The beauty of descriptive models – and bodyfat (the beast)

Georg Heinze

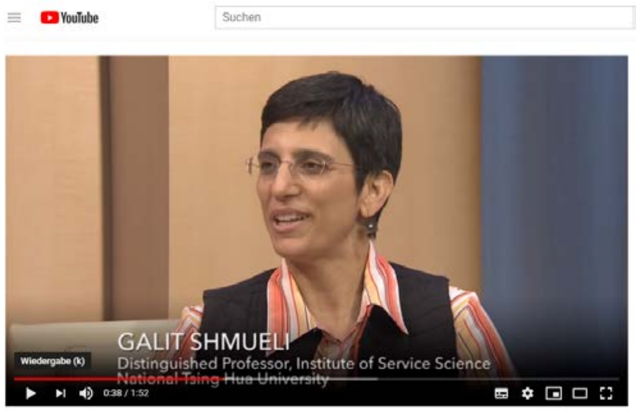
Institute of Clinical Biometrics

Outline

- On descriptive modeling
- The (ir)relevance of variable selection
- New task:
  predicting the percentage of body fat with a 5-parameter linear model

Purpose of models: To Explain or to Predict?

- **Predictive models**
  - Interest in predicting outcome for future application.
  - “Predict how outcomes will be, given the predictors.”
- **Explanatory models**
  - Interest in inferring causal effects of interventions on outcome.
  - “Explain why outcomes differ depending on the intervention.”
- **Descriptive models**
  - Interest in describing the data structure parsimoniously.
  - “Describe how outcome varies with predictors.“
- **A given modeling task may have several dimensions!** (e.g. descriptive-predictive)
- Similar considerations by Hernan et al, 2019; and Carlin and Moreno-Betancur, 2025



Galit Shmueli discusses the distinction between explaining and predicting (Preview)

(Shmueli, 2010)

Statistics in Medicine

FEATURED ARTICLE | Open Access

On the Uses and Abuses of Regression Models: A Call for Reform of Statistical Practice and Teaching

John B. Carlin, Margarita Moreno-Betancur

First published: 24 June 2025 | https://doi.org/10.1002/sim.10244 | Citations: 7

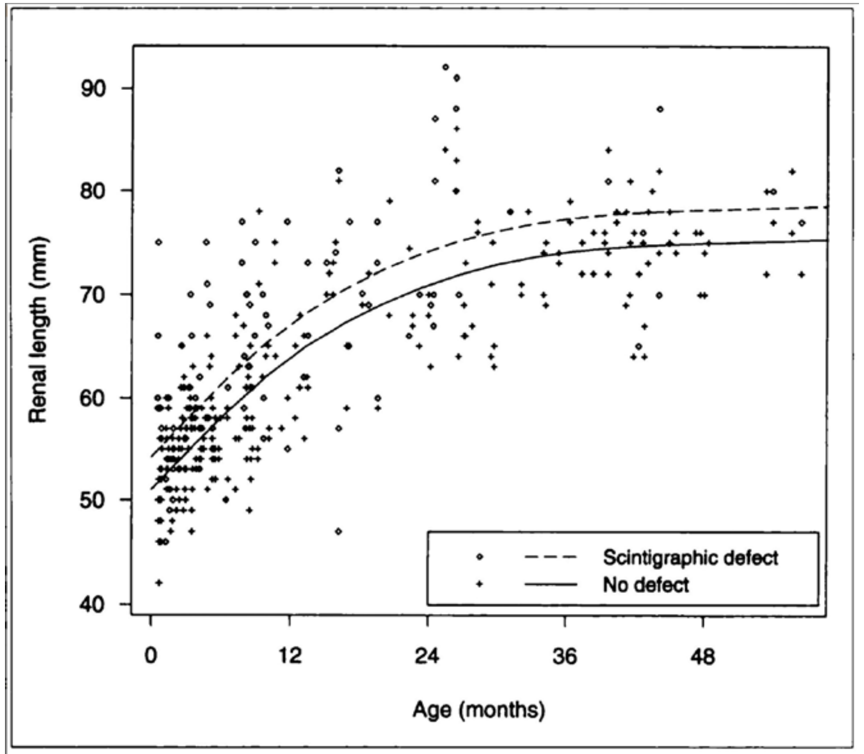


Fig. 1.—Scatter plot of renal length measured on sonograms versus age for kidneys with and without defects shown on scintigrams. Curved lines represent cubic model used in analysis of covariance calculations. Because curves are parallel, there is a similar absolute increase in renal length at all ages.

