OptiMod

SCIENTIFIC REPORT

How are fetal ultrasound measurements associated with birth weight?

Real observations. A fully automatic analysis.
Every check on record.

ULTRASOUND DATA · GAMLSS.DATA::ULTRA

Royal Hospital for Women, Sydney, Australia
Second ultrasound scans and birth weights of 1,038 births, 2008–2013

1,038
births analysed, none removed or imputed

11
candidate models fitted, checked and compared

78.8%
of the variation in birth weight accounted for in the fitting sample

Question. How are abdominal circumference, biparietal diameter and femur length at the second ultrasound scan associated with birth weight, once head circumference, parity, maternal age and the days from scan to delivery are taken into account?

Key finding. A larger abdominal circumference goes with a clearly higher average birth weight: 583 g between the lower and the upper quartile (95% CI 536 to 629 g). Femur length and biparietal diameter contribute 188 g and 111 g. All three associations keep their direction, with intervals clear of zero, in all four fitted models and in the analysis without two unusual births. The engine also found, without being asked, that the association with the time to delivery depends on femur length.

Model. The location-scale model LS1 fits these data best: first of eleven candidates by cross-validated log loss, with the lowest AIC and the highest R^2 of the four models fitted to all births. Its final fit shows no major contradiction of the model assumptions. OptiMod's formal recommendation follows a stricter standard and is not issued until further checks are complete; Section 06 names them.

Analysed automatically by OptiMod. Data screening, the model search, assumption checks, findings and limitations in this report were computed by the OptiMod engine (backend 0.4.0, unmodified default request).

Sample report in the layout under development, prepared from the engine's saved results of 16 November 2026. Exploratory associations; not a validated clinical prediction.

01 / DATA AND INTEGRITY

Start with the observations

One row is one birth. All 1,038 rows enter the analysis; nothing is imputed, winsorized or silently removed.

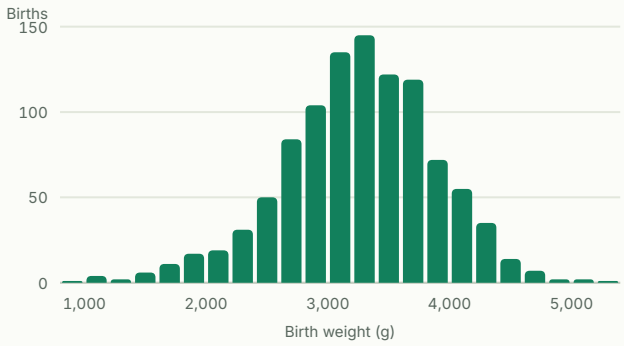
Scope	Observed data
Source	R package `gamlss.data` 6.0.7, dataset `ultra`; analysed in Stasinopoulos, Kneib, Klein, Mayr and Heller (2024), Chapter 8.
Setting	Second ultrasound scans, Royal Hospital for Women, Sydney, 2008–2013.
Outcome	`birthweight` in grams: mean 3,249, median 3,275, SD 634, range 940 to 5,280.
Records	1,038 births · 8 variables · no missing values.

Variable	Role	Meaning	Mean	Median	SD	Range
AC	focal	Abdominal circumference	31.18	31.50	3.65	18.8 to 41.8
BPD	focal	Biparietal diameter	8.62	8.80	0.76	5.3 to 10.8
FL	focal	Femur length	6.64	6.70	0.67	3.9 to 8.2
HC	control	Head circumference	30.99	31.45	2.48	20.7 to 39.6
age	control	Age of the mother, years	34.92	35.00	5.63	17 to 52
DBD	control	Days from scan to delivery	28.11	24.00	21.58	0 to 112
parity	control	Number of previous births				0: 545 (52.5%) · 1: 335 (32.3%) 2: 108 (10.4%) · 3+: 50 (4.8%)

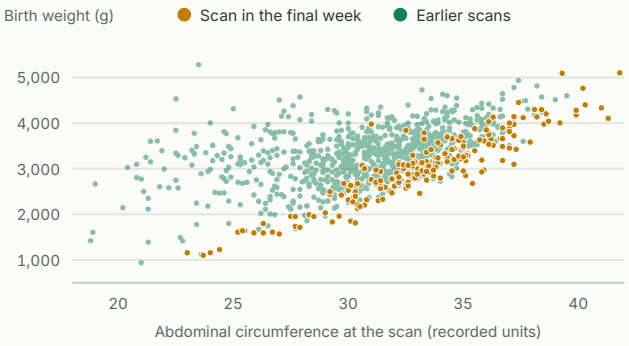
Ultrasound measurements are shown in their recorded units; the package documentation does not state them. Role: focal variables are the subject of the question, controls are adjusted for.

