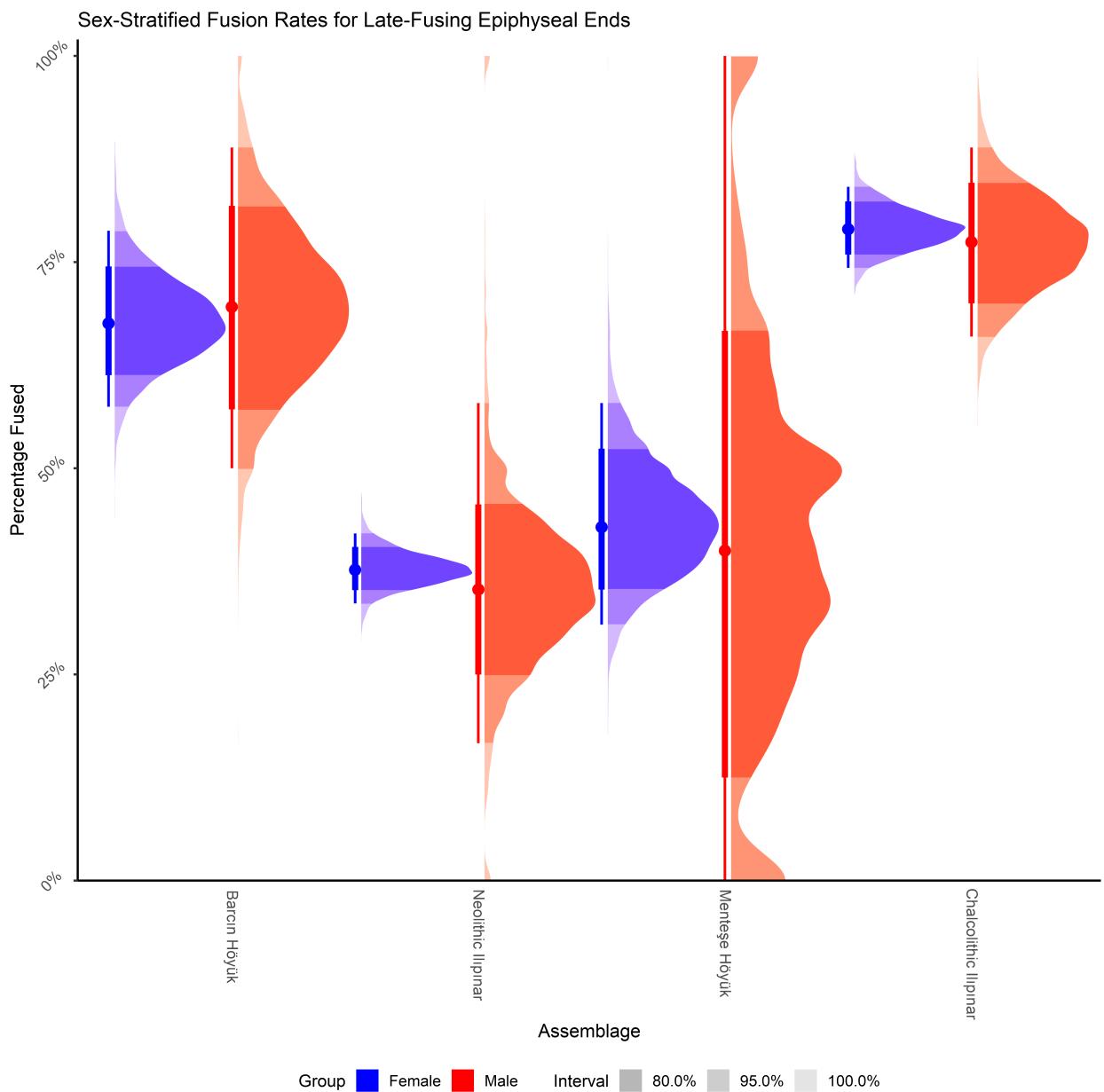


Figure 13: Posterior distributions of sex-specific fusion rates for late-fusing elements in Northwest Anatolian cattle full assemblages

2020), which could help develop prior distributions relevant to different taxa and to understand how variable different parameters, especially size variability (σ) parameters, are across populations.

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

The ability to create accurate simulated estimates of age and sex composition provides many opportunities for further analyses. For instance, comparison of the composition of animals in different depositional contexts could support contextual taphonomic analyses (e.g., Meier 2020). Access to certain kinds of animals could highlight systems of provisioning or status-related restrictions (Arbuckle 2012; Twiss 2019: 73–97). Differences in the age and sex composition of different body parts could also highlight ritual behaviors reflected in the use of certain contexts or sites (e.g., Madgwick and Mulville 2015). On a more practical level, providing specimen-specific probabilities of being immature, female, or male can provide a useful baseline for sampling strategies focused on ancient DNA or stable isotopes, allowing researchers to explore potential sex differences in diets (e.g., Post et al. 2001) or more easily identify male specimens to isolate Y-chromosomal DNA to explore sex-specific selection (e.g., McGrory et al. 2012; Daly et al. 2021).

The archaeological case studies highlight the importance of considering the presence of immature specimens and elemental variation in body size when summarizing the biometry and composition of an assemblage. Variation in the proportion of immature animals in the assemblage, as in the multisite case study for Northwest Anatolia, may point to differences in culling strategies or even the seasonality of animal presence at

the sites. Most of the Pinarbaşı B material derives from immature specimens, which could complicate inter-assemblage biometric analyses that do not use sex-specific size estimates (e.g., Arbuckle et al. 2014; Helmer et al. 2005). Restricting measurements only to fused specimens removes useful information, particularly when fusion rates may differ between male and female animals (Zeder and Hesse 2000); further, it does not resolve the problem of immature animals in the measurement assemblage if early-fusing elements like the distal humerus are still included because of post-fusion growth (Popkin et al. 2012). The ability to create sex-specific biometric estimates is important to document large-scale spatial and temporal dynamics in animal body size (e.g., Arbuckle et al. 2016; Wolfhagen et al. 2021).

Examining LSI_e (or LSI_{10}) values of Pinarbaşı B sheep without accounting for the impact of immature animals would mislead a researcher into believing that sheep were smaller, on average, than sheep from contemporaneous sites (e.g., late Çatalhöyük: Baird et al. 2011). Sex-specific size estimates, accounting for the presence of immature animals, show that the Pinarbaşı B sheep are similar in size to contemporaneous Çatalhöyük sheep (Wolfhagen et al. 2021). This bolsters the interpretation that Pinarbaşı B and Çatalhöyük sheep are the same animal population, and thus that the site was occupied seasonally by Çatalhöyük herders. While the estimated average LSI_e for female sheep (μ_2) at Pinarbaşı B is relatively broad (95% credible interval: -0.14 - -0.05 LSI_e) because of the small proportion of adult-sized females in the assemblage (95% credible interval of the proportion of adult-sized females π_2 : 1-16%), it can still be fruitfully compared to other assemblages [95% credible interval of μ_2 for late Çatalhöyük sheep: -0.11 – -0.06 LSI_e ; Wolfhagen et al. (2021)].

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

6. Conclusions

This paper describes a new method for estimating the biometry and age/sex composition of faunal assemblages based on standard measurement data, Bayesian multilevel mixture modeling. The model produces accurate estimates of sex-specific biometry, which can provide a more useful framework for inter-assemblage analysis (e.g., Arbuckle et al. 2016; Helmer et al. 2005). Such a framework could better explore broad spatial and chronological patterns in animal biometry while accounting for differences in assemblage composition across the assemblages, ensuring reliable comparisons of animal body size in relation to other variables. These analyses could investigate the processes behind size fluctuation in animals, particular in relation to changing human-animal interactions and adaptation to new lifeways and anthropogenic environments.

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

Acknowledgements

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

References

- Arbuckle, Benjamin S. 2012. "Animals and Inequality in Chalcolithic Central Anatolia." Journal Article. *Journal of Anthropological Archaeology* 31 (3): 302–13. <https://doi.org/10.1016/j.jaa.2012.01.008>.
- Arbuckle, Benjamin S., and Levent Atici. 2013. "Initial Diversity in Sheep and Goat Management in Neolithic South-Western Asia." Journal Article. *Levant* 45 (2): 219–35.
- Arbuckle, Benjamin S., Sarah Whitcher Kansa, Eric C. Kansa, David C. Orton, Canan Cakirlar, Lionel