- No other modeling task than the ‚trichonomy‘ (D/P/E)
- Descriptive purpose of modeling:
  „characterizing the distribution of a feature or health outcome in a population”
- Here: “curve fitting”
- Model makes an important assumption. *Which one?*

Statistics in Medicine

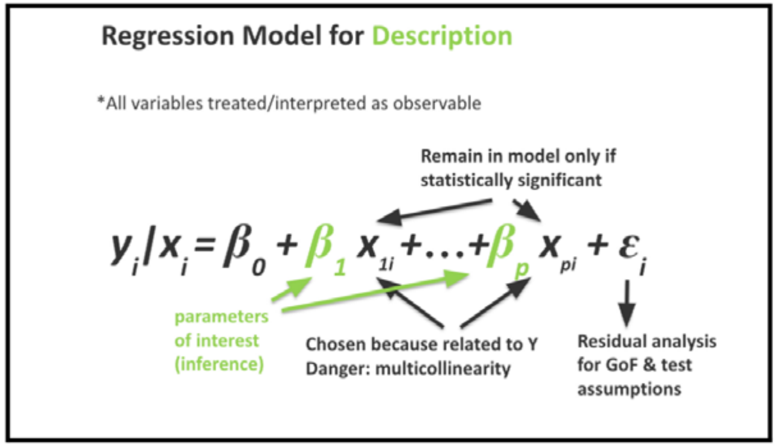
COMMENTARY | Full Access

To Explain, to Predict, or to Describe: Figuring out the Study Goal [Commentary on "On the Uses and Abuses of Regression Models" by Carlin and Moreno-Betancur]

Galit Shmueli

First published: 24 June 2025 | https://doi.org/10.1002/sim.10307 | Citations: 2

Funding: This work was partially supported by the Taiwan National Science and Technology Council, 111-2410-H-007-030-MY3.



- Unit of action: individual or collective?
  - Description of means in (sub)populations (collective),
    description of expected values with given covariates (individual, see Mariana's task),
    or description of differences between means (=associations, collective)?
- Variable names matter:
  - Predictors, confounder/collider/mediator/instrument, covariate/independent variable?

Here are two typical tasks of descriptive modeling

- To describe the association of an „outcome“ Y with the essential covariates X using a model, selected from a larger set of possible models
  - Parsimonious description using a simple statistical model
  - Variable selection, simple functional forms
  - Easy to derive partial associations from the model
  - Easy to make statements about expected values of Y|X
  - *Risk of overestimation/overinterpretation of partial associations, neglecting others*
  - *Risk of underestimation of uncertainty*
- To describe a non-causal association of an outcome Y with an independent variable/exposure X, possibly adjusted for covariates
  - ‚Curve fitting‘:
    - smoothing as variance-bias trade-off
    - separate systematic from unsystematic variation
    - simple equations preferred for description
      (*try to publish a formula with truncated power functions in a medical journal!*)
  - Example: Kidney length difference by age, CV risk by Lipoprotein(a), BMI by age+sex

A new task: predicting body fat

- Background:
  - Body fat measurement difficult; needs underwater density measurements to estimate lean body mass;
    fat mass = total mass – lean mass
    bodyfat% = fat mass / total mass x 100%
  - Various models to predict body fat have been published, e.g. based on NHANES data (Stevens et al, 2016)
  - Two data sets were published by Johnson (1996, 2021)

- Research question:
  The percentage of bodyfat should be predicted using 5 out of a larger set of anthropometric variables.
  The model should be ‚describable' but also predictive: descriptive-predictive modeling task.
- Data set:
  - Training: Data set provided by Johnson (2021): 184 college women aged 18-25
  - Validation: Data set provided by Johnson (1996): 252 men aged 18-80

Predictor to choose from

- Weight (kg)
- Height (m)
- BMI: (Body Mass Index) Weight divided by the square of Height
- Age
- Neck: Minimal circumference perpendicular to the long axis of the neck (cm)
- Chest: Horizontal plane measurement at the sixth rib, at the end of a normal expiration (cm)
- Calf: Horizontal maximal calf measurement (cm)
- Biceps: Measurement with arm extended (cm)
- Hips: Horizontal maximal measurement around buttocks (cm)
- Waist: Horizontal minimal measurement, at the end of a normal expiration (cm)
- Forearm: Maximal measurement perpendicular to long axis (cm)
- PThigh: (Proximal Thigh) Horizontal measurement immediately distal to the gluteal furrow (cm)
- MThigh: (Middle Thigh) Measurement midway between the midpoint of the inguinal crease and the proximal border of the patella (cm)
- DThigh: (Distal Thigh) Measurement proximal to the femoral epicondyles (cm)
- Wrist: Measurement perpendicular to the long axis of the forearm (cm)
- Knee: Measurement at the mid-patellar level, with the knee slightly flexed (cm)
- Elbow: A minimal circumference measurement with the elbow extended (cm)
- Ankle: Minimal circumference measurement perpendicular to the long axis of the calf (cm)