Checks against the file

Check	Finding	Treatment
Missing values	None in the eight variables.	No deletion, no imputation.
Exact duplicates	5 of 1,038 rows repeat an earlier row in every analysed field.	Kept as given; recorded as a limitation.
Unusual observations	Rows 464 and 815 flagged in the development screen: large conditional residual and high influence.	Kept; a separate sensitivity analysis excludes them.
Meaning of DBD	The package help labels it “date of birth”; the textbook defines it as days from scan to delivery.	Textbook definition used; the discrepancy stays disclosed.
Measurement units	Not documented for the four ultrasound measurements.	Recorded values used unchanged.



Birth weight. 1,038 births in 200 g classes.



Timing matters. At the same abdominal circumference, births soon after the scan are lighter than births weeks later. Unadjusted view; 194 scans fall in the final week.

02 / QUESTION AND DESIGN

Fix the design before the fit

The analyst supplies the data, the outcome and the variables to consider. The engine determines everything else and keeps a record of each decision.

Specified by the analyst

Setting	Value
Goal	Explain a relationship; exploratory analysis.
Outcome	`birthweight`, continuous, in grams.
Focal variables	`AC`, `BPD`, `FL`
Controls	`HC`, `parity`, `age`, `DBD`
Request	Unmodified default: 120 s, 24 candidates, at most 2 batches.

All seven selected variables stay in every model. The engine does not drop variables on the basis of significance.

Determined by the engine

Decision	Record
Evaluation design	831 births for development in 5 folds, 207 held out; seed 20260911. The holdout is not used during the search.
Predictors	Numeric mean predictors standardized, z = (value − mean) / SD; parity as categories with reference level 0.
Outcome scale	Original and logarithmic outcome examined; original-scale means of log models include the variance correction.
Populations	Original population, n = 1,038. The two flagged rows define a separate sensitivity population, n = 1,036.
Candidates	11 fitted for the original population, 13 for the sensitivity population.
Inference	Model-conditional 95% confidence intervals; no adjustment for model selection or multiple comparisons.
Recommendation	Policy `research_v3`: formally recommended only when every required check is supported.

What ran automatically

01 Screen — Types, missing values, duplicates and unusual rows.

02 Split — 831 births to develop the model, 207 held out.

03 Search — Eleven candidates, each a modelling question.

04 Check — Assumption checks in every fold and on the final fit.

05 Select — A fixed rule; no recommendation without complete checks.

06 Audit — The locked model meets the holdout once.

07 Report — Refit on all births; findings, limits and provenance.

831 development births · five folds for cross-validation | 207 holdout

Boundaries set by the design

Associations, not causes

The data are observational. The estimates describe how birth weight differs between births with different measurements; they do not describe the effect of changing a measurement.

Retrospective timing

`DBD` is the actual number of days from scan to delivery. It is unknown at the scan, so the model describes a retrospective association and is not a prospective clinical predictor.

Exploratory inference

Intervals and p-values hold within the fitted model. They are not adjusted for the model search or for the number of comparisons and do not confirm a hypothesis.

Reference profile for adjusted estimates

Adjusted means and contrasts hold the other variables at their medians and at the reference category: AC 31.5, BPD 8.8, FL 6.7, HC 31.45, age 35, DBD 24, parity 0 (as for 545 of the 1,038 births). Contrasts for numeric variables compare the upper with the lower quartile; each parity category is compared with parity 0.

03 / MODEL SEARCH

Eleven candidates, and the one that fits best

Each candidate answers a modelling question. All are scored on the same 831 development births by five-fold cross-validation.

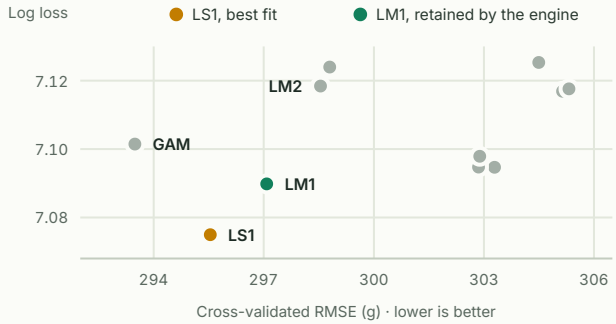
#	Candidate	Role in the search	Checks	CV R^2	RMSE	MAE	Log loss	EDF
1	LS1 · Location-scale · log · DBD × FL · σ: AC	Targeted refinement	incomplete	0.7879	295.54	224.63	7.0750	13
2	LM1 · Linear · log · DBD × FL	Interaction scout	incomplete	0.7857	297.08	225.61	7.0898	11
3	LS2 · Location-scale · log · σ: AC	Matched comparison	incomplete	0.7766	303.29	231.75	7.0947	12
4	Location-scale · log · σ: DBD, FL	Competing explanation	incomplete	0.7773	302.85	231.38	7.0947	13
5	Location-scale · log · σ: AC, DBD, FL	Competing explanation	incomplete	0.7772	302.88	231.51	7.0979	14
6	Additive model with smooth terms (GAM)	Smooth mean, constant scale	incomplete	0.7908	293.49	223.41	7.1015	11.95
7	Linear · log	Lognormal linear location	not adequate	0.7739	305.14	233.61	7.1169	10
8	Linear · log · quadratic AC	Targeted refinement	not adequate	0.7736	305.32	233.75	7.1176	11
9	LM2 · Linear · original	Linear mean, constant scale	not adequate	0.7835	298.55	228.60	7.1184	10
10	Location-scale · original · σ: AC, BPD, FL	Linear mean, linear scale	incomplete	0.7832	298.80	228.55	7.1240	14
11	Gamma regression	Gamma distribution question	incomplete	0.7748	304.50	232.93	7.1253	10.66

