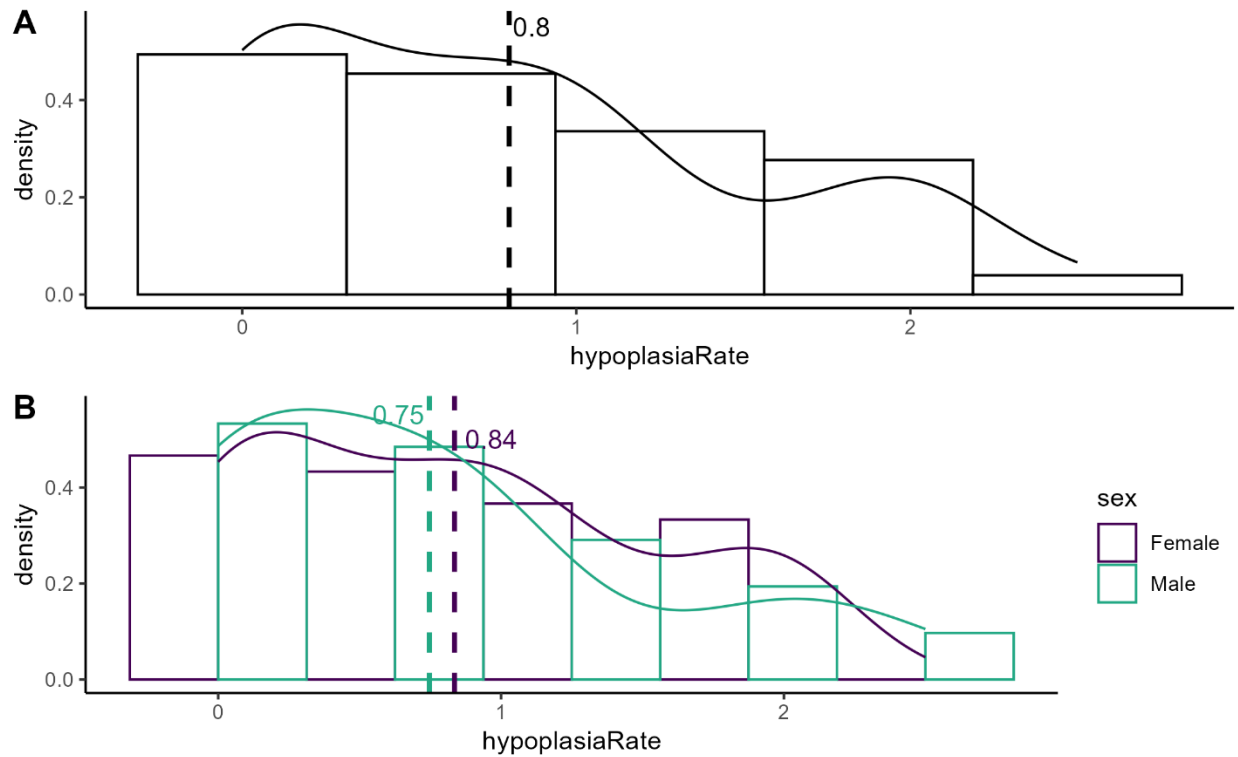


Supplementary Materials for: Market integration improves traditional Hadza male childhood health with no effect on females

Section 1: Data Distribution

Data distributions visualized via histograms with density plots for the entire sample (**S1A**) and by sex (**S1B**). **Figure 1A** shows a clear peak in the first bin, indicating a high frequency of individuals with a hypoplasia rate around 0. The frequency decreases in subsequent bins as the hypoplasia rate increases. The density plot confirms this right-skewed distribution, with a primary mode near 0 and a tail extending towards higher hypoplasia rates. There might be a very slight secondary bump or shoulder around a hypoplasia rate of 2, but the primary concentration is at the lower end. The vertical line for the sample mean is to the right of the primary peak, consistent with a right-skewed distribution where the mean is pulled towards the tail. **Figure 1B** shows distinct distributions by sex. Both have a peak in the first bin (hypoplasia rate around 0), but the shapes differ. The female distribution appears more concentrated around 0 with a longer tail extending to higher rates. The male distribution seems to have a less pronounced peak at 0 and potentially a more noticeable secondary peak or broader distribution around a hypoplasia rate of 1-2. The density plots further illustrate these differences. The female density plot has a higher peak at 0 and a more gradual decline. The male density plot has a lower peak at 0 and a more prominent shoulder or secondary mode in the higher rate range. The female mean is slightly lower than the male mean, and both vertical mean lines are positioned to the right of the primary peak for their respective distributions. Overall, the highest frequency is at or near a hypoplasia rate of 0, indicating good developmental health for a majority of the sample. There is evidence of some individuals with higher hypoplasia rates, creating a longer tail and potentially contributing to the slight secondary mode or shoulder, particularly in the male distribution. The distribution indicates a classic linear model is not appropriate.

SF1: Distribution of data for the sample and by sex



Section 2: LLM Model Centering

Centering the mean separates the effects of predictors into within-group and between-group components and improves the interpretability of the model and significance of the effects studied. Models with different centered means for the year of birth variable were compared (mean year of birth, and two years observed as important break points in the data). Centering on mean year of birth is less meaningful than centering on a year of birth that has inherent significance in the observed patterns, which is why those results are compared to two additional years. The male slope crosses the female slope fully by 1970, and individual male values are at or below the female slope by 1975. **ST1** shows that there are no differences in the results obtained using different mean centers.

Because males born after 1975 have values at or below the female slope, this year was used for centering.

ST1: Model results and anova, comparing centers

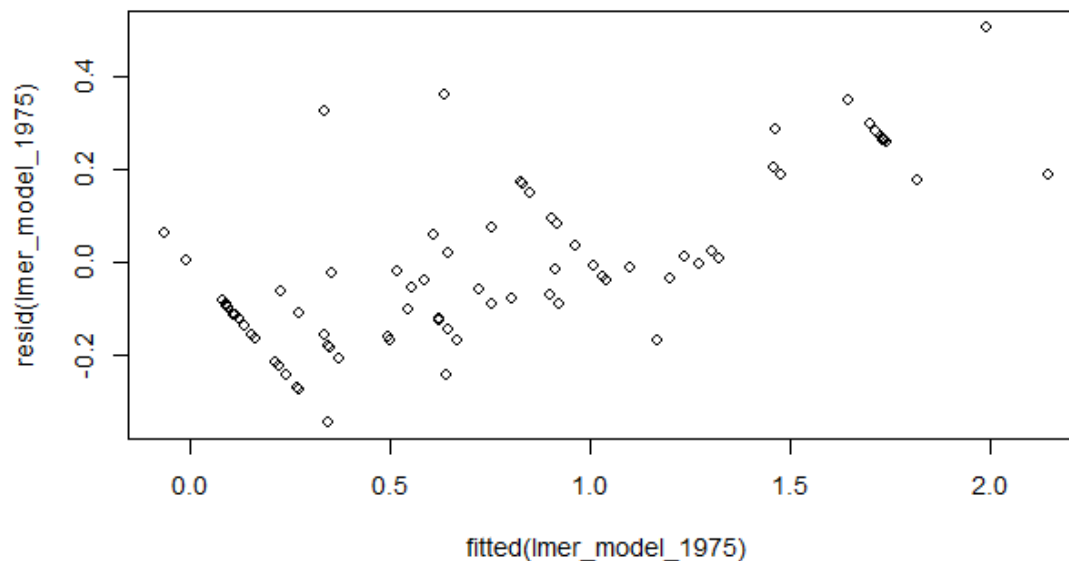
Models	Effects	Est.	Std.Err.	df	t value	p-value
Mean Centered	(Intercept)	1.25	0.14	76.03	8.82	<.01
	yob	-0.01	0.01	76.03	-0.65	0.51
	Male	-0.10	0.22	75.60	-0.43	0.67
	Intersection	-0.04	0.02	75.88	-2.37	0.02
1970 Centered	(Intercept)	1.25	0.14	76.03	8.83	<.01