Estimands

- Selection of predictors (binary)
- Regression coefficients of predictors and their uncertainty

descriptive

- External validation: performance and transportability of the model
  - R-squared
  - Root mean squared prediction error
  - Calibration slope + intercept

predictive

Proposed model and modeling approach

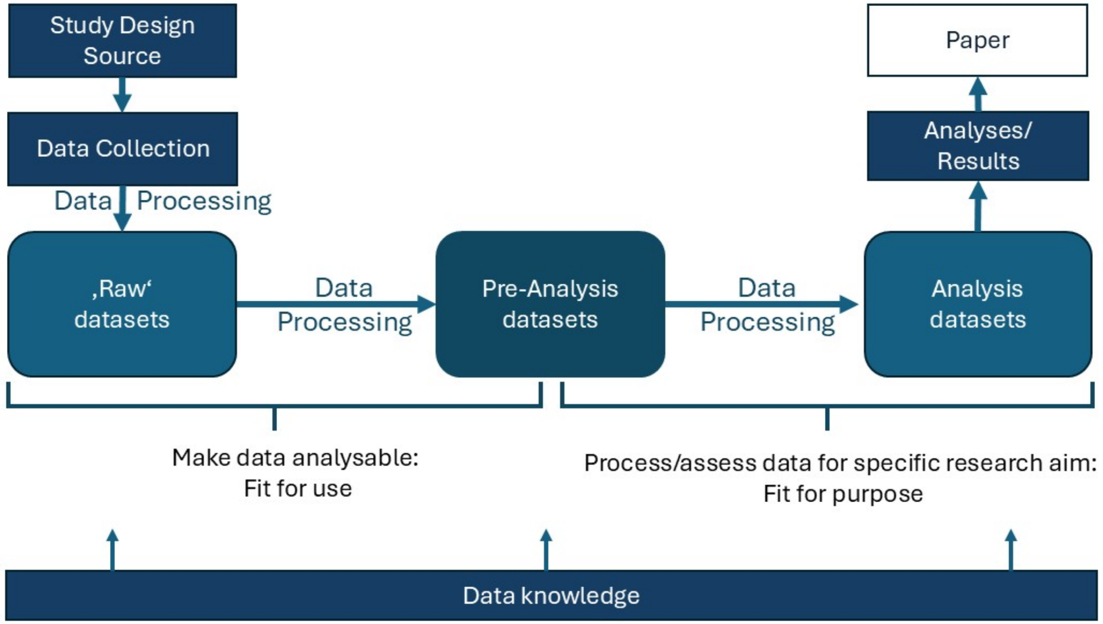
- Linear regression model with linear functional forms only
  - One could argue that %bodyfat is restricted between 0-100% -> zero-inflated beta model?
  - Variable selection to obtain max. 5 predictors
    - Simple and describable model
    - Relaxed Lasso with fixed model size (5)
- Assessing apparent validity
  - Overall (local) calibration: residuals vs. linear predictor
  - Validity of linearity assumption: residuals vs. included predictors
  - Independence of residuals: residuals vs. excluded predictors
  - Distributional assumption: normality of residuals

Model description

- Model should be ‚describable' (some call that ‚explainable':
  - Regression coefficients
  - Standardized regression coefficients (Δ%bodyfat per SD(x))
  - Drop-in-R^2
- Model stability (using the bootstrap):
  - Selection probabilities of each predictor
  - Variability of predicted probabilities (prediction stability plot, Riley & Collins 2023)
  - Probability of selecting the final set of predictors