Ranked by cross-validated log loss; RMSE, MAE in grams. log, original: outcome scale (GAM, Gamma: original). σ: dispersion variables. EDF: effective parameters. "Incomplete" (engine label: inconclusive): all completed checks are supported; pattern checks, in four cases one more, could not run.

The model this report presents

LS1 fits best. It ranks first of the eleven candidates by cross-validated log loss and has the lowest AIC and the highest R^2 of the four fitted models. It also models the spread of birth weights, the point on which the simpler LM1 was flagged in its holdout audit (Section 07).

The engine's default rule is more cautious. It keeps the simpler model unless the gain is clear-cut: LS1 improves the log loss over LM1 by 0.015 per observation with 2 additional parameters, just short of the margin of 0.016. The engine therefore retains LM1; this report presents LS1 for its fit. Their findings agree closely (Section 07).



Two criteria, two rankings. Each point is one candidate; lower left is better. The additive model (GAM) predicts the average slightly better, LS1 describes the spread of individual births better.

The four fitted models

Model	AIC	Δ AIC	Fit R^2	CV RMSE g	CV MAE g	CV log loss	Role
LM1	14,716.97	38.28	0.7857	297.08	225.61	7.0898	Retained by the engine's default rule
LS1	14,678.69	0.00	0.7881	295.54	224.63	7.0750	Best fit; presented in this report
LM2	14,767.40	88.71	0.7846	298.55	228.60	7.1184	Comparison
LS2	14,721.21	42.52	0.7768	303.29	231.75	7.0947	Comparison

Original population, n = 1,038. AIC on the scale of the original outcome, including the Jacobian of the log transform. None of the four carries OptiMod's formal recommendation yet (Section 06).

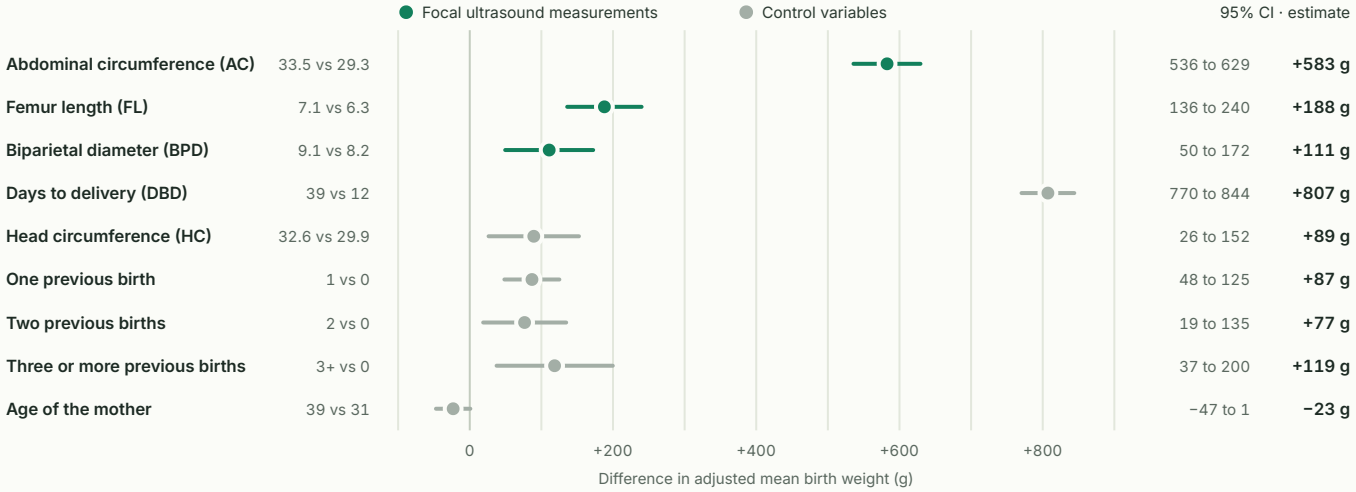
Scope of the search

11 candidates completed, 0 failed. When the candidate budget was reached, 280 further model questions were recorded as deferred, among them smooth and quadratic terms combined with the interaction, other dispersion structures and the interactions AC × HC and AC × BPD. The search is bounded, not exhaustive; deferred questions are not ruled out.