	yob	-0.01	0.01	76.03	-0.65	0.51
	Male	-0.06	0.22	75.66	-0.28	0.78
	Intersection	-0.04	0.02	75.88	-2.37	0.02
1975 Centered	(Intercept)	1.22	0.15	76.03	8.37	<.01
	yob	-0.01	0.01	76.03	-0.65	0.51
	Male	-0.27	0.24	75.36	-1.14	0.26
	Intersection	-0.04	0.02	75.88	-2.37	0.02

Section 3: LLM Model Diagnostics

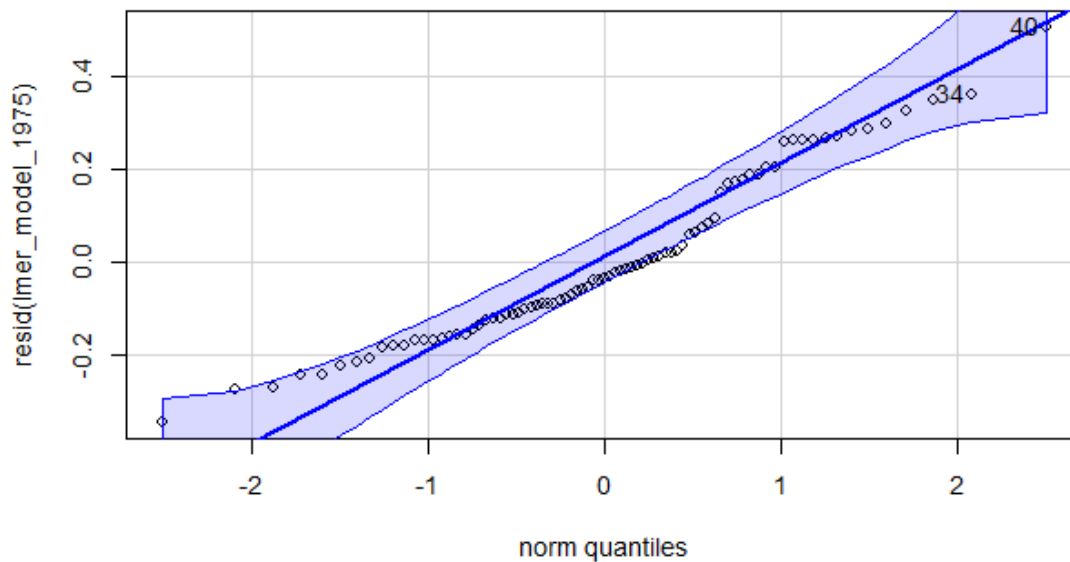
The plot of residuals versus fitted values (**SF2**) indicates the values are scattered around 0 with no obvious curvature. While the variance of residuals is slightly wider at higher fitted values (~1.5-2.0) compared to lower fitted values (~0-0.5), which suggests slight homoscedasticity. The randomness of the residuals is generally supported, suggesting the assumption of linearity is met.

SF2: Plot of residual vs. fitted values



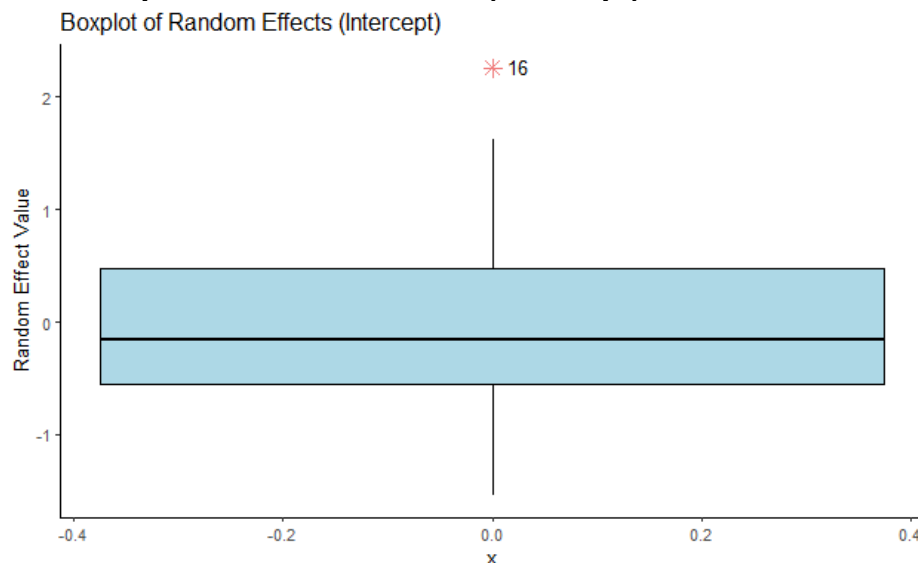
The quantile-quantile plot of the residuals (**SF3**) shows good agreement with the normal distribution line, with most points falling within the confidence interval. Overall, the distribution of residuals did not suggest a severe violation of the normality assumption, although two potential outliers were identified in the QQ-plot (Individuals 34 and 40).

SF3: QQ Plot of residuals



The boxplot summarizing the distribution of random intercepts for individual ID (**SF4**), indicates a median close to zero, and an interquartile range showing the spread of the central 50% of the individual-level deviations. There is a notable outlier at a high positive random effect value (row 16, Individual ID 861), suggesting this individual has a considerably higher baseline hypoplasia rate compared to the rest of the sample.

SF4: Boxplot of random effects (intercept)



Sensitivity testing was conducted on each of the outliers previously identified (**SF3**, **SF4**). These individuals impact the intercept and sex estimates, but not year of birth and interaction effects, which remain stable (**ST2**). Notably, the hypoplasia rates of these

outliers align with the overall temporal trend observed in the data, with the two individuals born before 1975 (Individual 34, born 1929, rate = 1.00; Individual 40, born 1963, rate = 2.50) exhibiting much higher rates of hypoplasia compared to the individual born after 1975 (Individual 16, born 1977, rate = 0.55).