Data preparation

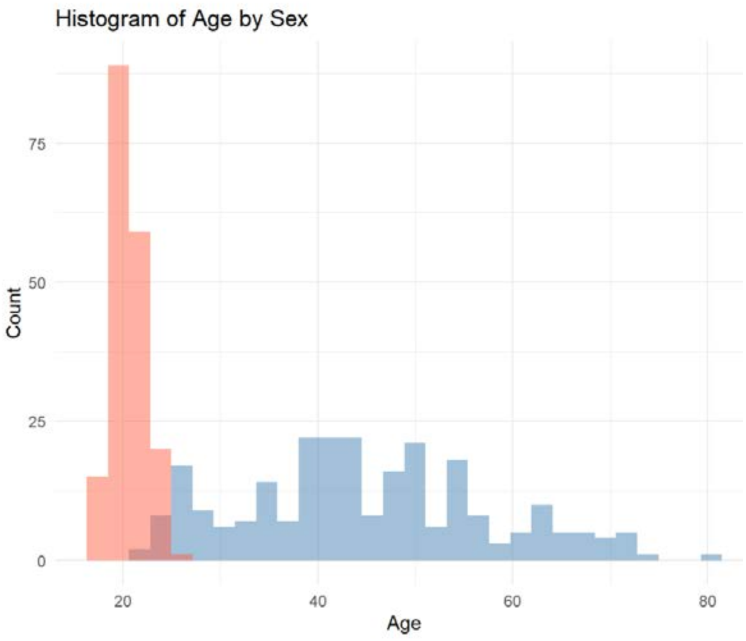
- From Schmidt, Klinger, Sauerbrei, Heinze (2025), Biometric Bulletin 42(2) (current issue):

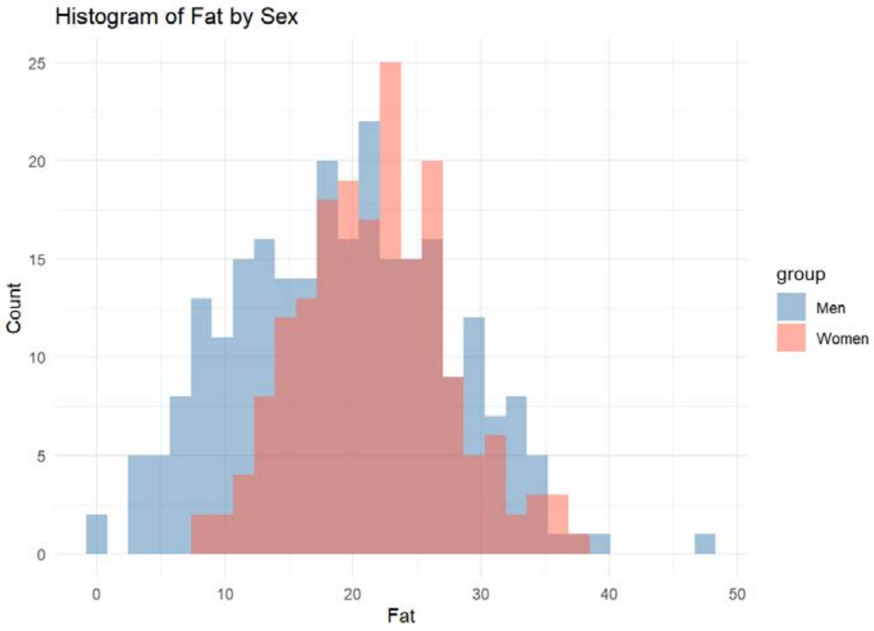


- The two data sets have in-built data errors, which need to be corrected
- Some variables have different names, some have somewhat different definitions

Data preparation

- Data errors were corrected (set to NA)
- Data sets were harmonized
- IDA was performed





Statistical Analysis Plan

- We created a SAP according to SAPI guideline (currently under development, see https://osf.io/zfjah/
- The SAP integrates main analysis and Bayesian robustness check
- Analyses performed in our shared GitHub repository

Statistical analysis plan (simplified): Predicting body fat proportion using anthropometric measures

Georg Heinze

SAPI Version 1.1, June 2025

1 Administrative information

1.1 Project title

Prediction of the percentage of body fat and validation

1.2 Project documents

This project is embedded in STRATOS TG2 project P6, Bayesian-Frequentist contrast, in which four case studies are analysed with hypothetical task descriptions, analysis plans, analyses and replication analyses using either Bayesian or frequentist paradigms for analysis and the respective other paradigm for the replication.

This analysis project links to the task description "Predicting body fat proportion using anthropometric measures - simplified task description" and to the GitHub repository `https://github.com/aliaksah/Bayes_for_STRATOS`, subfolder `https://github.com/aliaksah/Bayes_for_STRATOS/TaskBodyfat`, where data preparation and Initial Data Analysis can be found in the Quarto and html documents `Task Bodyfat simplified.qmd` and `Task Bodyfat simplified.html`.

1.3 Ethics approval

Since this project is related to published data set, ethics approval is not needed.