04 / FINDINGS

What goes with a higher birth weight

Together, the modelled variables account for about 78.8% of the observed variation in birth weight in the fitting sample. Each estimate holds the other variables at the reference profile.



Adjusted contrasts, model LS1. Numeric variables: upper against lower quartile. Parity: against no previous birth. Age: 39 against 31 years. Intervals are model-conditional and not adjusted for multiple comparisons. Every contrast keeps its direction in all four fitted models (Section 07).

The findings in plain language

Variable	Component	What the model finds	Coefficient	95% CI	p
AC	average	Higher AC is associated with higher average birth weight.	+0.166	0.152 to 0.179	< 0.001
AC	dispersion	Higher AC is associated with lower dispersion of log birth weight.	−0.0438	−0.0575 to −0.0300	< 0.001
BPD	average	Higher BPD is associated with higher average birth weight.	+0.0306	0.0136 to 0.0476	< 0.001
FL	average	Higher FL is associated with higher average birth weight.	+0.0478	0.0336 to 0.0620	< 0.001
DBD × FL	combined effect	The relationship between DBD and birth weight depends on FL. Found automatically.	−0.0170	−0.0219 to −0.0121	< 0.001
DBD	average	Higher DBD is associated with higher average birth weight.	+0.207	0.198 to 0.216	< 0.001
HC	average	Higher HC is associated with higher average birth weight.	+0.0266	0.0077 to 0.0454	0.006
parity	average	Average birth weight differs between parity categories (overall test p < 0.001).	+0.0277 · +0.0245 · +0.0377		< 0.01
age	average	No statistically significant association with average birth weight.	−0.0053	−0.0108 to 0.0002	0.058

Coefficients of the mean: log birth weight per standard deviation of the predictor. Dispersion coefficient: log(σ − 0.01) per recorded unit of AC. Parity: one, two and three or more previous births against none.

FOUND AUTOMATICALLY · EXPLORATORY

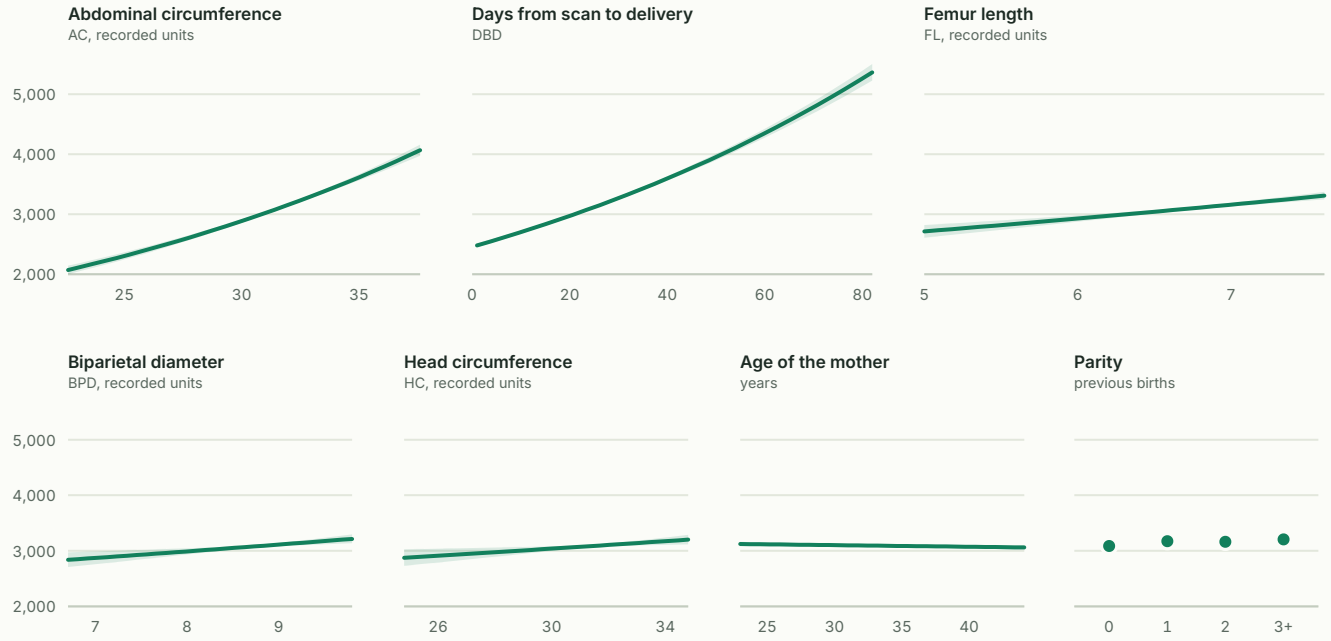
A combined effect of time to delivery and femur length

The engine screened all 21 pairs of variables for interactions and examined DBD × FL. Adding it lowered the cross-validated prediction error (RMSE) by 2.6% and the log loss by 0.020 against the otherwise identical model, in all five folds. The coefficient is negative: the longer the time between scan and delivery, the less femur length tells about birth weight. The folds that suggested the interaction also assessed it, so it remains exploratory until it is confirmed on independent data. Section 05 shows its shape and the second model component, the spread of individual births.