ST2: Sensitivity testing outliers (individuals: #34, #40, #16)

Models	Effects	Est.	Std. Error	t value	CI 5%	CI 95%
LLM	(Intercept)	1.22	0.15	8.37	0.94	1.51
	yob_centered_1975	-0.01	0.01	-0.65	-0.02	0.01
	sexMale	-0.27	0.24	-1.14	-0.74	0.20
	yob_centered_1975:sexMale	-0.04	0.02	-2.37	-0.08	-0.01
Minus Ind 16 (Row 40)	(Intercept)	13.37	17.60	0.76	-21.12	47.87
	yob	-0.01	0.01	-0.69	-0.02	0.01
	sexMale	72.86	33.19	2.20	7.82	137.91
	yob:sexMale	-0.04	0.02	-2.20	-0.07	0.00
Minus Ind 450 (Row 34)	(Intercept)	-5.14	19.75	-0.26	-43.85	33.58
	yob	0.00	0.01	0.32	-0.02	0.02
	sexMale	100.81	34.80	2.90	32.60	169.02
	yob:sexMale	-0.05	0.02	-2.90	-0.09	-0.02
Minus Ind 861 (Row 16)	(Intercept)	13.37	18.64	0.72	-23.17	49.91
	yob	-0.01	0.01	-0.65	-0.02	0.01
	sexMale	81.67	35.15	2.32	12.77	150.57
	yob:sexMale	-0.04	0.02	-2.33	-0.08	-0.01

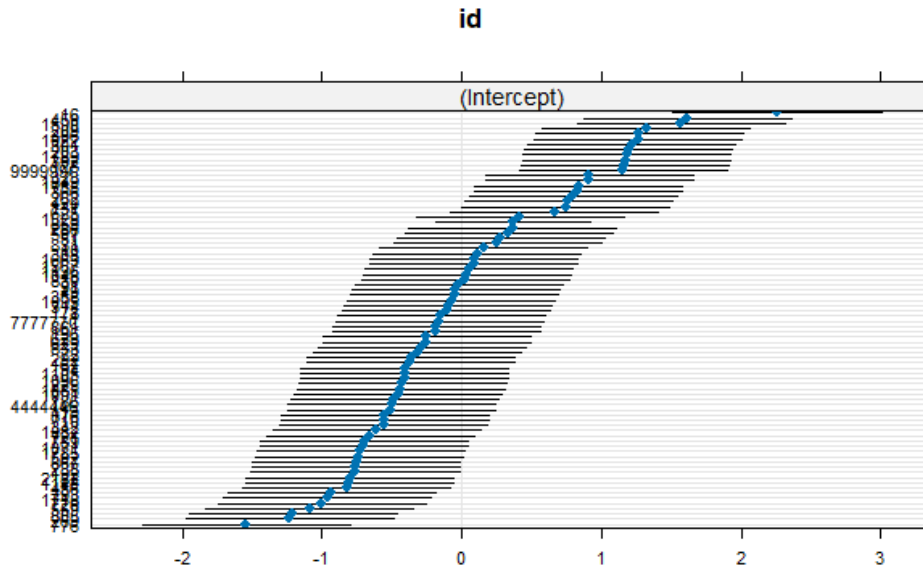
Despite their influence on the main effects (see **ST3** for individual level data), the robustness of the interaction term highlights a consistent pattern of differential change in hypoplasia rates over time between males and females, a phenomenon that warrants further consideration in the context of existing knowledge about sex-based health differences in hunter-gatherer societies.

ST3: Raw data for outliers

Row#	IndID	Age	yob	yob_ctr	Sex	#LEH	#Teeth	hypoRate	tooth_offset
34	450	78	1929	-46	Female	1	1	1.00	-2.48
16	861	30	1977	2	Male	6	11	0.55	-0.09
40	16	45	1963	-12	Male	5	2	2.50	-1.79

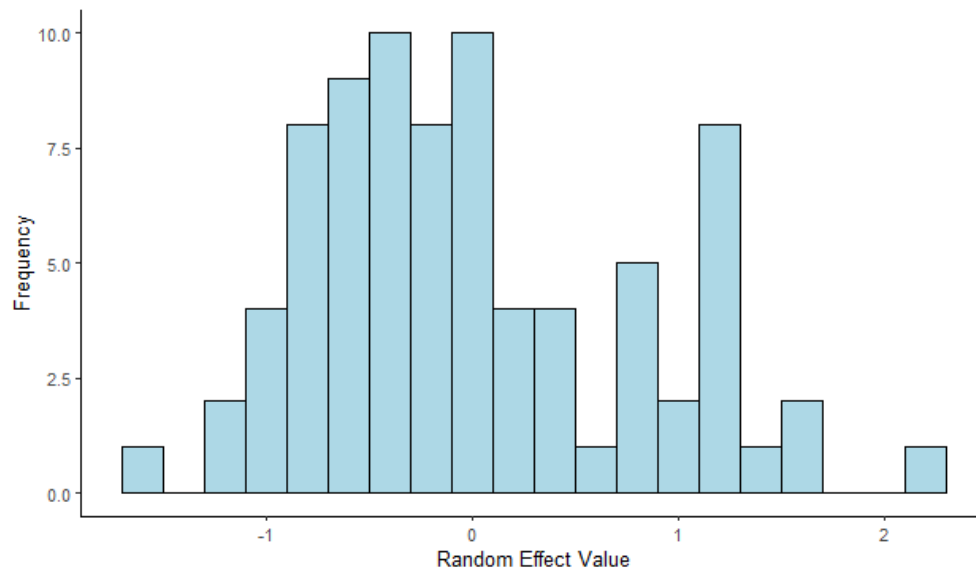
The conditional modes of the random intercepts appear symmetrically distributed around zero. Some clustering or breaks in the pattern can be observed at higher positive values suggesting groups of individuals with similarly elevated baseline hypoplasia rates (**SF5**).

SF5: Conditional Modes of Random Intercepts by Individual ID



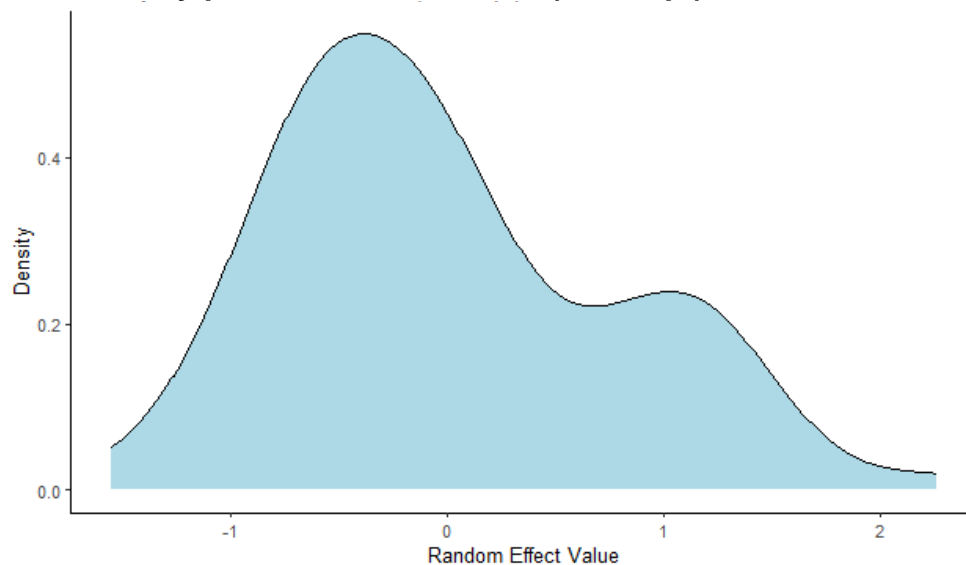
SF6 shows the distribution of the random intercepts for individual ID. While centered around zero, the histogram reveals a somewhat uneven distribution with distinct peaks around -0.25 to 0 and 0 to 0.25, suggesting potential deviations from a perfectly normal distribution of individual-level baseline hypoplasia rates.