1.4 Names and contact

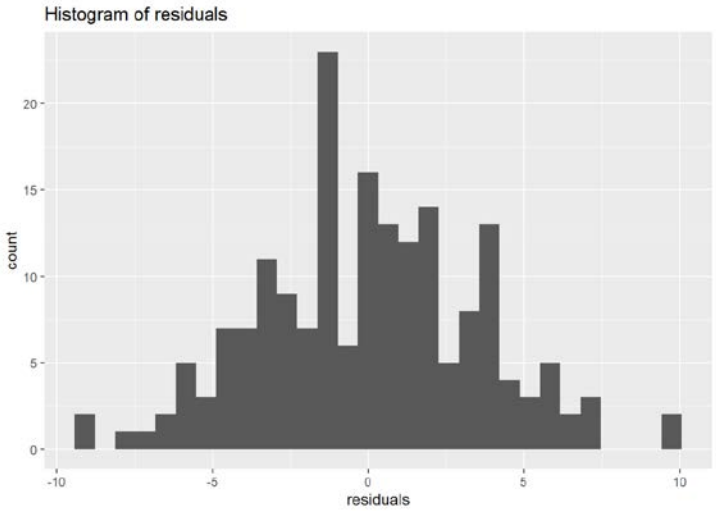
- Main Author: Georg Heinze (`georg.heinze@meduniwien.ac.at`)
- Author of Bayesian replication: Mariana Nold (`mariana.nold@uni-jena.de`)

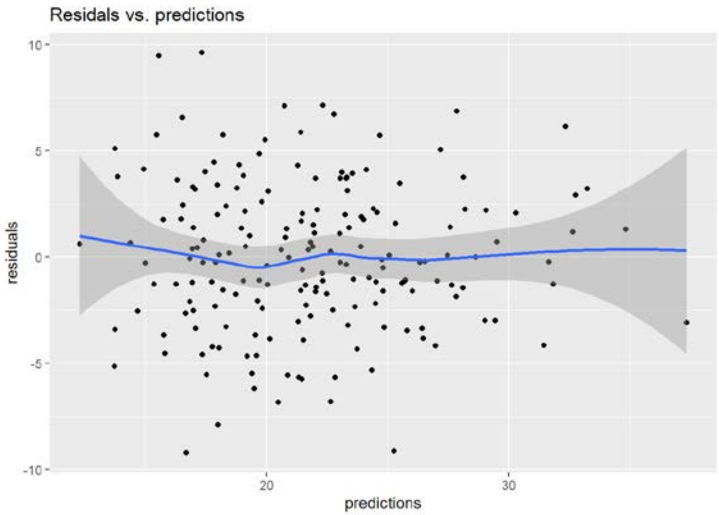
Results

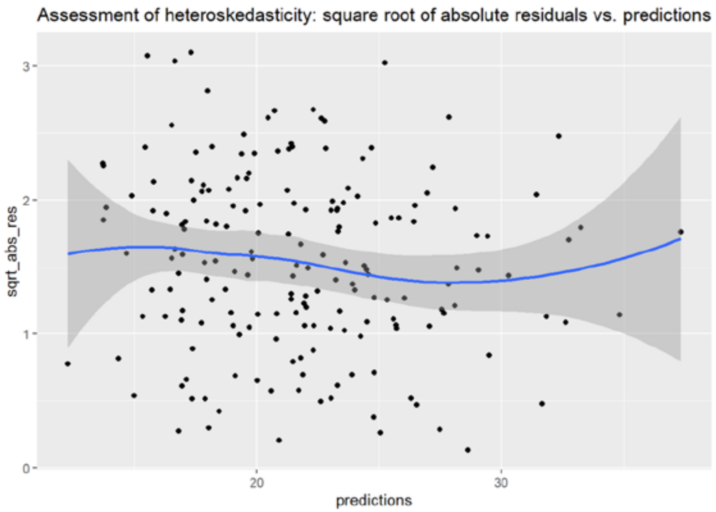
Regression coefficients and stability measures

	Regression coefficient	Standardized beta	Variable inclusion frequency	Lower 95% confidence limit	Upper 95% confidence limit
(Intercept)	-35.8567775	-35.8567775	1.000	-46.0949017	-0.8635916
Waist	0.2074769	1.3338429	0.978	0.0524680	0.4220547
BMI	0.7215750	1.9475939	0.977	-0.2001433	1.2139484
Hips	0.1632254	0.9994807	0.958	0.0000000	0.4697224
Knee	0.2594271	0.5612850	0.806	0.0000000	0.7050692
Biceps	0.0942895	0.2318775	0.570	0.0000000	0.4737582
Height	0.0000000	0.0000000	0.277	-28.0185071	0.0000000
Ankle	0.0000000	0.0000000	0.084	-1.0689316	0.0000000
MThigh	0.0000000	0.0000000	0.080	0.0000000	0.1375382
Wrist	0.0000000	0.0000000	0.068	-1.6003007	0.0000000
Age	0.0000000	0.0000000	0.056	0.0000000	0.2471138
Neck	0.0000000	0.0000000	0.018	0.0000000	0.0000000
Chest	0.0000000	0.0000000	0.011	0.0000000	0.0000000
Weight	0.0000000	0.0000000	0.004	0.0000000	0.0000000
Forearm	0.0000000	0.0000000	0.001	0.0000000	0.0000000