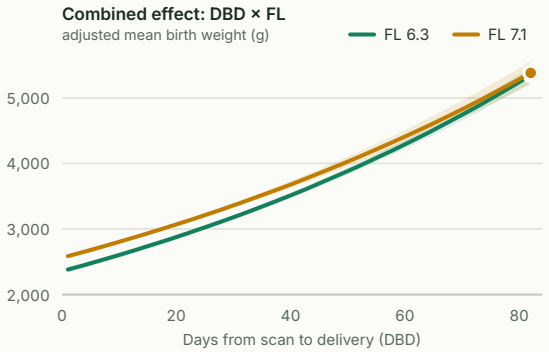
05 / RELATIONSHIPS

The shape of each relationship

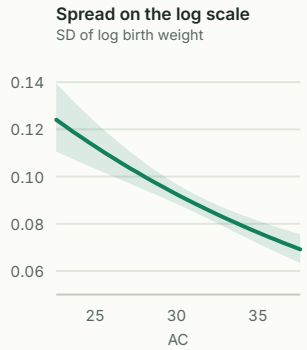
Adjusted mean birth weight in grams along each variable, with pointwise 95% confidence bands. All panels share one vertical scale, so slopes can be compared directly.



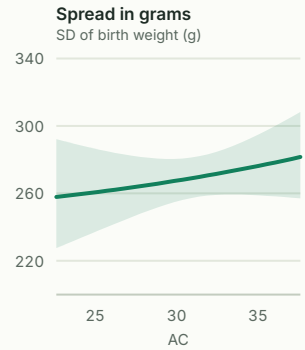
Other variables at the reference profile: AC 31.5, BPD 8.8, FL 6.7, HC 31.45, age 35, DBD 24, parity 0. Curves cover the 2.5th to 97.5th percentile of each variable; all rows remain in the fit. Bands describe the uncertainty of the mean, not the range of individual births.



The interaction. With a longer femur the curve along the time to delivery is flatter in relative terms; the two curves approach each other. Evaluated from the fitted model equation of Section 08 at the lower and upper quartile of FL.



Relative spread falls as AC increases.



Absolute spread stays about the same.

SECOND MODEL COMPONENT

How much individual births vary

LS1 models the spread as well as the average; its dispersion equation contains AC only. On the log scale the standard deviation falls from about 0.124 at AC 22.6 to 0.069 at AC 37.6 (coefficient −0.0438, $p < 0.001$): relative to their expected weight, births after a larger measured circumference are more predictable.

In grams the standard deviation is the mean multiplied by $(\exp(\sigma^2) - 1)^{1/2}$. It therefore rises with every variable that raises the mean, for example by 71 g between 12 and 39 days to delivery. Along AC the falling relative spread and the rising mean roughly cancel: +7 g between 29.3 and 33.5 (95% CI −8 to +22 g).

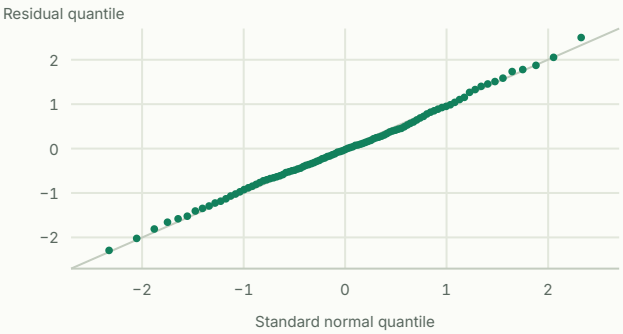
06 / CHECKS

No major contradiction; three checks open

Every required check is on record with its method and result. “Supported” means that the automated screen found no major contradiction, not that an assumption is proven.