SF6: Distribution of random effects (intercept)



SF7 presents the density plot of the random intercepts for individual ID. Consistent with the histogram, the distribution appears bimodal, with a primary peak near zero and a secondary, smaller peak around a random effect value of 1. This bimodality suggests that the individual-level deviations from the average baseline hypoplasia rate are not uniformly distributed and may reflect distinct subgroups within the population.

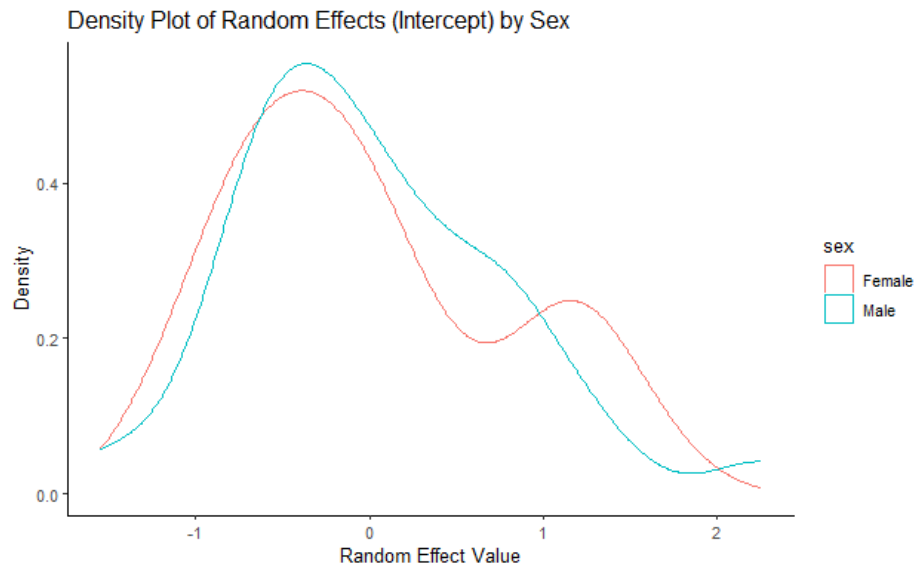
SF7: Density plot of random effects (intercept)



SF8 displays the density plots of the random intercepts for individual ID, separated by sex. The distribution for females is unimodal, centered slightly below zero. In contrast, the distribution for males exhibits a clear bimodality, with a primary peak like that of females and a distinct secondary peak around a random effect value of 1. This suggests

that the presence of a subgroup of males (born before 1975, see main paper **Figure 1B**) with higher baseline hypoplasia rates is the primary driver of the bimodality observed in the overall distribution of random intercepts.

SF8: Density plot of random effects (intercept) by sex



Section 4: Bayesian Model (brms)

LMM comparison (with and without random effect)

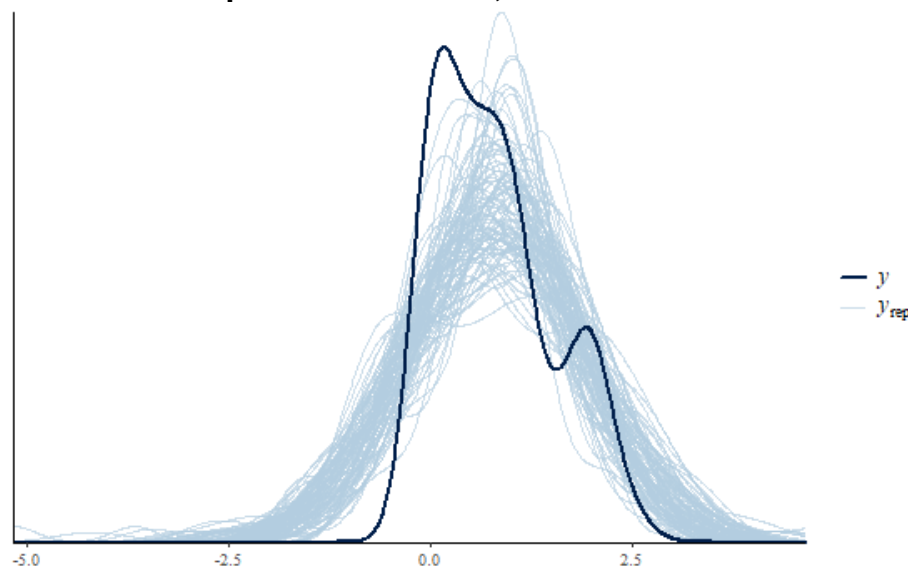
The LMM included a random intercept for individual ID, based on the understanding that individual-level variation might influence hypoplasia rates, but initial attempts to fit a similar Bayesian model with a random intercept encountered convergence issues, suggesting near-zero variance for this random effect. Given these challenges, we explored alternative Bayesian model specifications. We attempted to improve sampling by adjusting `max_treedepth` and `adapt_delta`, as well as increasing iterations, but convergence problems persisted. We also experimented with different weakly informative half-normal priors with varying widths (narrow, wide, and wider) for the standard deviation of the random intercept, but the near-zero variance and convergence issues remained and the posterior estimates consistently indicated near-zero variance for this random effect.

We investigated the necessity of the random intercept for individualID by comparing models with the same fixed-effect structure, the previously described and one without a random effect using `lm` in base R. We compared results using a likelihood ratio test in base R `anova`. The results of the ANOVA comparison (`anova(lmer_model, lmer_model_fixed)`) showed no significant difference in model fit ($\text{Chisq} = 0.46$, $\text{df} = 1$, $p = 0.5$). This statistical evidence supports the finding from our Bayesian attempts that the random intercept had minimal impact on explaining the variance in hypoplasia rates beyond the fixed effects of year of birth and sex.

ST4: Model comparison, with and without random effect

Models	Effects	Est.	Std. Error	t value
Random	(Intercept)	1.22	0.15	8.37
	yob_centered_1975	-0.01	0.01	-0.65
	sexMale	-0.27	0.24	-1.14
	yob_centered_1975:sexMale	-0.04	0.02	-2.37
No random	(Intercept)	1.22	0.15	8.40
	yob_centered_1975	-0.01	0.01	-0.66
	sexMale	-0.25	0.23	-1.08
	yob_centered_1975:sexMale	-0.04	0.02	-2.31

The posterior predictive checks (**SF9**) revealed some discrepancies between the observed and predicted data distributions. The density overlay indicated that while the model captured the general range of hypoplasia rates, the predicted densities tended to smooth over the observed bimodal distribution (a primary peak at zero and a secondary peak around two, driven by females). The predicted densities showed a tendency to fill in the lower values to the left of the primary peak, where the observed data had fewer observations.