Model diagnostics

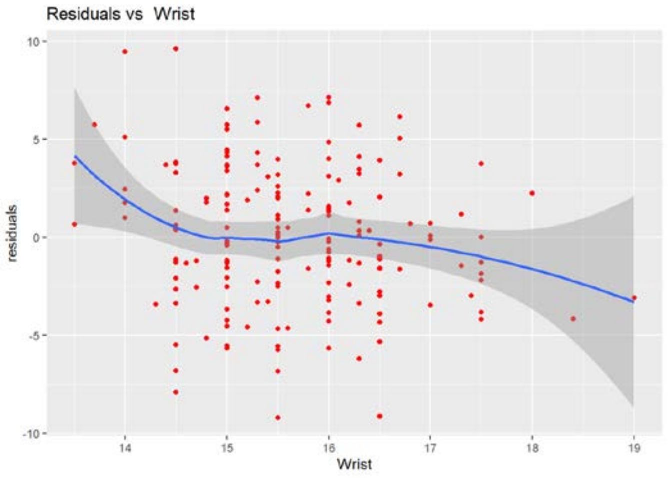


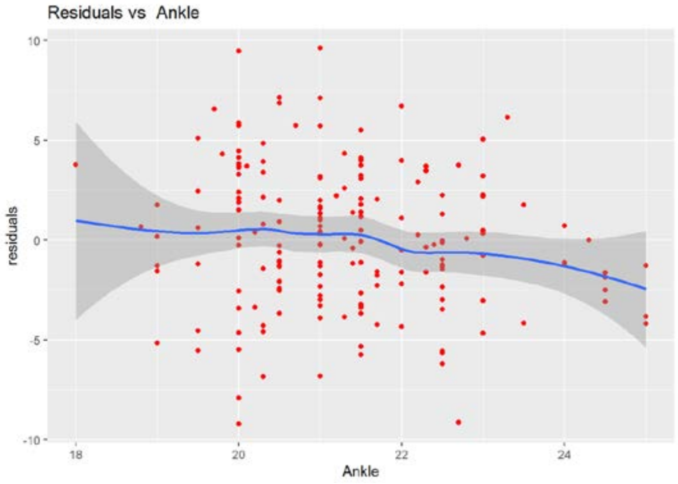


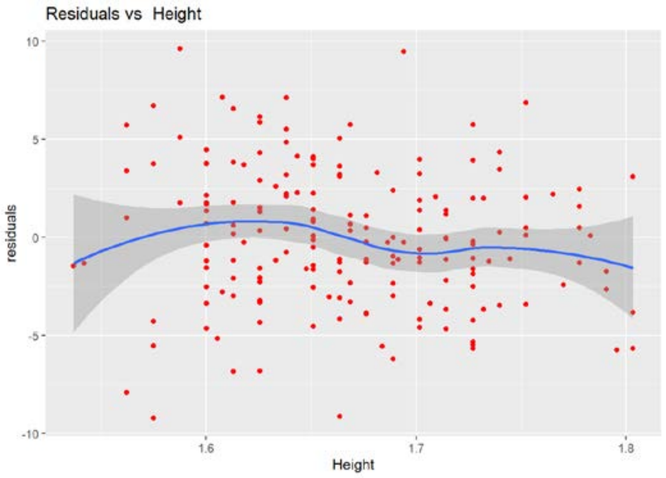


Residuals vs. excluded predictors

- The residuals vs included predictors plots did not indicate major problems.
- We also assessed residuals vs the top-3 excluded predictors (if Lasso path is extended to 8 variables):