- Gourichon, A. Levent Atici, et al. 2014. "Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals Across Neolithic Turkey." Journal Article. *PLoS One* 9 (6): 1–11. <https://doi.org/10.1371/journal.pone.0099845>.
- Arbuckle, Benjamin S., Max D. Price, Hitomi Hongo, and Banu Öksüz. 2016. "Documenting the Initial Appearance of Domestic Cattle in the Eastern Fertile Crescent (Northern Iraq and Western Iran)." Journal Article. *Journal of Archaeological Science* 72: 1–9. <https://doi.org/10.1016/j.jas.2016.05.008>.
- Baird, Douglas, Denise Carruthers, Andrew Fairbairn, and Jessica A. Pearson. 2011. "Ritual in the Landscape: Evidence from Pinarbaşı in the Seventh-Millennium Cal BC Konya Plain." Journal Article. *Antiquity* 85: 380–94.
- Betancourt, Michael. 2017. "A Conceptual Introduction to Hamiltonian Monte Carlo." Journal Article. *arXiv* 1701.02434 [Preprint]: Available from <https://arxiv.org/abs/1701.02434v1>.
- Breslawski, Ryan P. 2022. "Minimum Animal Units and the Standardized Count Problem." Journal Article. *Journal of Archaeological Method and Theory*. <https://doi.org/10.1007/s10816-022-09563-9>.
- Breslawski, Ryan P., and David A. Byers. 2015. "Assessing Measurement Error in Paleozoological Osteometrics with Bison Remains." Journal Article. *Journal of Archaeological Science* 53: 235–42. <https://doi.org/10.1016/j.jas.2014.10.001>.
- Buitenhuis, Hijlke. 2008. "Ilipinar: The Faunal Remains from the Late Neolithic and Early Chalcolithic Levels." Book Section. In *Archaeozoology of the Near East VIII*.
- . 2013. "Ilipinar Zooarchaeology." Dataset. <http://opencontext.org/projects/D297CD29-50CA-4B2C-4A07-498ADF3AF487>. <https://doi.org/https://doi.org/10.6078/M76H4FB>.
- Çakırlar, Canan. 2013. "Rethinking Neolithic Subsistence at the Gateway to Europe With New Archaeozoological Evidence from Istanbul." Book Section. In *Barely Surviving or More Than Enough?: The Environmental Archaeology of Subsistence, Specialisation and Surplus Food Production*, edited by Maaike Groot, Daphne Lentjes, and Jørn Zeiler, 59–80. Leiden: Sidestone Press.
- Carruthers, Denise. 2005. "Hunting and Herding in Central Anatolian Prehistory: The Sites at Pinarbasi." Book Section. In *Archaeozoology of the Near East VI*.
- . 2006. "Pinarbaşı 1994: Animal Bones." Dataset. <http://opencontext.org/projects/TESTPRJ0000000004>.
- Corti, Paulo, and David M. Shackleton. 2002. "Relationship Between Predation-Risk Factors and Sexual Segregation in Dall's Sheep (*Ovis Dalli Dalli*).". Journal Article. *Canadian Journal of Zoology* 80 (12): 2108–17. <https://doi.org/10.1139/z02-207>.
- Crema, Enrico R. 2011. "Modelling Temporal Uncertainty in Archaeological Analysis." Journal Article. *Journal of Archaeological Method and Theory* 19 (3): 440–61. <https://doi.org/10.1007/s10816-011-9122-3>.

- Dahl, Gudrun, and Anders Hjort. 1976. *Having Herds: Pastoral Herd Growth and Household Economy*. Book. Stockholm Studies in Social Anthropology. Stockholm: University of Stokholm.
- Daly, K. G., V. Mattiangeli, A. J. Hare, H. Davoudi, H. Fathi, S. B. Doost, S. Amiri, et al. 2021. “Hherded and Hunted Goat Genomes from the Dawn of Domestication in the Zagros Mountains.” Journal Article. *Proc Natl Acad Sci U S A* 118 (25). <https://doi.org/10.1073/pnas.2100901118>.
- Davis, Simon J. M. 1981. “The Effects of Temperature Change and Domestication on the Body Size of Late Pleistocene to Holocene Mammals of Israel.” Journal Article. *Paleobiology* 7 (1): 101–14.
- . 1996. “Measurements of a Group of Adult Female Shetland Sheep Skeletons from a Single Flock: A Baseline for Zooarchaeologists.” Journal Article. *Journal of Archaeological Science* 23: 593–612.
- Degerbøl, Magnus. 1970. “The Urus (*Bos Primigenius Bojanus*) and Neolithic Domesticated Cattle (*Bos Taurus Domesticus Linné*) in Denmark.” Journal Article. *Det Kongelige Danske Videnskabernes Selskab Biologiske Skrifter* 17 (1): 5–177.
- Dong, Zhuan. 1997. “Mixture Analysis and Its Preliminary Application in Archaeology.” Journal Article. *Journal of Archaeological Science* 24: 141–61.
- Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of ‘Data.frame’*. <https://CRAN.R-project.org/package=data.table>.
- Driesch, Angela von den. 1976. *A Guide to the Measurement of Animal Bones from Archaeological Sites*. Book. Peabody Museum Bulletins. Cambridge, MA: Harvard University.
- Evershed, Richard P., Sebastian Payne, Andrew G. Sherratt, Mark S. Copley, Jennifer Coolidge, Duska Urem-Kotsu, Kostas Kotsakis, et al. 2008. “Earliest Date for Milk Use in the Near East and Southeastern Europe Linked to Cattle Herding.” Journal Article. *Nature* 455 (7212): 528–31. <https://doi.org/10.1038/nature07180>.
- Fernández, Hélène, and Hervé Monchot. 2007. “Sexual Dimorphism in Limb Bones of Ibex (*Capra Ibex*): Mixture Analysis Applied to Modern and Fossil Data.” Journal Article. *International Journal of Osteoarchaeology* 17: 479–91. <https://doi.org/10.1002/oa.876>.
- Gabry, Jonah, and Rok Cešnovar. 2022. *Cmdstanr: R Interface to ‘CmdStan’*.
- Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. “Visualization in Bayesian Workflow.” Journal Article. *Journal of the Royal Statistical Society, Series A* 182 (2): 389–402.
- Galik, Alfred. 2013. “Barçın Höyük Zooarchaeology.” Dataset. <http://opencontext.org/projects/74749949-4FD4-4C3E-C830-5AA75703E08E>. <https://doi.org/10.6078/M78G8HM0>.
- Gelman, Andrew. 2006a. “Multilevel (Hierarchical) Modeling: What It Can and Cannot Do.” Journal Article. *Technometrics* 48 (3): 432–35. <https://doi.org/10.1198/004017005000000661>.

- . 2006b. “Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper).” Journal Article. *Bayesian Analysis* 3: 515–34.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2020. *Bayesian Data Analysis*. Book. Third Edition.
- Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. “A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models.” Journal Article. *The Annals of Applied Statistics* 2 (4): 1360–83. <https://doi.org/10.1214/08-aoas191>.
- Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. “Bayesian Workflow.” Journal Article. *ArXiv*.
- Gerritsen, Fokke, and Rana Özbal. 2019. “Barcin Höyük, a Seventh Millennium Settlement in the Eastern Marmara Region of Turkey.” Journal Article. *Documenta Praehistorica* 46: 58–67. <https://doi.org/10.4312/dp.46.4>.
- Gillis, R., R. M. Arbogast, J. F. Piningre, K. Debue, and J. D. Vigne. 2015. “Prediction Models for Age-at-Death Estimates for Calves, Using Unfused Epiphyses and Diaphyses.” Journal Article. *International Journal of Osteoarchaeology* 25 (6): 912–22. <https://doi.org/10.1002/oa.2377>.
- Gillis, R., I. Carrère, M. Saña Seguí, G. Radi, and J. D. Vigne. 2014. “Neonatal Mortality, Young Calf Slaughter and Milk Production During the Early Neolithic of North Western Mediterranean.” Journal Article. *International Journal of Osteoarchaeology* 26 (2): 303–13. <https://doi.org/10.1002/oa.2422>.
- Gourichon, Lionel, and Daniel Helmer. 2008. “Etude de La Faune Neolithique de Menteşe.” Book Section. In *Life and Death in a Prehistoric Settlement in Northwest Anatolia. The Ilipinar Excavations, Volume III.*, edited by Jacob Roodenberg and Songül Alpaslan Roodenberg, 435–48. Leiden: Nederlands Instituut voor he Nabije Oosten.
- . 2014. “Faunal Data from Neolithic Menteşe.” Dataset. <http://opencontext.org/projects/42EAD4DB-BAED-4A58-9B9B-7EC85266D2A9>. <https://doi.org/https://doi.org/10.6078/M7S46PVN>.
- Grigson, Caroline. 1982. “Sex and Age Determination of Some Bones and Teeth of Domestic Cattle: A Review of the Literature.” Book Section. In *Ageing and Sexing Animal Bones from Archaeological Sites*, edited by Bob Wilson, Caroline Grigson, and Sebastian Payne, 7–23. BAR British Series. Oxford: British Archaeological Reports.
- . 1989. “Size and Sex: Evidence for the Domestication of Cattle in the Near East.” Book Section. In *People and Culture in Change: Preceedings of the 2nd Symposium on Upper Paleolithic, Mesolithic, and Mesolithic Populations of Europe and the Mediterranean Basin*, edited by Israel Hershkovitz, 77–109. BAR International Series. Oxford: Archaeopress.

- Helmer, Daniel, Lionel Gourichon, Hervé Monchot, Joris Peters, and Maria Saña Segui. 2005. "Identifying Early Domestic Cattle from Pre-Pottery Neolithic Sites on the Middle Euphrates Using Sexual Dimorphism." Book Section. In *First Steps of Animal Domestication: New Archaeozoological Approaches*, edited by Jean-Denis Vigne, Joris Peters, and Daniel Helmer, 86–95. Proceedings of the 9th Conference of the International Council of Archaeozoology, Durham, August 2002. Oxford: Oxbow Books.
- Jasra, A., C. C. Holmes, and D. A. Stephens. 2005. "Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling." Journal Article. *Statistical Science* 20 (1): 50–67. <https://doi.org/10.1214/088342305000000016>.
- Kabukcu, Ceren. 2017. "Woodland Vegetation History and Human Impacts in South-Central Anatolia 16,000–6500 Cal BP: Anthracological Results from Five Prehistoric Sites in the Konya Plain." Journal Article. *Quaternary Science Reviews* 176: 85–100. <https://doi.org/10.1016/j.quascirev.2017.10.001>.
- Karul, Necmi. 2019. "Early Farmers in Northwestern Anatolia in the Seventh Millennium." Book Section. In *Concluding the Neolithic: The Near East in the Second Half of the Seventh Millennium BC*, edited by Arkadiusz Marciniak, 269–86. Atlanta, Georgia: Lockwood Press.
- Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr>.
- Kay, Matthew. 2022. *ggdist: Visualizations of Distributions and Uncertainty*. <https://doi.org/10.5281/zenodo.3879620>.
- Lebenzon, Roxanne, and Natalie D. Munro. 2022. "Body Size Variation in a Modern Collection of Mountain Gazelle (*Gazella Gazella*) Skeletons." Journal Article. *Journal of Archaeological Science: Reports* 41. <https://doi.org/10.1016/j.jasrep.2021.103285>.
- Lyman, R. Lee. 2008. *Quantitative Paleozoology*. Book. Cambridge Manuals in Archaeology. Cambridge: Cambridge University Press.
- Madgwick, Richard, and Jacqui Mulville. 2015. "Reconstructing Depositional Histories Through Bone Taphonomy: Extending the Potential of Faunal Data." Journal Article. *Journal of Archaeological Science* 53: 255–63. <https://doi.org/10.1016/j.jas.2014.10.015>.
- Manning, Katie, Adrian Timpson, Stephen Shennan, and Enrico Crema. 2015. "Size Reduction in Early European Domestic Cattle Relates to Intensification of Neolithic Herding Strategies." Journal Article. *PLoS One* 10 (12): e0141873. <https://doi.org/10.1371/journal.pone.0141873>.
- Marom, Nimrod, and Guy Bar-Oz. 2013. "The Prey Pathway: A Regional History of Cattle (*Bos Taurus*) and Pig (*Sus Scrofa*) Domestication in the Northern Jordan Valley, Israel." Journal Article. *PLoS One* 8 (2): e55958. <https://doi.org/10.1371/journal.pone.0055958.g001>.
- Martín, Patricia, and Ricardo García-González. 2015. "Identifying Sheep (*Ovis Aries*) Fetal Remains in