Check	Final fit, n = 1,038	Development folds	Status
Numerical fit and estimability	Converged with finite, estimable coefficients; stationarity 2.6×10^{-9}, limit 1.0×10^{-4}.	No discrepancy in 5 of 5 folds.	**supported**
Outcome support	All outcomes finite and strictly positive, as the model family requires.	No discrepancy in 5 of 5 folds.	**supported**
Coefficient inference	Model-specific covariance calculated (mgcv Bayesian covariance).	No discrepancy in 5 of 5 folds.	**supported**
Residual distribution shape	Q-Q distance 0.054, limit 0.25. Shapiro-Wilk W = 0.9944, p = 0.00065; this test alone does not veto.	Q-Q distance 0.080 to 0.130; no fold flagged.	**supported**
Mean functional form	Residual-pattern regression: p = 0.00074; the pattern explains 4.4% of the residual squared magnitude. Strongest contribution: FL.	Not assessable: a fold is too small for the planned check.	**follow-up**
Modelled dispersion	p = 0.018; pattern share 2.9%. No flag, because p is not below 0.01.	Not assessable: a fold is too small for the planned check.	**supported** final fit only
Original-scale mean calibration	Relative bias −0.02% (99% CI −0.74% to +0.70%); tolerance ±5%.	Bias −1.1% to +0.7%; no fold flagged.	**supported**
Original-scale mean after retransformation	p = 0.0011; pattern share 4.0%. Strongest contribution: FL.	Not assessable: a fold is too small for the planned check.	**follow-up**
Original-scale mean sensitivity	Effective number of contributions 228.5; minimum 20.	25.0 to 53.8; no fold flagged.	**supported**
Independence of errors	Requires study-design, observation-order or grouping information.		**not assessed**

Decision policy research_v3. A residual pattern is flagged when p < 0.01 and the pattern explains more than 2% of the residual squared magnitude; above 20% it counts as major. Development results need agreement in 3 of 5 folds. The pattern checks need at least 8 evaluation rows for each of their 24 tested terms, which a fold of 166 or 167 births does not provide.



Residual shape. 99 quantiles of the centred, unit-SD residuals of log birth weight against the normal distribution; summarized by the Q-Q distance in the table.



Observed against fitted. Fitted means from the model equation of Section 08 for all 1,038 births; fit R^2 0.7881, RMSE 291.7 g.

What the formal recommendation still requires

OptiMod issues its formal recommendation only once every required check is complete, in the folds and in the final fit. LS1’s three pattern checks could not run in the folds, nor could those of any other candidate; two final-fit checks suggest a follow-up on FL. The status “no adequate model” records these open checks, not a major contradiction in LS1.

Supplementary evidence and gaps

Rerun with a relaxed row floor, the three open checks of LS1 were supported in aggregate, with no major pattern in any fold (one of five folds flagged in each). This relaxation is not calibrated and does not replace the required checks. Breusch-Pagan, variance inflation factors, Durbin-Watson and Cook’s distance for LS1 were not computed.

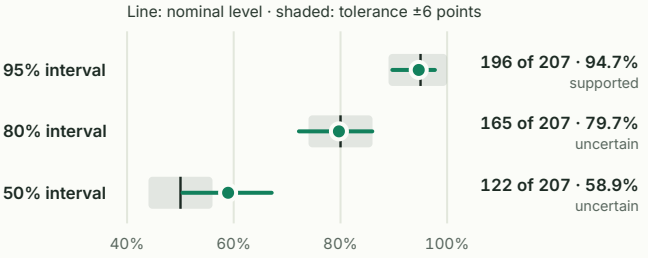
07 / VALIDATION AND SENSITIVITY

How far the result travels

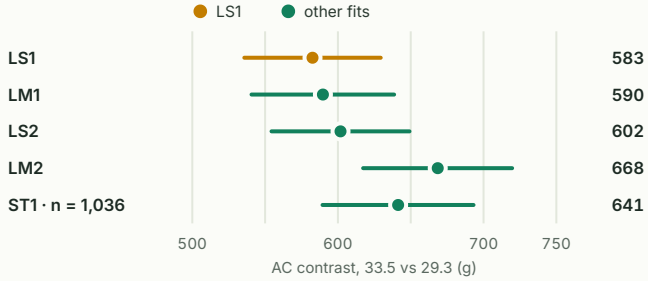
Cross-validation guided the search; 207 births were held out. A separate analysis repeats the search without the two flagged observations.

Model	Evaluation	n	R^2	RMSE g	MAE g	Log loss
LM1 · retained by the engine	Five-fold cross-validation	831	0.7857	297.08	225.61	7.0898
	Holdout audit	207	0.7577	293.74	223.29	7.0839
LS1 · best fit, presented here	Five-fold cross-validation	831	0.7879	295.54	224.63	7.0750
	Holdout, post hoc	207	0.7594	292.69	222.24	7.0520

On the held-out births LS1's prediction error matches its cross-validation error. Computed after the results were known, these figures describe LS1 and cannot validate it independently: the protocol's single audit belonged to LM1, locked before the holdout was seen. It found a major pattern in LM1's constant dispersion (21.8%, boundary 20%; 18.2% without row 742). LS1 models the dispersion; its dispersion check on the final fit is supported.



Prediction intervals of LM1 on the holdout. Point: observed coverage with exact interval (Clopper-Pearson, Bonferroni-adjusted). The 95% interval is supported; the 80% and 50% intervals stay uncertain. LS1, post hoc: 194 of 207 births (93.7%) inside its 95% interval.



One finding across five fits. The abdominal-circumference contrast with its 95% interval in the four original-population fits and in the model retained for the sensitivity population.