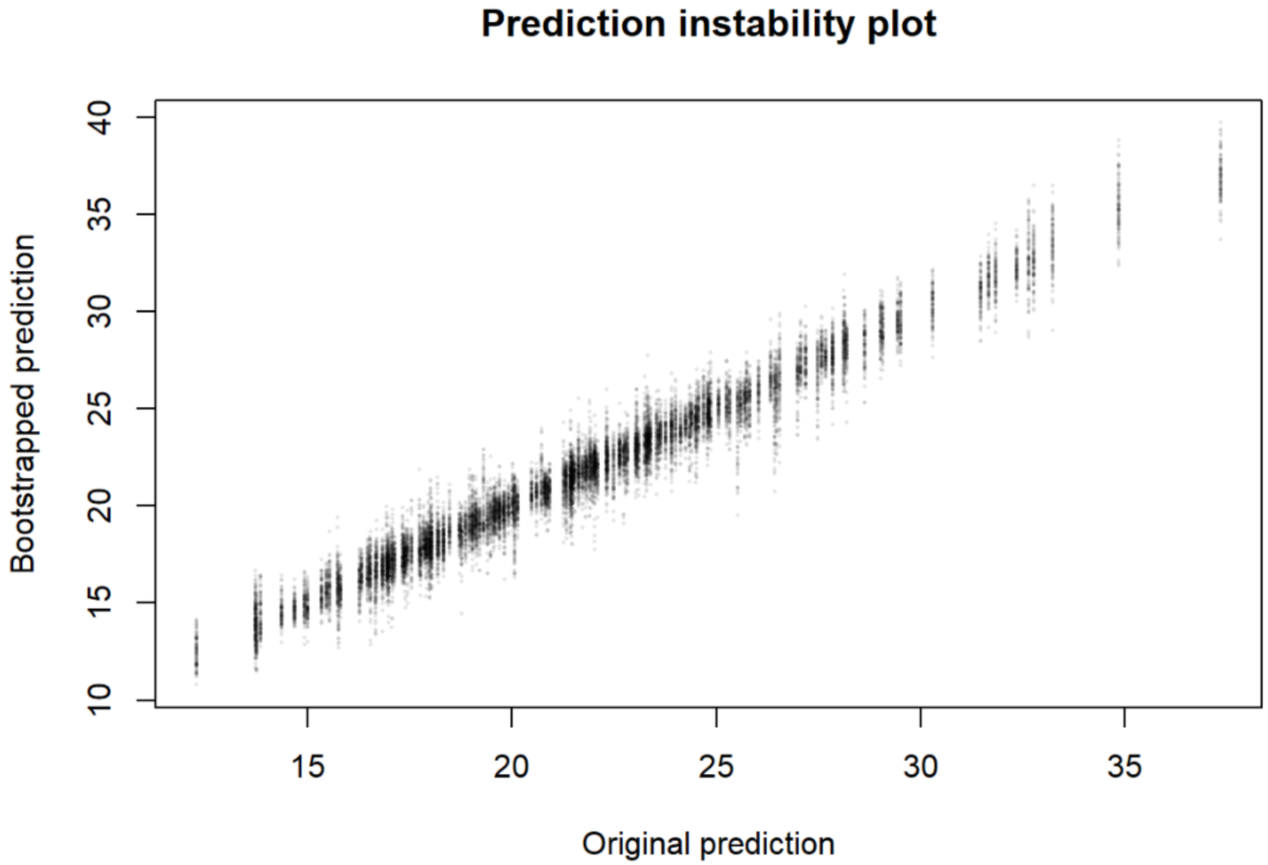




- These variables are most likely to show issues, as they were ‚next on the list to be included‘

Prediction instability plot

- Bootstrap RMSE of fitted values: 0.823
- Model-based RMSE of fitted values: 0.640
- → Variable selection produced ‚unnoticed uncertainty'!



Model validation

- Apparent:

	x
Residual error	3.5438128
Bootstrap RMSE of fitted	0.8231749
(Model-based RMSE of fitted)	0.6399374
R-squared adj.	0.6282604
Corrected R-squared	0.6139631

- At validation in men:

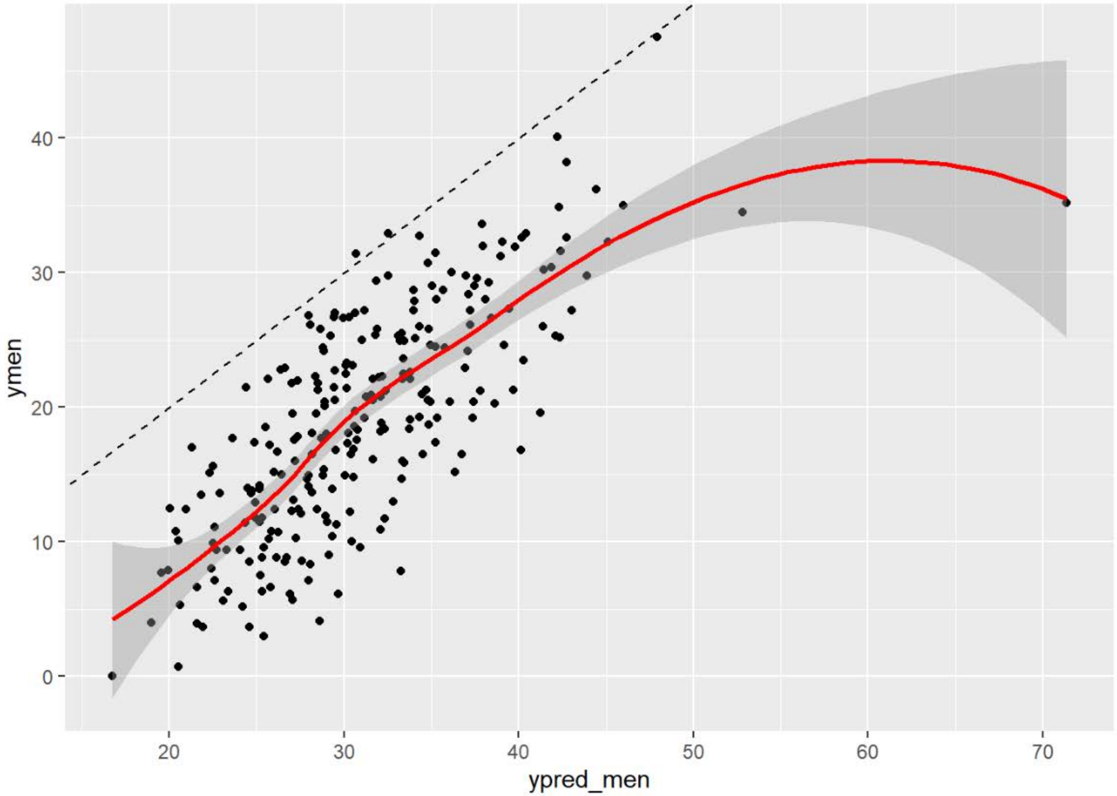
	x
Validation R-squared	-1.468229
Root mean squared prediction error	13.141202

What went wrong? Calibration plot to the rescue!

- Calibration analysis revealed that bodyfat% is overpredicted 11.9%-points
- But Calibration slope is almost ideal

Calibration statistics

	x
Calibration intercept	-11.9228028
Calibration intercept, lower 95% CL	-12.6097611
Calibration intercept, upper 95% CL	-11.2358445
Calibration slope	0.9582249
Calibration slope, lower 95% CL	0.8534096
Calibration slope, upper 95% CL	1.0630403



- →The same formula can be used for women and men, but intercept has to be recalibrated

Predictive Projection – Motivation for Robustness Analysis: **Why this approach?**

It directly addresses the research question: balancing **predictive power** and **model simplicity** by combining feature selection with a **reference model** LOO-L1 strategy within predictive projection (projpred), using elastic net with relaxation — close to relaxed Lasso and original analysis. see: Piironen et al. (2020), *Projective inference in high-dimensional problems*

Additional key Advantages:

1. Post-selection inference: accounts directly for variable selection uncertainty via the reference model → the projected posterior for, for the final submodel retains uncertainty from the reference model's posterior.

2. Clean separation of selection and validation: Selection on training data only; evaluation on validation data only → Prevents selection-induced bias (automatically handled by cv_varsel(), based on **relative cross-validated pointwise utilities** computed exclusively on validation data) → Potentially addresses key objectives of bootstrap-based internal validation

Robustness Analysis – Comparison with Original Study

Consistency with Original Study:

- Same key variables selected: *BMI*, *Waist*, *Hips*, *Knee*, *Biceps*
- Final regression model shows similar coefficient estimates and uncertainty intervals

Bayesian Enhancements:

- Enables uncertainty intervals for out-of-sample R^2 via posterior predictive distribution
- Also allows uncertainty quantification for calibration intercept and slope

Validation Perspective:

- Concepts like optimism correction and resampling align more with **frequentist thinking** (sampling distributions)
- Traditional Bayesian view treats data as fixed and focuses on posterior summaries
- **Modern Bayesian prediction culture** embraces internal validation → Can be combined with methods like **bootstrap** to assess generalization

Summary of example

- Set of selected variables in agreement
- Calibration intercept and slope at external validation similar → similar overall conclusion
- Variable selection has been said to generate uncertainty that often goes unnoticed (see also Heinze et al, 2018)
- Compared to their results, we see less variation if the number of predictors is fixed in advance
  - E.g., four predictors with VIF>80%, 1 with 50%, next one with <25%
  - Still, additional uncertainty in predictions (compared to assuming the selected model was fixed in advance)
  - Bootstrap may overestimate this variability (see also Wallisch et al, 2021)
- Concept of ‚selection probability' not well defined in frequentist statistics
  - Pragmatic way of quantifying uncertainty
  - Multi-model inference not easy to explain from frequentist viewpoint