- Archaeological Contexts.” Journal Article. *Journal of Archaeological Science* 64: 77–87. <https://doi.org/10.1016/j.jas.2015.10.003>.
- Martín, Patricia, Ricardo García-González, Jordi Nadal, and Josep Maria Vergès. 2015. “Perinatal Ovi-caprine Remains and Evidence of Shepherding Activities in Early Holocene Enclosure Caves: El Mirador (Sierra de Atapuerca, Spain).” Journal Article. *Quaternary International*. <https://doi.org/10.1016/j.quaint.2015.08.024>.
- Marwick, Ben. 2017. “Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation.” Journal Article. *Journal of Archaeological Method and Theory* 24 (2): 424–50. <https://doi.org/10.1007/s10816-015-9272-9>.
- Marwick, Ben, and Suzanne E. Pilaar Birch. 2018. “A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing.” Journal Article. *Advances in Archaeological Practice* 6 (02): 125–43. <https://doi.org/10.1017/aap.2018.3>.
- McCarthy, Michael A., and P. I. P. Masters. 2005. “Profiting from Prior Information in Bayesian Analyses of Ecological Data.” Journal Article. *Journal of Applied Ecology* 42 (6): 1012–19. <https://doi.org/10.1111/j.1365-2664.2005.01101.x>.
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. Book. Texts in Statistical Science. Boca Raton, FL: CRC Press.
- McGrory, S., E. M. Svensson, A. Götherström, J. Mulville, A. J. Powell, M. J. Collins, and T. P. O’Connor. 2012. “A Novel Method for Integrated Age and Sex Determination from Archaeological Cattle Mandibles.” Journal Article. *Journal of Archaeological Science* 39 (10): 3324–30. <https://doi.org/10.1016/j.jas.2012.05.021>.
- Meadow, Richard. 1999. “The Use of Size Index Scaling Techniques for Research on Archaeozoological Collections from the Middle East.” Book Section. In *Historia Animalium Ex Ossibus: Beiträge Zur Paläoanatomie, Archäologie, Ägyptologie, Ethnologie Und Geschichte Der Tiermedizin*, edited by Cornelia Becker, Henriette Manhart, Joris Peters, and Jörg Schibler, 285–300. Leidorf: Verlag Marie.
- Meier, Jacqueline. 2020. “The Contextual Taphonomy of Middens at Neolithic Kfar HaHoresh.” Journal Article. *Journal of Archaeological Science: Reports* 33. <https://doi.org/10.1016/j.jasrep.2020.102531>.
- Microsoft Corporation, and Steve Weston. 2022. *doParallel: Foreach Parallel Adaptor for the 'Parallel' Package*. <https://CRAN.R-project.org/package=doParallel>.
- Milner, J. M., E. B. Nilsen, and H. P. Andreassen. 2007. “Demographic Side Effects of Selective Hunting in Ungulates and Carnivores.” Journal Article. *Conserv Biol* 21 (1): 36–47. <https://doi.org/10.1111/j.1523-1739.2006.00591.x>.
- Monchot, Hervé, and Jacques Léchelle. 2002. “Statistical Nonparametric Methods for the Study of Fossil

- 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).
- Otárola-Castillo, Erik, and Melissa G. Torquato. 2018. “Bayesian Statistics in Archaeology.” Journal Article. *Annual Review of Anthropology* 47 (1). <https://doi.org/10.1146/annurev-anthro-102317-045834>.
- Otárola-Castillo, Erik, Melissa G. Torquato, Jesse Wolfhagen, Matthew E. Hill, and Caitlin E. Buck. 2022. “Beyond Chronology, Using Bayesian Inference to Evaluate Hypotheses in Archaeology.” Journal Article. *Advances in Archaeological Practice*.
- Özdoğan, Mehmet. 2011. “Archaeological Evidence on the Westward Expansion of Farming Communities from Eastern Anatolia to the Aegean and the Balkans.” Journal Article. *Current Anthropology* 52 (S4): S415–30. <https://doi.org/10.1086/658895>.
- . 2019. “Early Farmers in Northwestern Turkey: What Is New?” Book Section. In *Concluding the Neolithic: The Near East in the Second Half of the Seventh Millennium BC*, edited by Arkadiusz Marciniak, 307–28. Atlanta, Georgia: Lockwood Press.
- Payne, Sebastian. 1969. “A Metrical Distinction Between Sheep and Goat Metacarpals.” Book Section. In *The Domestication and Exploitation of Plants and Animals*, edited by Peter J. Ucko and Geoffrey W. Dimbleby, 295–305. London: Duckworth.
- . 1973. “Kill-Off Patterns in Sheep and Goats: The Mandibles from Aşvan Kale.” Journal Article. *Anatolian Studies* 23: 281–303.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pérez-Barbería, F. Javier, Iain J. Gordon, and M. Pagel. 2002. “The Origins of Sexual Dimorphism in Body Size in Ungulates.” Journal Article. *Evolution* 56 (6): 1276–85.
- Peters, Joris, Angela von den Driesch, and Daniel Helmer. 2005. “The Upper Euphrates-Tigris Basin: Cradle of Agro-Pastoralism?” Book Section. In *The First Steps of Animal Domestication*, edited by Jean-Denis Vigne, Joris Peters, and Daniel Helmer, 96–124. Proceedings of the 9th Conference of the International Council of Archaeozoology, Durham, August 2002. Oxford: Oxbow Books.
- Popkin, Peter R. W., Polydora Baker, Fay Worley, Sebastian Payne, and Andy Hammon. 2012. “The Sheep Project (1): Determining Skeletal Growth, Timing of Epiphyseal Fusion and Morphometric Variation in Unimproved Shetland Sheep of Known Age, Sex, Castration Status and Nutrition.” Journal Article. *Journal of Archaeological Science* 39 (6): 1775–92. <https://doi.org/10.1016/j.jas.2012.01.018>.
- Post, Diane M., Trent S. Armbrust, Eva A. Horne, and Jacob R. Goheen. 2001. “Sexual Segregation Results in Differences in Content and Quality of Bison (*Bos Bison*) Diets.” Journal Article. *Journal of Mammalogy* 82 (2): 407–13.

- Pozo, Jose M, Valenzuela-Lamas, Silvia, Trentacoste, Angela, Nieto-Espinet, Ariadna, Guimarães Chiarelli, and Silvia. 2021. “Zoolog: Zooarchaeological Analysis with Log-Ratios.” *Journal of Statistical Software* In preparation. <https://josempozo.github.io/zoolog/>.
- Proaktor, G., T. Coulson, and E. J. Milner-Gulland. 2007. “Evolutionary Responses to Harvesting in Ungulates.” Journal Article. *J Anim Ecol* 76 (4): 669–78. <https://doi.org/10.1111/j.1365-2656.2007.01244.x>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Redding, Richard W. 1984. “Theoretical Determinants of a Herder’s Decisions: Modeling Variation in the Sheep/Goat Ratio.” Book Section. In *Animals and Archaeology*, edited by Juliet Clutton-Brock and Caroline Grigson, 3. Early Herders and their Flocks:223–41. BAR International Series. Oxford: British Archaeological Reports.
- Roodenberg, Jacob. 2012a. “Change in Food Production and Its Impact on an Early 6th Millennium Community in Northwest Anatolia. The Example of Ilipinar.” Journal Article. *Prähistorische Zeitschrift* 87 (2). <https://doi.org/10.1515/pz-2012-0015>.
- . 2012b. “Ilipinar: A Neolithic Settlement in the Eastern Marmara Region.” Book Section. In *The Oxford Handbook of Ancient Anatolia (10,000-323 BCE)*, edited by Gregory McMahon and Sharon Steadman. <https://doi.org/10.1093/oxfordhb/9780195376142.013.0044>.
- Roodenberg, Jacob, A. van As, L. Jacobs, and M. H. Wijnen. 2003. “Early Settlement in the Plain of Yenisehir (NW Anatolia): The Basal Occupation Layers at Mentese.” Journal Article. *Anatolica* XXIX (1): 18–58.
- RStudio Team. 2022. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. <http://www.rstudio.com/>.
- Ruckstuhl, K. E. 2007. “Sexual Segregation in Vertebrates: Proximate and Ultimate Causes.” Journal Article. *Integr Comp Biol* 47 (2): 245–57. <https://doi.org/10.1093/icb/icm030>.
- Ruckstuhl, K. E., and P. Neuhaus. 2002. “Sexual Segregation in Ungulates: A Comparative Test of Three Hypotheses.” Journal Article. *Biological Reviews* 77 (1): 77–96. <https://doi.org/10.1017/s1464793101005814>.
- Said, Sonia, Vincent Tolon, Serge Brandt, and Eric Baubet. 2011. “Sex Effect on Habitat Selection in Response to Hunting Disturbance: The Study of Wild Boar.” Journal Article. *European Journal of Wildlife Research* 58 (1): 107–15. <https://doi.org/10.1007/s10344-011-0548-4>.
- Smith, A. F. M., and Alan E. Gelfand. 1992. “Bayesian Statistics Without Tears: A Sampling-Resampling Perspective.” Journal Article. *The American Statistician* 46 (2): 84–88.