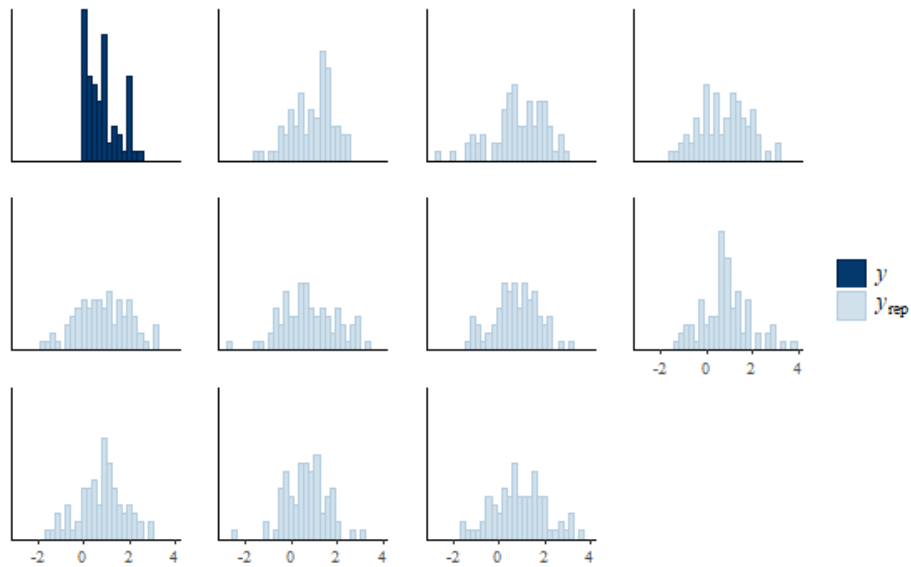
SF9: Posterior predictive checks, brms

Similarly, the histograms (**SF10**) showed that while the observed data had a high frequency in the first bin (zero hypoplasias) with decreasing frequency in subsequent

bins, the predicted data exhibited a more even distribution with some predictions extending to lower values than observed.

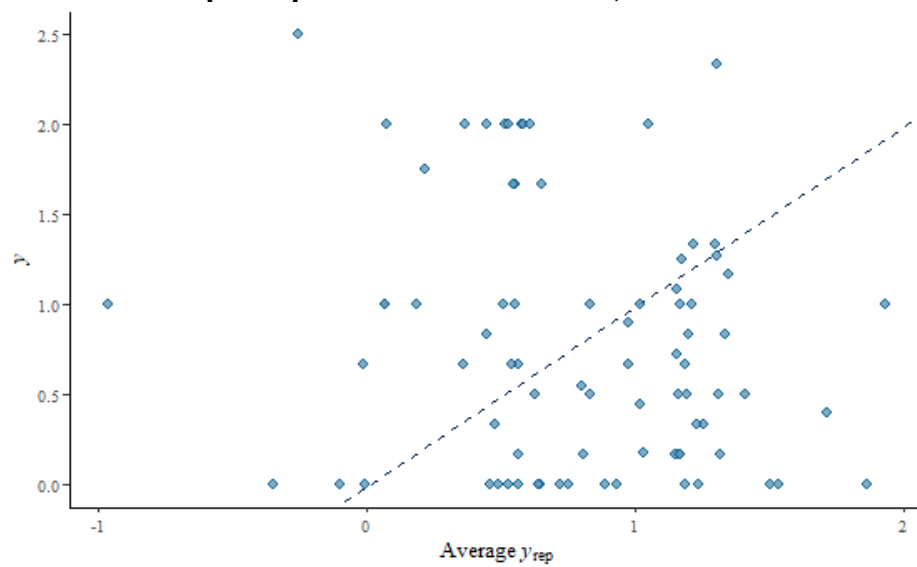
SF10: Histograms observed v predicted, brms

Predicted densities, brms



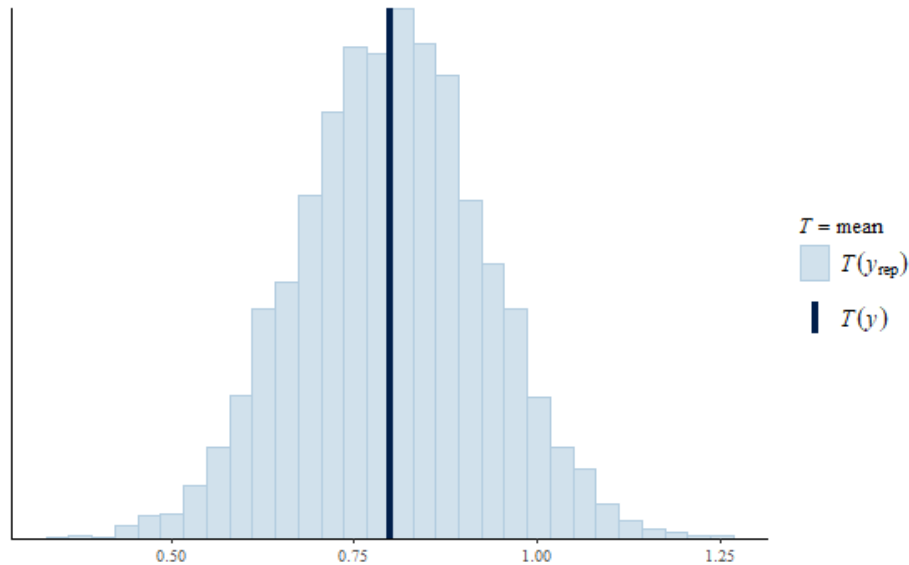
The average predicted vs. observed scatterplot showed a reasonable fit, with points scattered around the identity line, and denser areas at zero and two, reflecting the modes in the observed data (SF11).

SF11: Scatterplots predicted v observed, brms



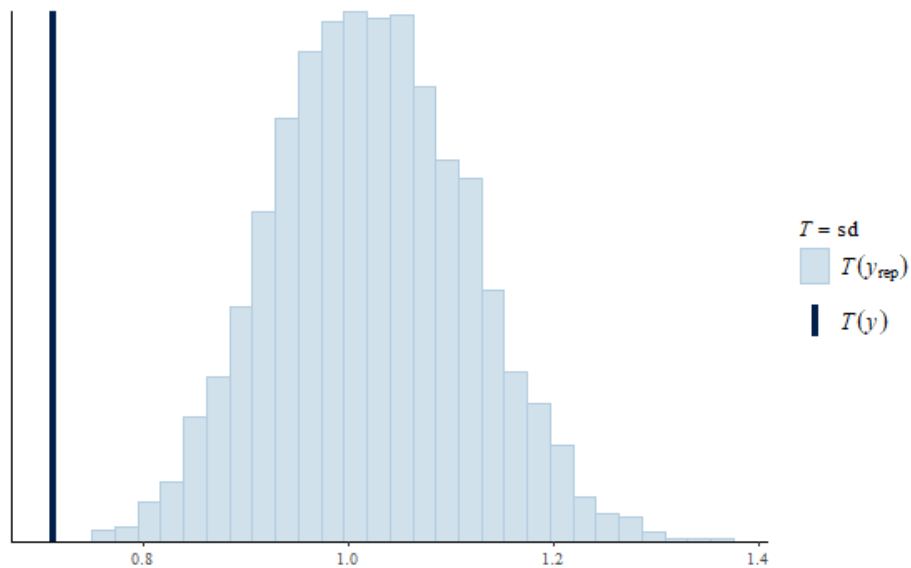
The posterior predictive checks for the mean (**SF12**) showed the mean of the observed data (0.71) lies within the distribution of the means of the predicted data, suggesting the model adequately captured the central tendency.