Stability across the four original-population fits

Contrast	LS1, g	Range, g	Interval excludes zero in
AC · 33.5 vs 29.3	+583	+583 to +668	all four fits
FL · 7.1 vs 6.3	+188	+91 to +205	all four fits
BPD · 9.1 vs 8.2	+111	+92 to +113	all four fits
DBD · 39 vs 12	+807	+788 to +853	all four fits
HC · 32.6 vs 29.9	+89	+41 to +105	LS1, LM1
Parity · 1 vs 0	+87	+77 to +98	all four fits
Age · 39 vs 31	−23	−37 to −23	LM1, LM2, LS2

Every contrast has the same direction in all four fits. The fits share the same births: this is a descriptive comparison, not an independent replication. The engine records finding stability as not assessed.

Without the two flagged births

Rows 464 and 815 combined a large conditional residual (4.4 and 4.9 standard units) with high influence (Cook's distance 0.026 and 0.063) in the development screen. A flag is not a confirmed error, so both stay in the main analysis.

A separate search without them retained the Student-t regression ST1. The focal associations keep their direction: AC +641 g (589 to 693), FL +117 g (62 to 171), BPD +85 g (21 to 150).

The interaction retained there is a different one, AC × FL (RMSE 1.5% lower). Both interaction findings are exploratory, and no winner across populations is declared.

Sensitivity population, n = 1,036	AIC	Δ AIC	Fit R^2	CV RMSE g	CV MAE g	CV log loss	Role
ST1 · Student-t · AC × FL	14,641.34	0.00	0.7996	284.06	220.64	7.0621	Retained; checks to review
LM3 · Linear	14,684.50	43.16	0.7938	288.56	224.89	7.0844	Comparison
LM4 · Linear · AC × FL	14,655.94	14.60	0.7998	284.12	221.03	7.0693	Comparison
LS3 · Location-scale · AC × FL · σ: AC	14,652.61	11.27	0.7997	284.15	220.96	7.0714	Comparison

All four on the original outcome scale. Development outcomes helped to select the excluded rows, so these scores are post-screened and descriptive; they do not validate the exclusion. Compare AIC only within one population.

08 / STATISTICAL SPECIFICATION

The fitted model, in full

LS1: Gaussian location-scale model for log birth weight; mgcv penalized likelihood with REML; no smooth terms; 13 parameters; 1,038 births.

$$\log Y = \mu(x) + \sigma(x) \cdot \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, 1)$$

$$\mu(x) = \beta_0 + \beta_{\mathrm{AC}}\, z_{\mathrm{AC}} + \beta_{\mathrm{BPD}}\, z_{\mathrm{BPD}} + \beta_{\mathrm{FL}}\, z_{\mathrm{FL}} + \beta_{\mathrm{HC}}\, z_{\mathrm{HC}} + \beta_{\mathrm{age}}\, z_{\mathrm{age}} + \beta_{\mathrm{DBD}}\, z_{\mathrm{DBD}} + \beta_{\mathrm{FL\times DBD}}\, z_{\mathrm{FL}}\, z_{\mathrm{DBD}} + \beta_1\,[\text{parity 1}] + \beta_2\,[\text{parity 2}] + \beta_{3+}\,[\text{parity 3+}]$$

$$\sigma(x) = 0.01 + \exp(\alpha_0 + \alpha_{\mathrm{AC}} \cdot \mathrm{AC})$$

$$\mathrm{E}(Y \mid x) = \exp(\mu + \sigma^2/2) \;\cdot\; \text{median} = \exp(\mu) \;\cdot\; \mathrm{SD}(Y \mid x) = \mathrm{E}(Y \mid x) \cdot (\exp(\sigma^2) - 1)^{1/2} \;\cdot\; z_v = (v - \text{centre}) \,/\, \mathrm{SD}$$

Term	Estimate	Std. error	z	p	95% confidence interval
Mean equation · scale log(birth weight) · numeric predictors standardized					
Intercept	8.03965	0.00424	1,896.19	<0.0001	8.03134 to 8.04796
AC	0.16559	0.00676	24.48	<0.0001	0.15233 to 0.17885
BPD	0.03060	0.00867	3.53	0.0004	0.01360 to 0.04759
FL	0.04777	0.00725	6.59	<0.0001	0.03356 to 0.06198
HC	0.02657	0.00961	2.76	0.0057	0.00773 to 0.04540
parity 1	0.02772	0.00624	4.45	<0.0001	0.01550 to 0.03995
parity 2	0.02450	0.00938	2.61	0.0090	0.00612 to 0.04288
parity 3+	0.03767	0.01295	2.91	0.0036	0.01229 to 0.06305
age	−0.00529	0.00279	−1.89	0.0584	−0.01077 to 0.00019
DBD	0.20699	0.00458	45.16	<0.0001	0.19800 to 0.21597
FL × DBD	−0.01701	0.00250	−6.81	<0.0001	−0.02191 to −0.01211
Dispersion equation · scale log(σ − 0.01) · AC in recorded units					
Intercept	−1.18161	0.21941	−5.39	<0.0001	−1.61164 to −0.75157
AC	−0.04377	0.00702	−6.24	<0.0001	−0.05753 to −0.03001

Reference category: parity 0. Standard errors, z tests and intervals come from the mgcv Bayesian covariance; they are conditional on this model and not adjusted for the model search or for multiple comparisons.