- South, Andy. 2017a. *Rnaturalearth: World Map Data from Natural Earth*. <https://CRAN.R-project.org/package=rnaturalearth>.
- . 2017b. *Rnaturalearthdata: World Vector Map Data from Natural Earth Used in 'Rnaturalearth'*. <https://CRAN.R-project.org/package=rnaturalearthdata>.
- Speth, John D. 2013. “Thoughts about Hunting: Some Things We Know and Some Things We Don’t Know.” Journal Article. *Quaternary International* 297: 176–85. <https://doi.org/10.1016/j.quaint.2012.12.005>.
- Stan Development Team. 2021. “RStan: The R Interface to Stan.” <https://mc-stan.org/>.
- . 2022. “Stan Modeling Language Users Guide and Reference Manual, Version 2.29.” Generic. <https://mc-stan.org/>.
- Stiner, Mary C. 1990. “The Use of Mortality Patterns in Archaeological Studies of Hominid Predatory Adaptations.” Journal Article. *Journal of Anthropological Archaeology* 9: 305–51.
- Tchernov, Eitan, and Liora Kolska Horwitz. 1991. “Body Size Diminution Under Domestication: Unconscious Selection in Primeval Domesticates.” Journal Article. *Journal of Anthropological Archaeology* 10: 54–75.
- Thissen, Laurens, Hadi Özbal, Ayla Türkekul Büyik, Fokke Gerritsen, and Rana Özbal. 2010. “The Land of Milk? Approaching Dietary Preferences of Late Neolithic Communities in NW Anatolia.” Journal Article. *Leiden Journal of Pottery Studies* 26: 157–72.
- Twiss, Katheryn C. 2019. *The Archaeology of Food*. Book. <https://doi.org/10.1017/9781108670159>.
- Uerpmann, Margarethe, and Hans-Peter Uerpmann. 1994. “Animal Bones.” Book Section. In *Qala’at Al-Bahrain*, edited by Flemming Hojlund and H. Hellmuth Andersen, 1: The Northern City Wall and the Islamic Fortress:417–54. Jutland Archaeological Society Publications. Aarhus: Aarhus University Press.
- Urbanek, Simon, and Jeffrey Horner. 2022. *Cairo: R Graphics Device Using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output*. <https://CRAN.R-project.org/package=Cairo>.
- Watkins, Trevor. 1996. “Excavations at Pınarbaşı: The Early Stages.” Book Section. In *On the Surface: Çatalhöyük 1993-1995*, edited by Ian Hodder, 47–57. Cambridge: McDonald Institute for Archaeological Research.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wolfhagen, Jesse. 2020. “Re-Examining the Use of the LSI Technique in Zooarchaeology.” Journal Article. *Journal of Archaeological Science* 123: 105254. <https://doi.org/10.1016/j.jas.2020.105254>.
- Wolfhagen, Jesse, Katheryn C. Twiss, Jacqui A. Mulville, and G. Arzu Demirergi. 2021. “Examining Caprine Management and Cattle Domestication Through Biometric Analyses at Çatalhöyük East (North

- and South Areas)." Book Section. In *Peopling the Landscape of Çatalhöyük: Reports from the 2009-2017 Seasons*, edited by Ian Hodder, 181–98. Los Angeles: Cotsen Institute of Archaeology Press.
- Wright, Elizabeth, and Sarah Viner-Daniels. 2015. "Geographical Variation in the Size and Shape of the European Aurochs (*Bos Primigenius*).” Journal Article. *Journal of Archaeological Science* 54: 8–22. <https://doi.org/10.1016/j.jas.2014.11.021>.
- Zeder, Melinda A. 2012. "Pathways to Animal Domestication." Book Section. In *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*, edited by P. Gepts, T. R. Famula, and R. L. Bettinger, 227–59. Cambridge: Cambridge University Press.
- Zeder, Melinda A., and Brian Hesse. 2000. "The Initial Domestication of Goats (*Capra Hircus*) in the Zagros Mountains 10,000 Years Ago." Journal Article. *Science* 287 (5461): 2254–57.
- Zeder, Melinda A., and Ximena Lemoine. 2020. "A Method for Constructing Demographic Profiles in *Sus Scrofa* Using Logarithm Size Index Scaling." Journal Article. *Journal of Archaeological Science* 116. <https://doi.org/10.1016/j.jas.2020.105115>.

Appendix 1 (Supplemental Table 6)

Posterior Summary Tables for Overall and Site-Level Model Parameters: Simulated Assemblages

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 π_1	0.15	0.15	0.04	0.10	0.22	1	3396	2582
Single Assemblage Model π_2	0.52	0.53	0.09	0.37	0.66	1	2806	3088
Single Assemblage Model π_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 π_1	0.16	0.16	0.02	0.12	0.19	1	3255	2865
Multisite Model π_2	0.47	0.47	0.06	0.37	0.56	1	2938	2854
Multisite Model π_3	0.38	0.38	0.06	0.28	0.48	1	2925	3047
Multisite Model μ_1	-0.10	-0.10	0.04	-0.17	-0.03	1	1357	1893
Multisite Model μ_2	-0.08	-0.08	0.04	-0.15	-0.01	1	1365	1867
Multisite Model μ_3	-0.01	-0.01	0.04	-0.08	0.07	1	1437	2207
Multisite Model σ_1	0.04	0.04	0.01	0.03	0.06	1	4776	3070
Multisite Model σ_2	0.04	0.04	0.00	0.03	0.05	1	3295	2950
Multisite Model σ_3	0.04	0.04	0.01	0.03	0.04	1	2266	3160

Table 8: Posterior fit summaries for site-specific 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 π_1	0.16	0.16	0.04	0.11	0.24	1	3968	3237
Site 2 π_1	0.48	0.49	0.10	0.31	0.65	1	3710	3496
Site 3 π_1	0.35	0.35	0.11	0.18	0.54	1	3992	3452
Site 4 π_1	0.18	0.17	0.05	0.11	0.27	1	4769	3036
Site 5 π_1	0.45	0.45	0.11	0.25	0.62	1	5229	3519

Table 8: Posterior fit summaries for site-specific 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 π_1	0.38	0.37	0.12	0.20	0.58	1	5383	3275
Site 7 π_1	0.14	0.14	0.04	0.08	0.20	1	3156	3556
Site 8 π_1	0.32	0.32	0.10	0.15	0.48	1	2188	2512
Site 9 π_1	0.54	0.53	0.10	0.38	0.71	1	2388	2859
Site 10 π_1	0.18	0.18	0.04	0.12	0.26	1	3658	3517
Site 11 π_1	0.45	0.46	0.10	0.28	0.61	1	4484	3839
Site 12 π_1	0.36	0.36	0.10	0.20	0.54	1	4906	3862
Site 13 π_1	0.15	0.15	0.04	0.09	0.21	1	3318	3642
Site 14 π_1	0.47	0.47	0.07	0.34	0.59	1	3887	3902
Site 15 π_1	0.38	0.38	0.07	0.26	0.50	1	4156	3361
Site 1 π_2	0.16	0.16	0.04	0.10	0.23	1	4480	3787
Site 2 π_2	0.59	0.59	0.10	0.41	0.74	1	3737	3305
Site 3 π_2	0.25	0.24	0.10	0.10	0.44	1	3704	3116
Site 4 π_2	0.16	0.16	0.04	0.10	0.24	1	5010	3631
Site 5 π_2	0.54	0.55	0.11	0.35	0.72	1	4399	3306
Site 6 π_2	0.29	0.28	0.12	0.11	0.51	1	4859	3441
Site 7 π_2	0.16	0.16	0.04	0.10	0.23	1	4094	3673
Site 8 π_2	0.53	0.53	0.10	0.35	0.69	1	3459	3065
Site 9 π_2	0.32	0.31	0.11	0.15	0.50	1	3762	3167
Site 10 π_2	0.19	0.19	0.05	0.13	0.28	1	3056	3571
Site 11 π_2	0.56	0.57	0.09	0.40	0.71	1	3889	3347
Site 12 π_2	0.25	0.24	0.09	0.11	0.41	1	4298	2874
Site 13 π_2	0.26	0.25	0.07	0.16	0.38	1	1760	2220
Site 14 π_2	0.61	0.61	0.08	0.48	0.73	1	2286	2320
Site 15 π_2	0.13	0.12	0.05	0.05	0.23	1	3240	2679