SF12: Posterior predictive checks for mean, brms



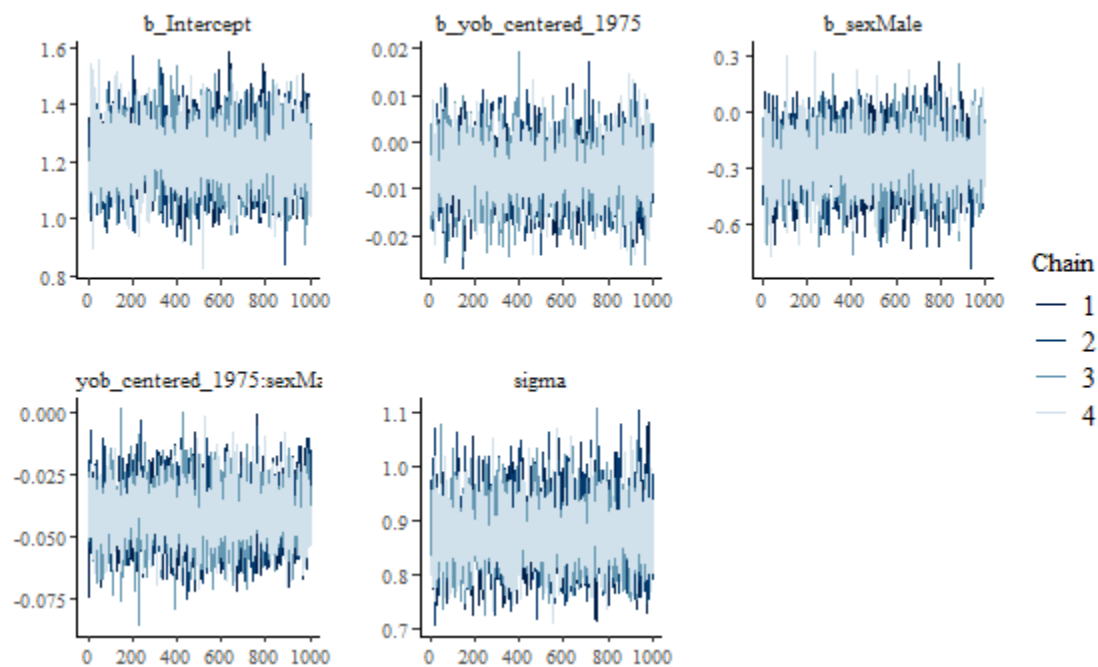
In contrast, the posterior predictive check for the standard deviation (**SF13**) indicated that the standard deviation of the observed data (0.71) was at the lower end of the distribution of the standard deviations of the predicted data, suggesting the model might be slightly overestimating the variability in the outcome.

SF13: Posterior predictive checks for std.dev., brms



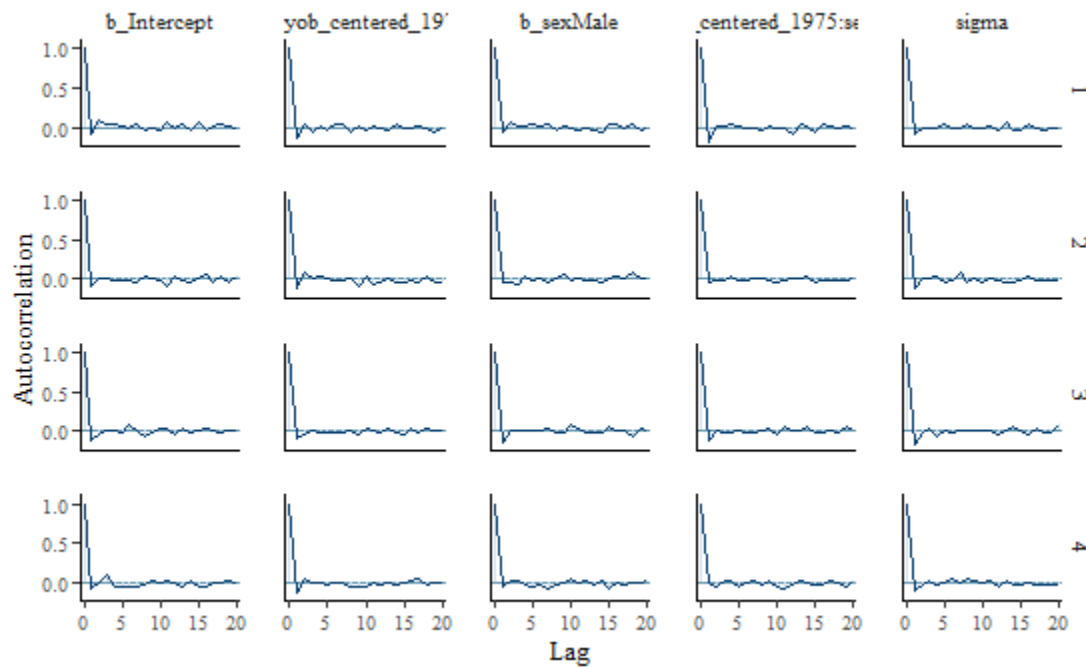
The convergence and mixing of the MCMC chains were assessed through visual inspection of trace plots and by examining Rhat values. Trace plots for all parameters (**SF14**) showed intermingling and stationarity across the four chains, further supported by Rhat values of 1.00. Effective sample sizes (Bulk_ESS and Tail_ESS) were also adequate (all above 1500, **Table 4**), indicating efficient sampling.