Centres and scales of the standardized predictors

Predictor	Centre	SD
AC	31.184817	3.651252
BPD	8.618497	0.761797
FL	6.637486	0.670434
HC	30.986609	2.476156
age	34.917148	5.626847
DBD	28.112900	21.580014

Fit statistics

Quantity	Value
Log-likelihood, original outcome	−7,326.344
AIC	14,678.688
Fit R^2, original scale	0.7881
RMSE, fitting sample	291.706 g
Cross-validated RMSE	295.54 g
Cross-validated log loss	7.0750

The numbers can be recomputed. Applied to the 1,038 rows of the public dataset, the coefficients and constants on this page reproduce the log-likelihood, AIC, fit R^2 and RMSE stated above; with the coefficient covariance matrix they also reproduce every contrast of Section 04 and its interval (delta method).

09 / LIMITS AND PROVENANCE

Scope, limits and next steps

The limits are part of the result. They state where the evidence ends and what a next analysis would have to add.

- **Associations, not causes.** Observational data; unmeasured confounding is not assessed.
- **Not a prospective predictor.** The time from scan to delivery is unknown at the scan.
- **Formal recommendation.** Not issued yet: three required checks could not be completed in the development folds; two final-fit checks suggest a follow-up.
- **Presentation choice.** LS1 was selected for this report after the results were seen. Its intervals and p-values are conditional and exploratory.
- **Bounded search.** 280 model questions were deferred; additional structure may remain undiscovered.
- **Optimism of cross-validation.** Adaptive cross-validation guided the search, so its scores can be optimistic. The refit on all births has not been revalidated.
- **Prediction intervals.** Coverage on 207 holdout births is uncertain for the 80% and 50% intervals and says nothing about subgroups.
- **Possible duplicates.** 5 rows repeat an earlier row exactly. If they are copies of one record, held-out performance can look optimistic and intervals too narrow.
- **Extrapolation.** Two holdout births (rows 1 and 934) lie outside the development range of AC and HC.
- **Population sensitivity.** Without the two flagged births another model and another interaction are retained; no winner across populations.
- **Documentation gaps.** Units of the ultrasound measurements are undocumented; the package labels DBD as date of birth.
- **Independence.** Not assessed; it requires study-design, observation-order or grouping information.

Suggested next steps

1. **Define the intended prediction time.** A model for use at the scan cannot contain the actual time to delivery.
2. **Assess the recorded follow-ups.** Compare alternative functional forms for FL in the mean; do not repair the model on the holdout.
3. **Use new validation data.** Holdout outcomes that were already inspected cannot validate the later preference for LS1.
4. **Review the two flagged births.** Decide on substantive grounds whether they belong to the target population; keep both analyses.
5. **Calibrate the policy more broadly.** Evaluate the provisional diagnostic thresholds and the selection rule before stronger inference or clinical claims.

Audit trail

Item	Record
Software	OptiMod backend 0.4.0 · R 4.6.1 · mgcv 1.9-4 · gamlss.data 6.0.7 · macOS arm64
Run	Results saved 16 November 2026 · seed 20260911 · assumption policy research_v3 · population policy fixed_populations_v1
Request	Unmodified default: 120 s, 24 candidates, at most 2 batches · 11 candidates used for the original and 13 for the sensitivity population
Data	gamlss.data::ultra · 1,038 rows, 8 variables · development 831, holdout 207, five folds · sensitivity population 1,036
Policy hash	`f210d06df1f87ab240901fd0167a0fddf8401091a07214885f1eecf0a02d853a`
Configuration hash	`4b0aa401b1102185484686199d61eb51ce1073d65e715cfbe8edfb1f232ff7bf`
Engine build	Pre-release revision `2ba74fbb` with a saved local patch; not a released main revision.
This document	Sample report. Text and layout were prepared from the engine's Scientific Result Report, Extended Technical Report and results window. No model was fitted, selected or re-estimated for it.

Sources

Data. Stasinopoulos, M., Rigby, R. and De Bastiani, F.: gamlss.data, Data for Generalised Additive Models for Location Scale and Shape, R package version 6.0-7, dataset `ultra`. **Reference analysis.** Stasinopoulos, M. D., Kneib, T., Klein, N., Mayr, A. and Heller, G. Z. (2024): Generalized Additive Models for Location, Scale and Shape. Chapter 8, Table 8.1. The published analysis is a reproducible reference, not known truth. **Method.** Wood, S. N.: mgcv, Gaussian location-scale family `gaulss`.