Table 8: Posterior fit summaries for site-specific 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 1 π_3	0.14	0.14	0.04	0.08	0.20	1	3172	3537
Site 2 π_3	0.46	0.46	0.11	0.28	0.63	1	4981	3387
Site 3 π_3	0.40	0.40	0.11	0.23	0.59	1	4854	3218
Site 4 π_3	0.12	0.12	0.04	0.05	0.18	1	3029	3113
Site 5 π_3	0.37	0.37	0.13	0.16	0.57	1	4797	3520
Site 6 π_3	0.51	0.51	0.13	0.29	0.74	1	4726	3414
Site 7 π_3	0.14	0.14	0.04	0.08	0.20	1	2730	3163
Site 8 π_3	0.33	0.32	0.11	0.14	0.52	1	3988	3599
Site 9 π_3	0.54	0.54	0.12	0.34	0.73	1	4545	3313
Site 10 π_3	0.14	0.14	0.04	0.08	0.21	1	3137	3113
Site 11 π_3	0.38	0.38	0.10	0.21	0.55	1	3594	3372
Site 12 π_3	0.48	0.47	0.11	0.30	0.66	1	3800	3123
Site 13 π_3	0.14	0.14	0.04	0.08	0.19	1	2782	3105
Site 14 π_3	0.30	0.30	0.08	0.18	0.43	1	4015	3721
Site 15 π_3	0.56	0.57	0.08	0.43	0.69	1	4046	3717
Site 1 μ_1	-0.08	-0.08	0.02	-0.12	-0.05	1	2861	2874
Site 2 μ_1	-0.08	-0.08	0.03	-0.13	-0.04	1	3426	3472
Site 3 μ_1	0.08	0.08	0.02	0.04	0.12	1	2464	3081
Site 4 μ_1	-0.30	-0.30	0.02	-0.34	-0.27	1	3619	3647
Site 5 μ_1	-0.10	-0.10	0.02	-0.13	-0.07	1	2861	3197
Site 6 μ_1	-0.12	-0.12	0.02	-0.15	-0.09	1	3034	3112
Site 7 μ_1	-0.09	-0.09	0.02	-0.13	-0.05	1	3465	3417
Site 8 μ_1	0.07	0.07	0.02	0.04	0.10	1	3184	3351
Site 9 μ_1	-0.34	-0.34	0.02	-0.38	-0.31	1	3044	3515
Site 10 μ_1	-0.11	-0.11	0.02	-0.13	-0.08	1	2910	3389

Table 8: Posterior fit summaries for site-specific 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 11 μ_1	-0.08	-0.07	0.02	-0.11	-0.04	1	3636	3474
Site 12 μ_1	-0.09	-0.09	0.03	-0.14	-0.04	1	3030	3289
Site 13 μ_1	0.09	0.09	0.02	0.05	0.12	1	3335	3627
Site 14 μ_1	-0.32	-0.32	0.02	-0.36	-0.28	1	3122	3479
Site 15 μ_1	-0.09	-0.09	0.02	-0.13	-0.06	1	3190	3404
Site 1 μ_2	-0.07	-0.07	0.02	-0.10	-0.03	1	2779	2995
Site 2 μ_2	-0.07	-0.07	0.03	-0.12	-0.02	1	3399	3471
Site 3 μ_2	0.10	0.10	0.02	0.06	0.14	1	2366	3030
Site 4 μ_2	-0.29	-0.29	0.02	-0.32	-0.26	1	3391	3483
Site 5 μ_2	-0.09	-0.09	0.02	-0.11	-0.06	1	2847	3078
Site 6 μ_2	-0.11	-0.11	0.02	-0.13	-0.08	1	2775	3035
Site 7 μ_2	-0.07	-0.07	0.02	-0.11	-0.04	1	3309	3298
Site 8 μ_2	0.09	0.09	0.02	0.06	0.12	1	3008	3269
Site 9 μ_2	-0.33	-0.33	0.02	-0.36	-0.30	1	2865	3329
Site 10 μ_2	-0.09	-0.09	0.02	-0.12	-0.06	1	2780	3165
Site 11 μ_2	-0.06	-0.06	0.02	-0.09	-0.03	1	3377	3501
Site 12 μ_2	-0.07	-0.07	0.03	-0.12	-0.03	1	2970	3227
Site 13 μ_2	0.10	0.11	0.02	0.07	0.14	1	3121	3314
Site 14 μ_2	-0.30	-0.30	0.02	-0.35	-0.27	1	2994	3138
Site 15 μ_2	-0.07	-0.07	0.02	-0.11	-0.04	1	3318	3491
Site 1 μ_3	-0.01	-0.01	0.02	-0.05	0.03	1	2904	3420
Site 2 μ_3	0.02	0.02	0.03	-0.03	0.08	1	3140	3047
Site 3 μ_3	0.19	0.19	0.03	0.15	0.23	1	1831	2511
Site 4 μ_3	-0.23	-0.23	0.02	-0.27	-0.19	1	3576	3343
Site 5 μ_3	0.17	0.17	0.02	0.14	0.20	1	2774	3329

Table 8: Posterior fit summaries for site-specific 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	3188	3353
Site 7 μ_3	-0.01	-0.01	0.05	-0.06	0.07	1	3619	3217
Site 8 μ_3	0.15	0.14	0.03	0.11	0.20	1	3127	3340
Site 9 μ_3	-0.27	-0.27	0.03	-0.31	-0.22	1	2744	3305
Site 10 μ_3	0.15	0.16	0.05	0.08	0.21	1	2159	1182
Site 11 μ_3	-0.01	-0.01	0.02	-0.04	0.03	1	3470	3394
Site 12 μ_3	0.00	0.00	0.03	-0.04	0.05	1	3553	3559
Site 13 μ_3	0.15	0.15	0.02	0.12	0.19	1	3248	3240
Site 14 μ_3	-0.24	-0.24	0.02	-0.27	-0.20	1	3331	3400
Site 15 μ_3	0.18	0.18	0.02	0.15	0.21	1	2887	3498
Site 1 σ_1	0.04	0.04	0.01	0.03	0.06	1	4266	3321
Site 2 σ_1	0.04	0.04	0.01	0.03	0.06	1	4280	3494
Site 3 σ_1	0.05	0.04	0.01	0.03	0.07	1	4165	3582
Site 4 σ_1	0.04	0.04	0.01	0.03	0.06	1	4368	3482
Site 5 σ_1	0.04	0.04	0.01	0.03	0.06	1	4479	3608
Site 6 σ_1	0.04	0.04	0.01	0.02	0.06	1	4023	3624
Site 7 σ_1	0.04	0.04	0.01	0.03	0.06	1	4212	3481
Site 8 σ_1	0.04	0.04	0.01	0.02	0.06	1	3836	3464
Site 9 σ_1	0.04	0.04	0.01	0.03	0.06	1	4282	3871
Site 10 σ_1	0.04	0.04	0.01	0.02	0.06	1	3347	3434
Site 11 σ_1	0.04	0.04	0.01	0.02	0.06	1	3912	3349
Site 12 σ_1	0.05	0.04	0.01	0.03	0.07	1	4401	3417
Site 13 σ_1	0.05	0.04	0.01	0.03	0.07	1	4424	3095
Site 14 σ_1	0.05	0.04	0.01	0.03	0.07	1	4547	3763
Site 15 σ_1	0.05	0.05	0.01	0.03	0.07	1	4199	3302

Table 8: Posterior fit summaries for site-specific 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	3000	2738
Site 2 σ_2	0.04	0.04	0.01	0.03	0.06	1	3048	3126
Site 3 σ_2	0.04	0.04	0.01	0.03	0.06	1	3505	3330
Site 4 σ_2	0.04	0.04	0.01	0.03	0.06	1	3421	3416
Site 5 σ_2	0.04	0.04	0.01	0.02	0.05	1	3356	2734
Site 6 σ_2	0.04	0.04	0.01	0.02	0.05	1	2406	2840
Site 7 σ_2	0.04	0.04	0.01	0.03	0.05	1	2968	3373
Site 8 σ_2	0.04	0.04	0.01	0.02	0.05	1	3047	3211
Site 9 σ_2	0.04	0.04	0.01	0.03	0.05	1	2708	2237
Site 10 σ_2	0.03	0.03	0.01	0.02	0.05	1	2740	3260
Site 11 σ_2	0.04	0.04	0.01	0.03	0.05	1	3421	3233
Site 12 σ_2	0.04	0.04	0.01	0.03	0.06	1	3745	3364
Site 13 σ_2	0.04	0.04	0.01	0.03	0.05	1	3507	3247
Site 14 σ_2	0.04	0.04	0.01	0.03	0.07	1	2526	3280
Site 15 σ_2	0.04	0.04	0.01	0.03	0.06	1	3829	2981
Site 1 σ_3	0.04	0.03	0.01	0.02	0.05	1	2482	2861
Site 2 σ_3	0.04	0.04	0.01	0.03	0.06	1	2623	2689
Site 3 σ_3	0.04	0.04	0.01	0.03	0.07	1	1457	2087
Site 4 σ_3	0.04	0.03	0.01	0.02	0.05	1	3098	2921
Site 5 σ_3	0.03	0.03	0.01	0.02	0.04	1	1783	2453
Site 6 σ_3	0.04	0.04	0.01	0.02	0.05	1	3349	2818
Site 7 σ_3	0.04	0.04	0.01	0.03	0.05	1	3119	2859
Site 8 σ_3	0.04	0.04	0.01	0.03	0.06	1	2563	3421
Site 9 σ_3	0.04	0.03	0.01	0.02	0.05	1	2519	2805
Site 10 σ_3	0.04	0.03	0.02	0.02	0.06	1	1730	1262

Table 8: Posterior fit summaries for site-specific 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	3443	3539
Site 12 σ_3	0.04	0.04	0.01	0.03	0.05	1	2306	2881
Site 13 σ_3	0.04	0.04	0.01	0.03	0.05	1	2947	3016
Site 14 σ_3	0.03	0.03	0.01	0.02	0.05	1	2178	2717
Site 15 σ_3	0.03	0.03	0.01	0.02	0.04	1	2331	3323

Appendix 2 (Supplemental Table 7)

Posterior Summary Tables for Overall and Site-Level Model Parameters: Archaeological Case 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 π_1	0.89	0.89	0.04	0.82	0.94	1	3681	2645
Pinarbaşı B Sheep π_2	0.07	0.07	0.04	0.02	0.14	1	2596	2523
Pinarbaşı B Sheep π_3	0.04	0.03	0.03	0.01	0.10	1	2696	2660
Pinarbaşı B Sheep μ_1	-0.13	-0.13	0.02	-0.16	-0.09	1	1753	2142
Pinarbaşı B Sheep μ_2	-0.10	-0.09	0.02	-0.13	-0.06	1	1709	1975
Pinarbaşı B Sheep μ_3	-0.03	-0.03	0.02	-0.07	0.01	1	1800	2043
Pinarbaşı B Sheep σ_1	0.05	0.05	0.00	0.04	0.06	1	5443	2963
Pinarbaşı B Sheep σ_2	0.05	0.04	0.00	0.04	0.05	1	5936	2987
Pinarbaşı B Sheep σ_3	0.05	0.05	0.00	0.04	0.05	1	5447	2295