SF14: Trace plots, brms



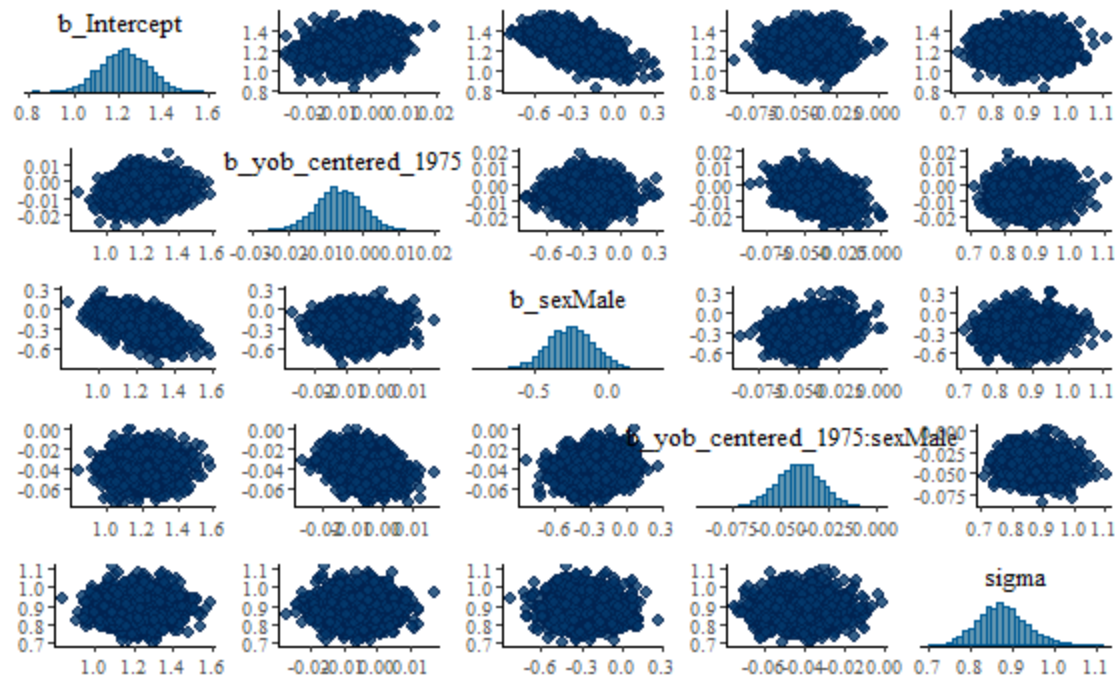
Further evaluation of the MCMC chains involved examining autocorrelation function plots for all parameters (**SF15**). These plots demonstrated a sharp decline to near zero within 1-2 lags, suggesting low autocorrelation and that the MCMC samples are largely independent, each providing new information about the posterior distribution. This efficient sampling behavior strengthens confidence in the reliability of the brms model's parameter estimates.

SF15: Autocorrelation function plots, brms



The pairs plot of the posterior distributions (**SF16**) showed generally reasonable bivariate distributions for the model parameters, with some expected correlations (e.g., between the intercept and yob, and between sexMale and yob:sexMale) but no indications of problematic strong or unexpected relationships.

SF16: Pairs plots of posterior distributions, brms



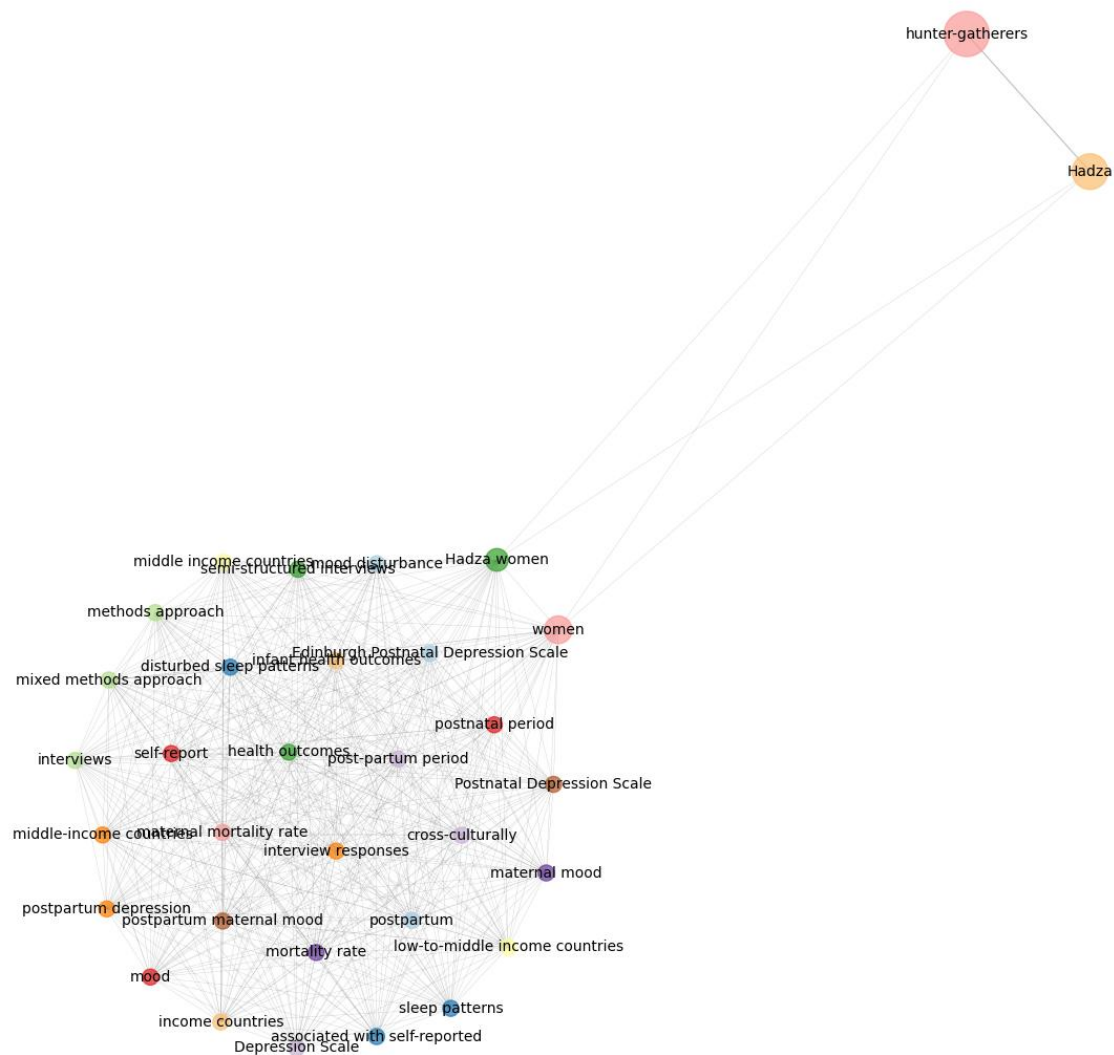
In summary, while the Bayesian model had some limitations in capturing the full bimodality of the data, the satisfactory convergence and mixing of the MCMC chains provide confidence in the reliability of the estimated fixed effects, which were consistent with the LMM results.

Section 5: Bibliometric Analysis

Bibliometric research focused on a timeline of publications and word co-occurrence/concept networks using data from dimensions.ai with the search parameters focused on 'Hadza AND men NOT women' or 'Hadza AND women NOT men'. Network metrics were presented in the main paper. This section includes the network visualizations and the terms that define the network clusters.

The women's network (**SP17**) shows a more tightly clustered and interrelated set of topics, which likely reflects the shorter time depth and interests of the collaborative networks producing the research.

SF17: Women's Co-occurrence Network



The cluster terms (**ST5**) provide insights into main topics including demography, mental health, physical health, legal matters, childcare, and social structure.