Table 10: Posterior fit summaries for model parameters (Northwestern 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 π_1	0.18	0.17	0.07	0.09	0.32	1	1173	1907
NW Anatolian Cattle π_2	0.63	0.64	0.09	0.48	0.76	1	1340	2010
NW Anatolian Cattle π_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-specific model parameters (Northwestern Anatolian cattle)

Parameter	Posterior Mean	Posterior Median	Posterior SD	5% Quantile	95% Quantile	\hat{R}	Effective Sample Size (Bulk)	Effective Sample Size (Tail)
Barçın π_1	0.34	0.34	0.04	0.27	0.41	1.00	3928	3135
Neolithic İlpınar π_1	0.48	0.49	0.07	0.37	0.59	1.00	1566	2807
Menteşe π_1	0.18	0.17	0.06	0.09	0.30	1.00	1457	2836
Chalcolithic İlpınar π_1	0.13	0.13	0.02	0.10	0.17	1.00	3868	3520
Barçın π_2	0.72	0.72	0.08	0.58	0.85	1.01	830	1861
Neolithic İlpınar π_2	0.14	0.14	0.08	0.02	0.28	1.01	767	1604
Menteşe π_2	0.24	0.24	0.06	0.15	0.35	1.00	5014	3080
Chalcolithic İlpınar π_2	0.60	0.60	0.08	0.47	0.73	1.00	2866	3255
Barçın π_3	0.16	0.15	0.06	0.06	0.27	1.00	1683	2551
Neolithic İlpınar π_3	0.07	0.07	0.02	0.05	0.11	1.00	3012	2548
Menteşe π_3	0.70	0.71	0.05	0.62	0.78	1.00	1518	2591
Chalcolithic İlpınar π_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

Appendix 3 (Supplemental Figures 1-4)

Traceplots of Posterior Distributions of Overall Model Parameters

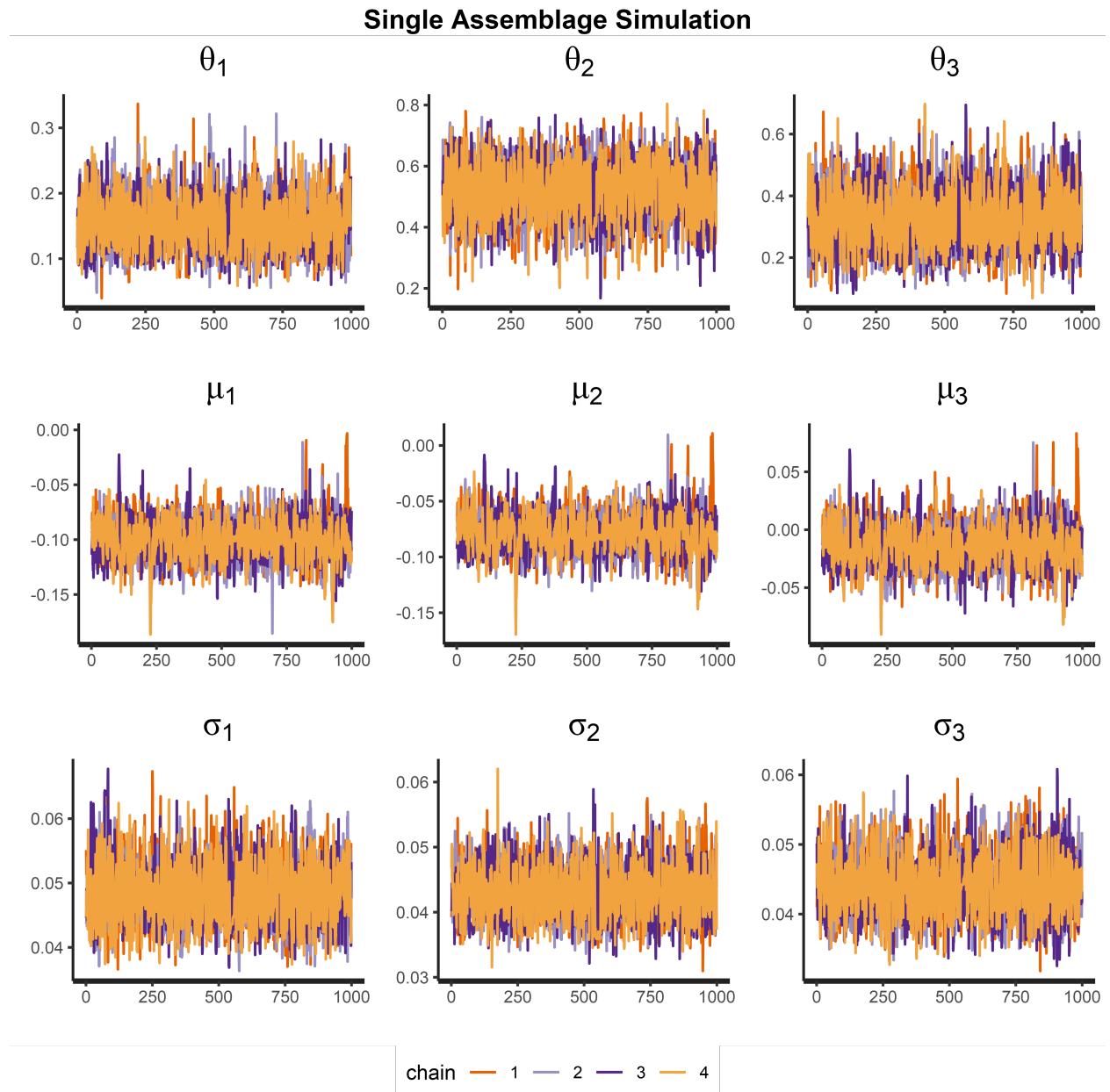


Figure 14: Traceplots of model parameters (single assemblage simulation)

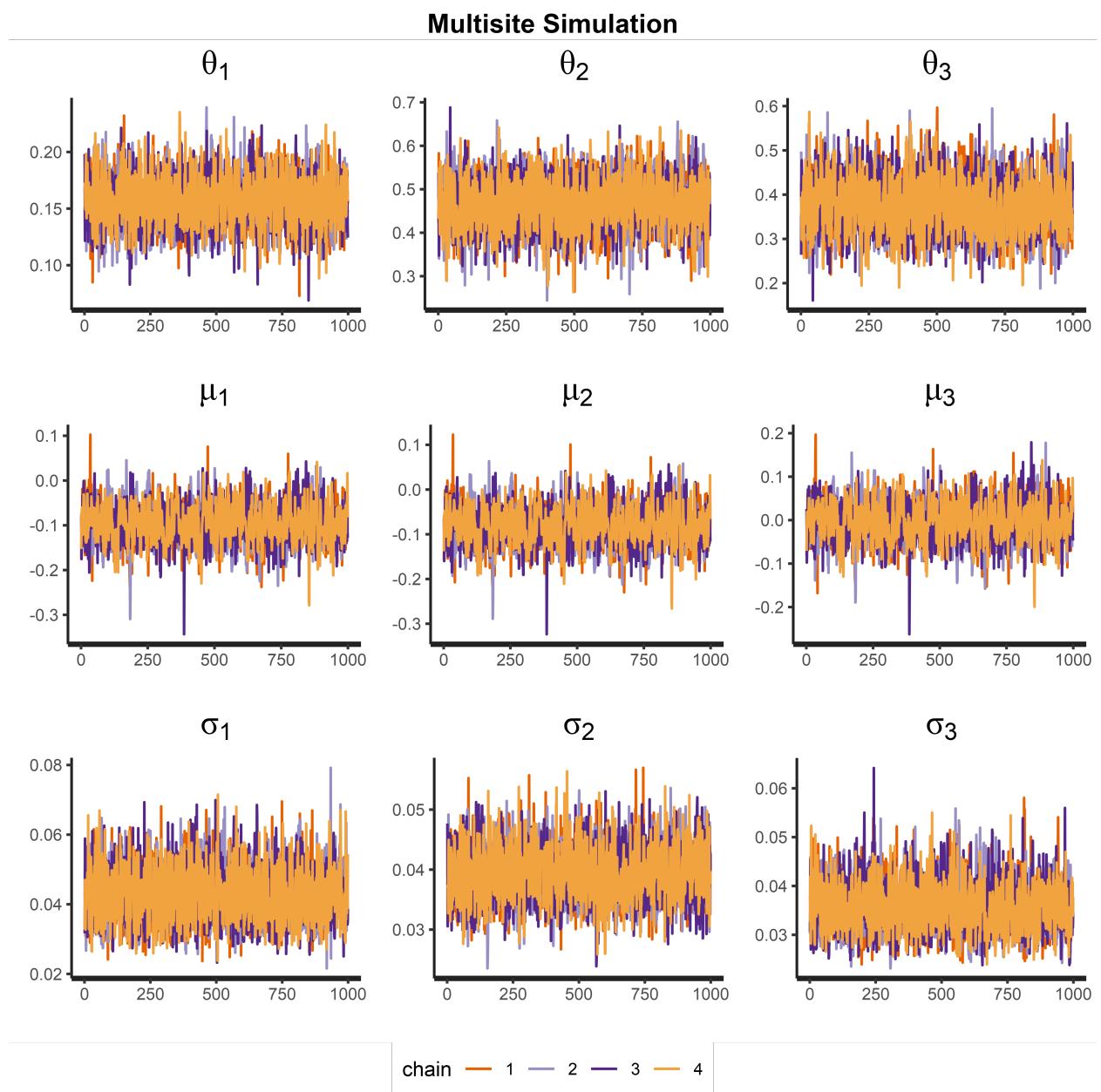


Figure 15: Traceplots of model parameters (multisite simulation)

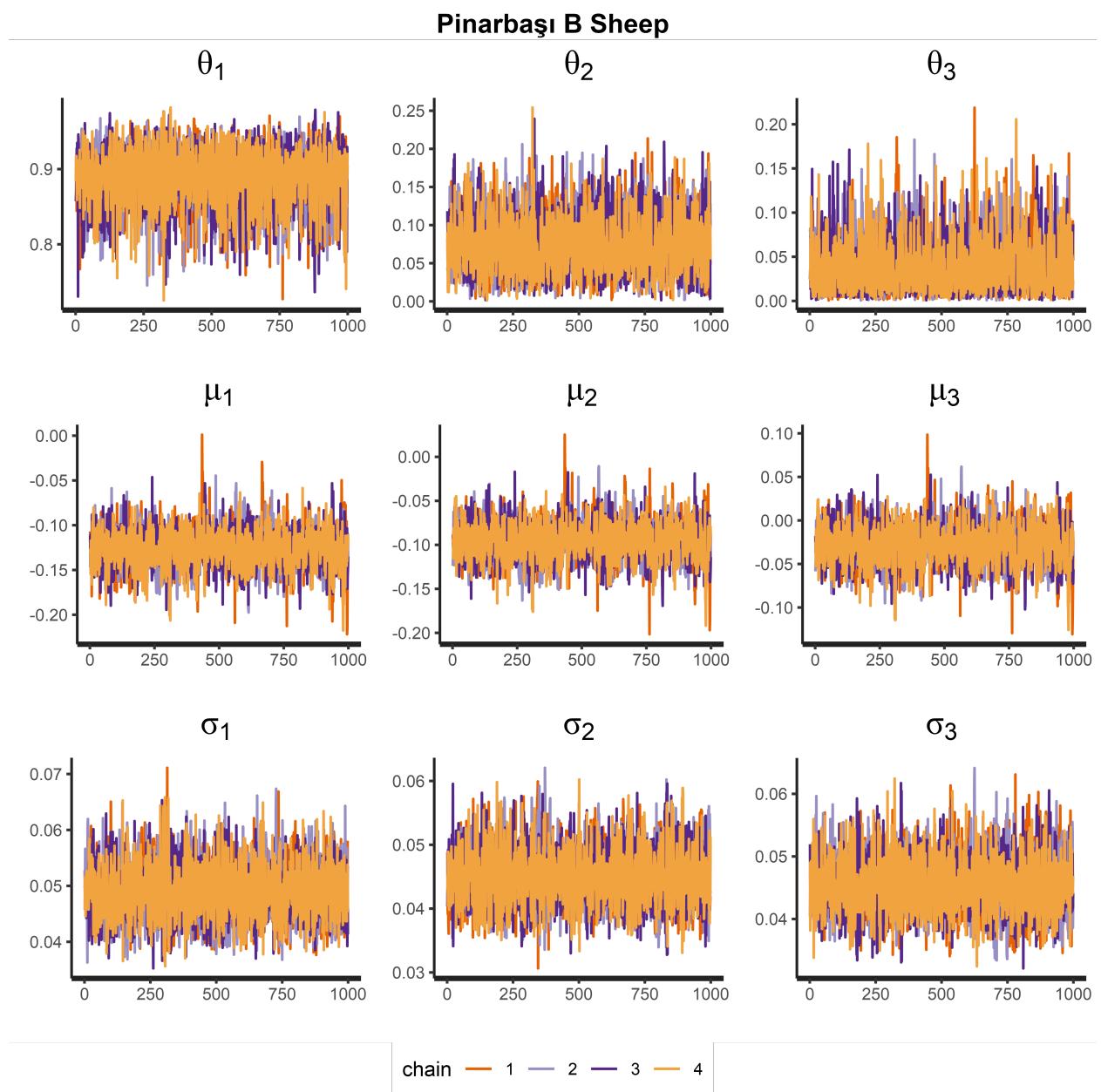


Figure 16: Traceplots of model parameters (Pınarbaşı B sheep)

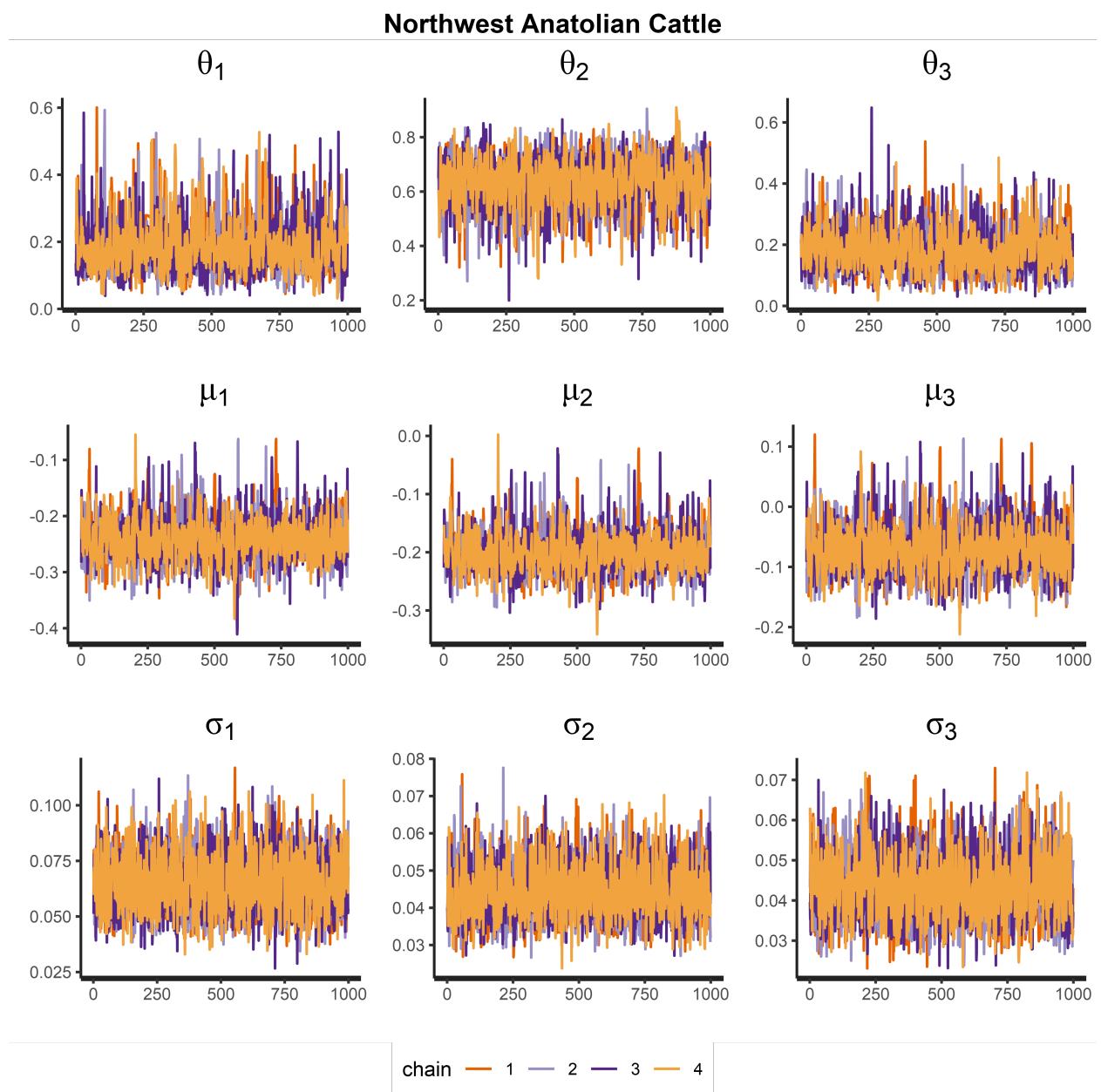


Figure 17: Traceplots of model parameters (Northwestern Anatolian cattle)

Appendix 4 (Supplemental Figures 5-7)