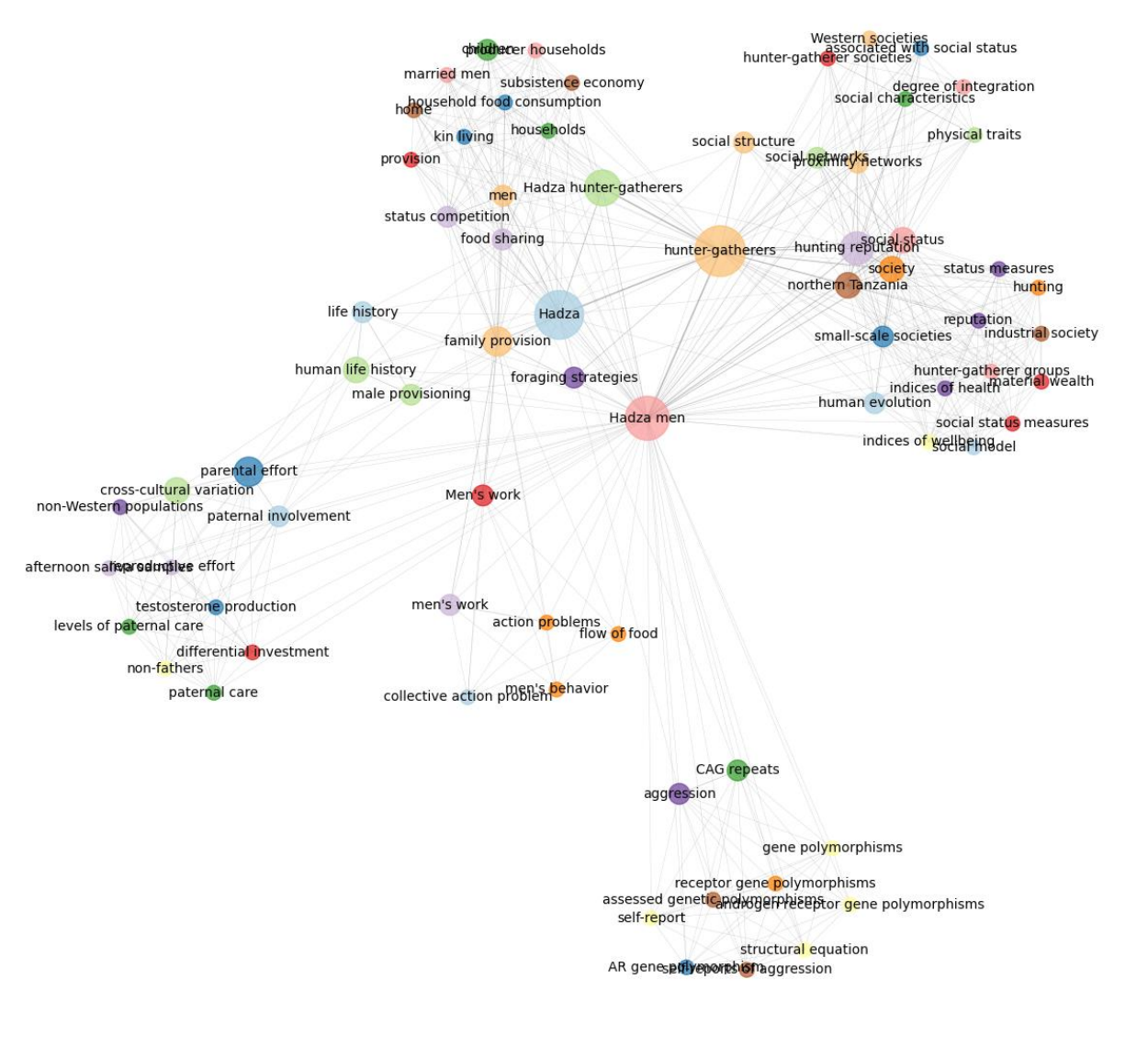
ST5: Cluster Terms for Hadza Women Network

Cluster 4	African hunter-gatherers, Hadza foragers, age of women, age structure, demographic differences, demography, estimates of age structure, higher fertility, living children, population growth rate, total population size
-----------	--

Cluster 1	Depression Scale, Edinburgh Postnatal Depression Scale, Hadza women, Postnatal Depression Scale, associated with self-reported, cross-culturally, disturbed sleep patterns, health outcomes, income countries, infant health outcomes, interview responses, interviews, low-to-middle income countries, maternal mood, maternal mortality rate, methods approach, middle income countries, middle-income countries, mixed methods approach, mood, mood disturbance, mortality rate, post-partum period, postnatal period, postpartum, postpartum depression, postpartum maternal mood, self-report, semi-structured interviews, sleep patterns, women
Cluster 7	East Africa, allomaternal nursing, biological kin, ethnographic reports, follower's observations
Cluster 2	Hadza, Hadza hunter-gatherers, Hadza hunter-gatherers of Tanzania, Sternberg's triangular theory, coital frequency, concept belief, fallback foods, human evolution, human evolutionary past, hunter-gatherer populations, hunter-gatherer societies, hunter-gatherers, low caloric value, people, reproductive success, rituals, women of reproductive age
Cluster 6	assault, court documents, legal problems, professional jealousy, sexual assault, sexual misconduct
Cluster 5	associated with hair cortisol concentrations, associated with higher hair cortisol concentrations, camp members, effects of social relationships, in-person, social network measures, social relationships
Cluster 3	egalitarian social structure, egalitarian society, foraging bands, hair cortisol concentrations, high social status, human evolutionary history, human society, individual social status, non-human animals, social groups, social hierarchy, social status, social structure, women's social status

The men's network (**SP17**) shows a sparse set of clusters with strong internal connections and a broader set of topics, which likely reflects the longer time depth of study, multiple generations of researchers, and broader collaborative networks producing the research.

SF18: Men’s Co-occurrence Network



The cluster terms (**ST6**) provide insights into main topics including genetics, men’s activities, subsistence, household, transition, skills, tools, status, and wealth.

ST 6: Cluster Terms for Hadza Men Network

Cluster 2	AR gene polymorphism, CAG repeats, Hadza, Hadza of Tanzania, aggression, androgen receptor gene polymorphisms, assessed genetic polymorphisms, chimpanzees, gene polymorphisms, human population, receptor gene polymorphisms, self-report, self-reports of aggression, structural equation
-----------	---

Cluster 1	Hadza data, Men's work, action problems, ancestral mothers, collective action problem, dependent offspring, early Homo, family provision, flow of food, hunter-gatherer ethnography, hunting hypothesis, men, men's behavior, men's work, primate males, status competition
Cluster 8	Hadza foragers, division of labor, foraging strategies, human life history, hunter-gatherers, life history, male provisioning, non-human primates, sexual division, sexual division of labor
Cluster 5	Hadza hunter-gatherers, children, food sharing, home, household food consumption, households, kin living, married men, producer households, provision, subsistence economy
Cluster 3	Hadza men, Western societies, associated with social status, degree of integration, hunter-gatherer societies, hunting reputation, northern Tanzania, physical traits, proximity networks, social characteristics, social networks, social status, social structure, society
Cluster 7	afternoon saliva samples, differential investment, levels of paternal care, male care, non-Western populations, non-fathers, parental effort, paternal care, paternal involvement, reproductive effort, testosterone production
Cluster 6	age of acquisition, cross-cultural differences, cross-cultural variation, oblique transmission, perceptions of task difficulty, ratings of teaching, social practices, subsistence skills, task difficulty, teaching labor, use of teaching
Cluster 4	human evolution, hunter-gatherer groups, hunting, indices of health, indices of wellbeing, industrial society, material wealth, reputation, small-scale societies, social model, social status measures, status measures