Prior-Posterior Comparisons for Model Hyper-Parameters

Multisite Simulation

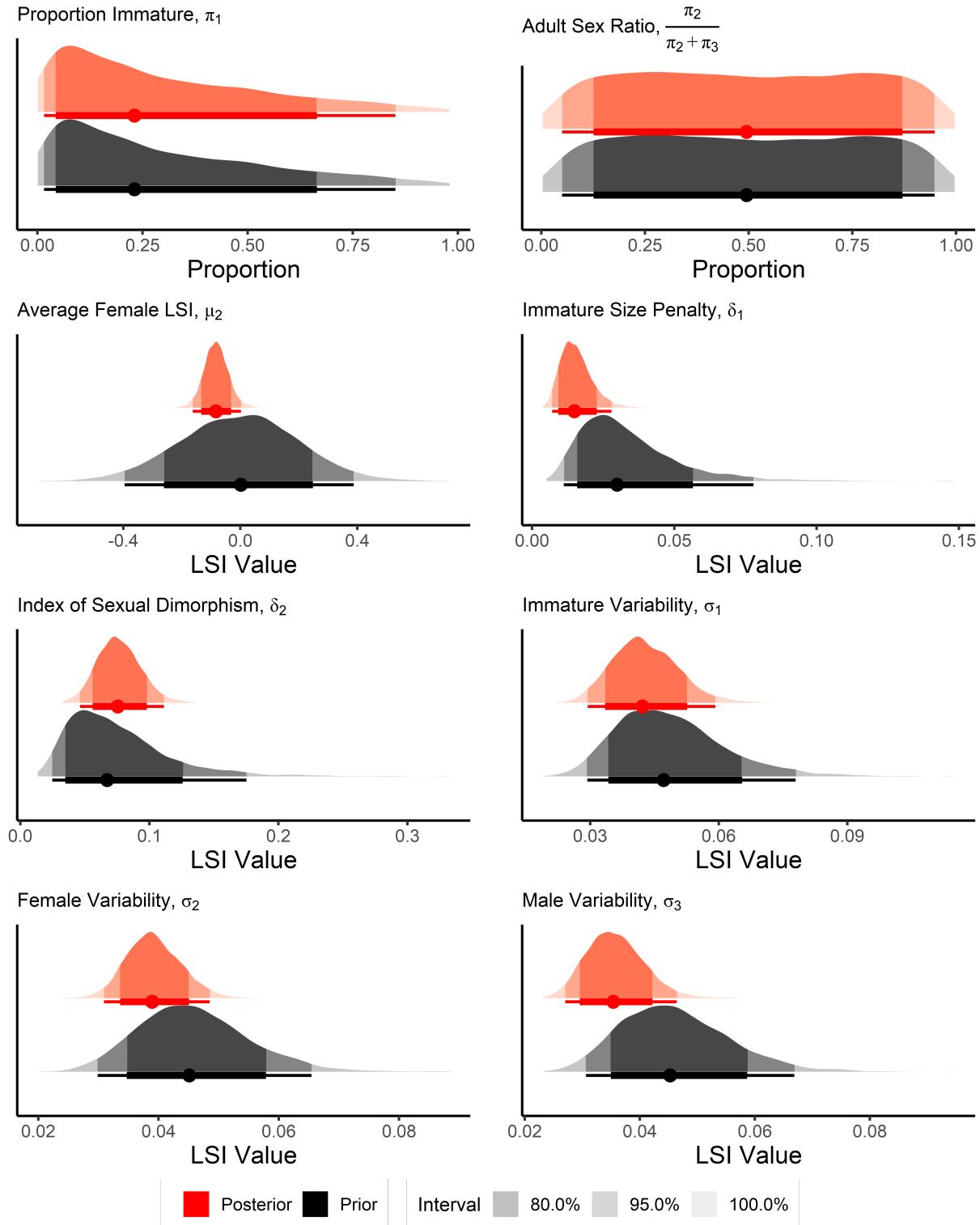


Figure 18: Prior-posterior comparison of multisite simulation model hyper-parameters

Pınarbaşı B Sheep

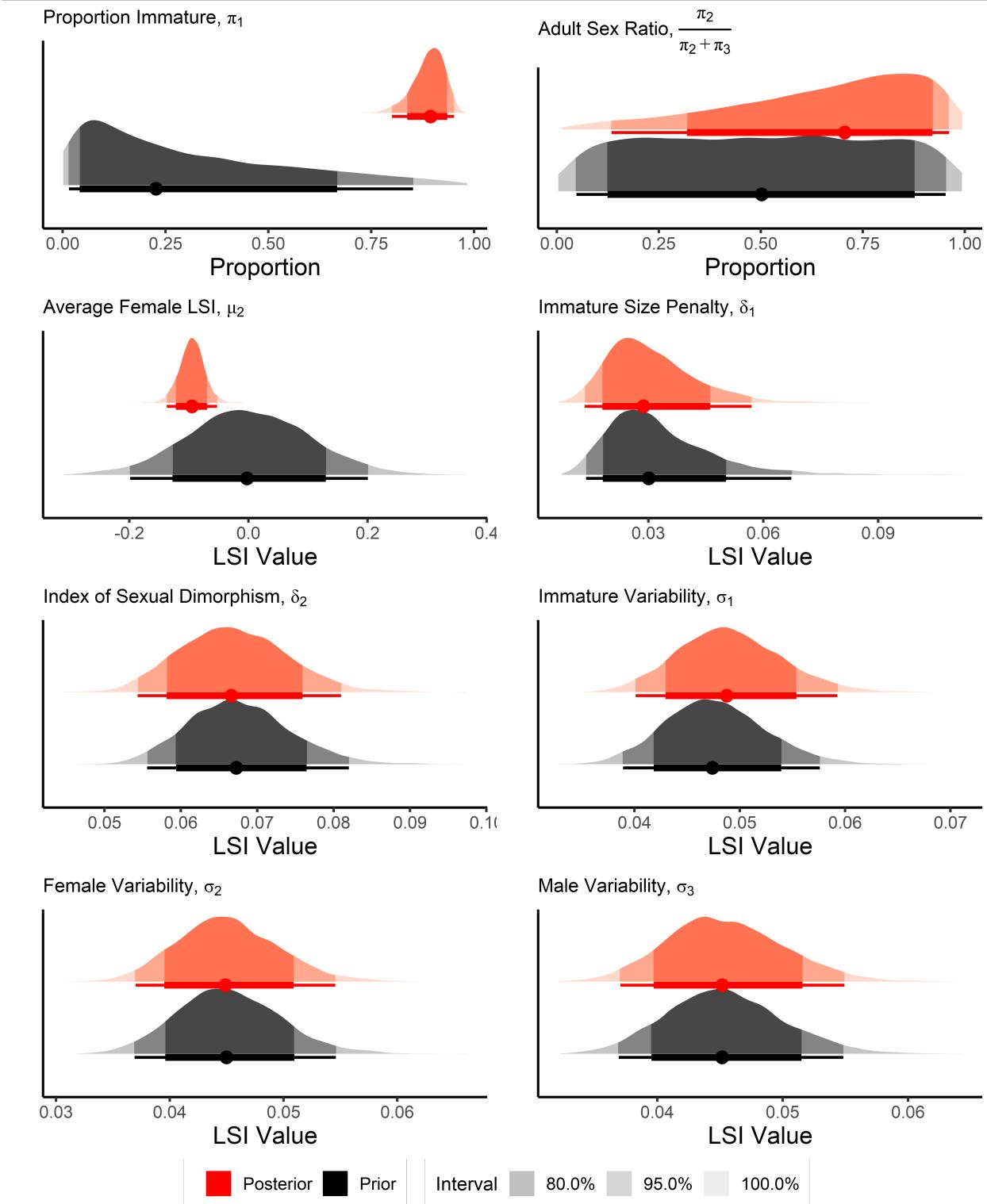


Figure 19: Prior-posterior Comparison of Pınarbaşı B sheep model hyper-parameters

Northwest Anatolian Cattle

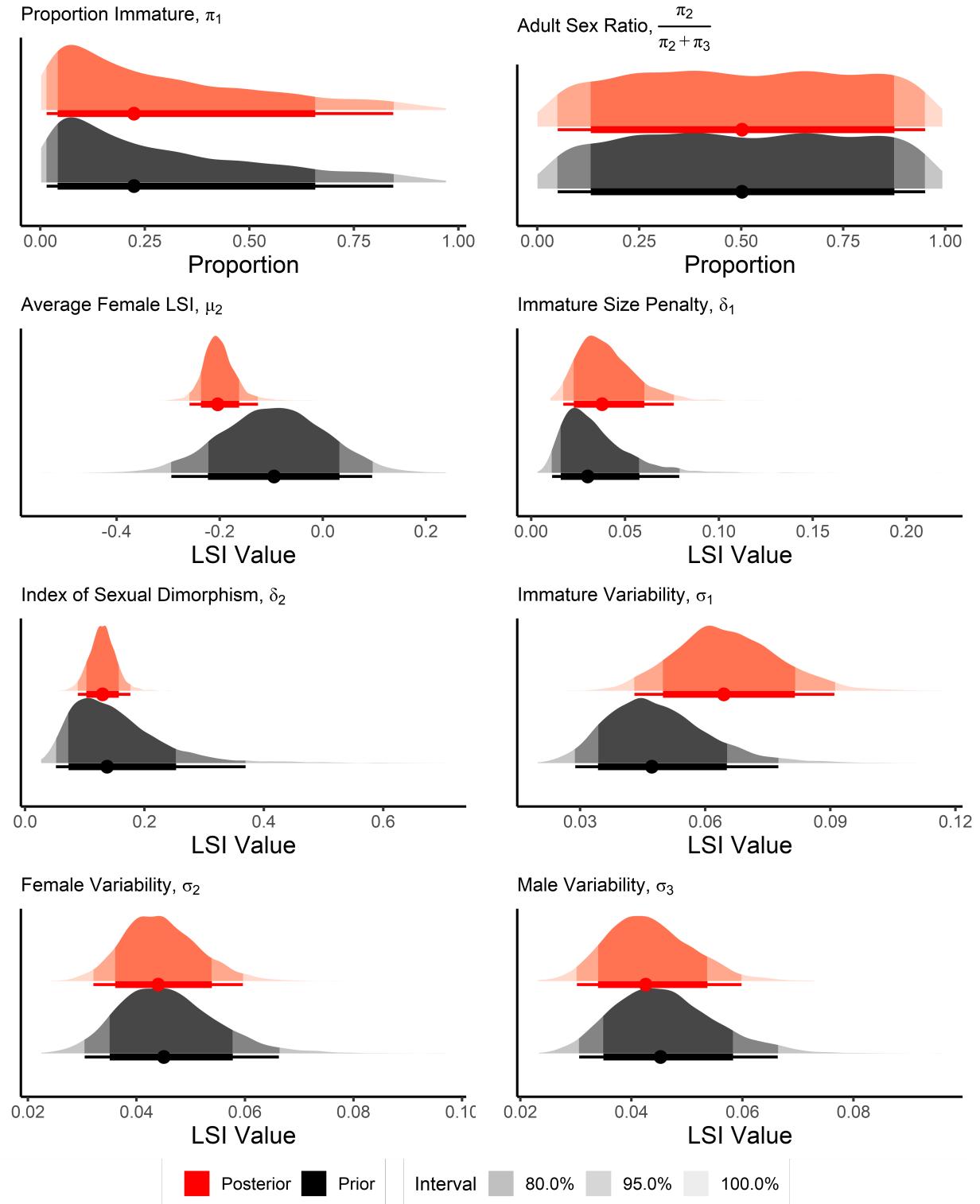


Figure 20: Prior-posterior comparison of Northwestern Anatolian cattle model hyper-parameters