

The limits of the additive model for adult height

Ivan A. Kuznetsov^{1,*}, Sergei A. Slavskii^{1,2,3}, Tatiana I. Shashkova^{2,3,4}, Georgii A. Bazykin^{1,4}, Tatiana I. Axenovich^{2,5}, Fyodor A. Kondrashov⁶, Yurii S. Aulchenko^{2,5,7}

¹ Skolkovo Institute of Science and Technology, Moscow, Russia, ² Novosibirsk State University, Novosibirsk, Russia, ³ Moscow Institute of Physics and Technology, Moscow, Russia, ⁴ Institute for Information Transmission Problems (Kharkevich Institute), Moscow, Russia, ⁵ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, ⁶ Institute of Science and Technology, Vienna, Austria, ⁷ PolyOmica, PA 's-Hertogenbosch, the Netherlands

*Ivan.Kuznetsov@skoltech.ru

Background

Adult height inspired the first biometrical and quantitative genetic studies and is a test case trait for understanding the genetic component of heritability.

In statistical textbooks and in human genetics height serves as an example of a trait that is shaped by a process whereby the independent contributions of many genetic and environmental factors are summed up, resulting in a normal distribution.

Results from anthropological and socioeconomic studies of distinct populations suggest that human height may be distributed log-normally, which could imply that genetic and environmental factors affecting human height multiply.

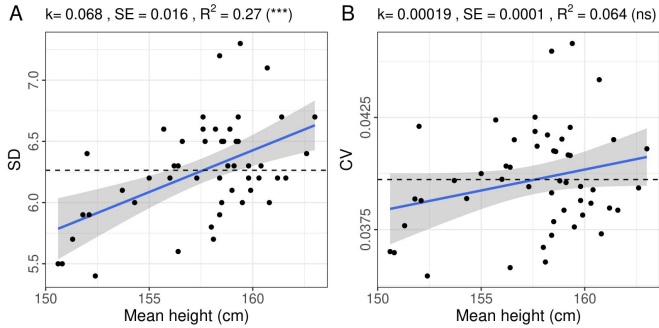
Materials

The regression analysis was conducted between SD or CV and mean height of females from 50 developing countries.

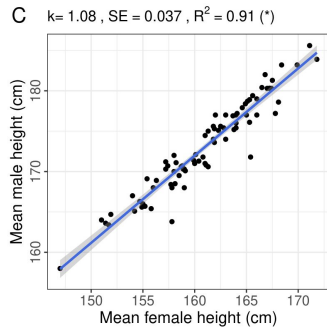
The regression analysis also was applied to explore dependence between male and female average height using data on 80 worldwide human populations.

Individual data on 369,153 white British UK Biobank participants belonging to six groups defined by place of birth were utilized for further analysis. We considered effects of sex, genotype, and residual effects. The genotype was included in analysis in the form of polygenic height score (PGHS), defined as the weighted prevalence of height-increasing alleles in the genotype. Factors related to socioeconomic status and study covariates were used to construct a single linear predictor (residual predictor, RP). All three predictors were strongly associated with height.

Results: population summary-level data



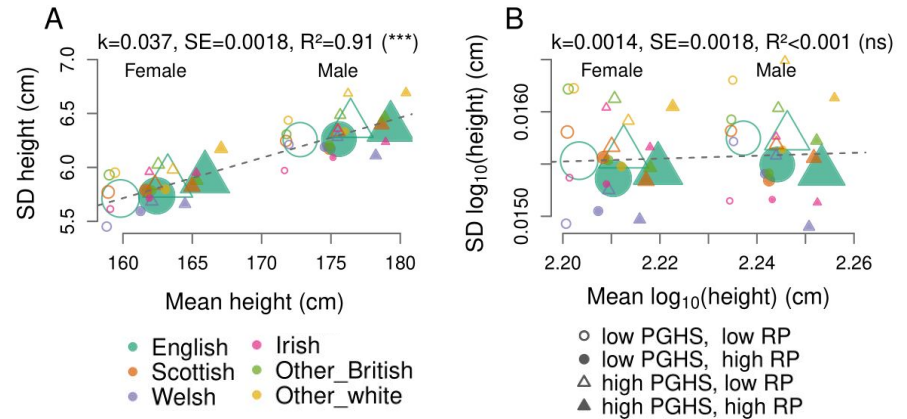
Linear regression of standard deviation (SD) and coefficient of variation (CV) of height on mean height of women from 50 countries. The dashed line shows the overall mean.



Linear regression of mean male height on mean female height in populations from 80 populations.

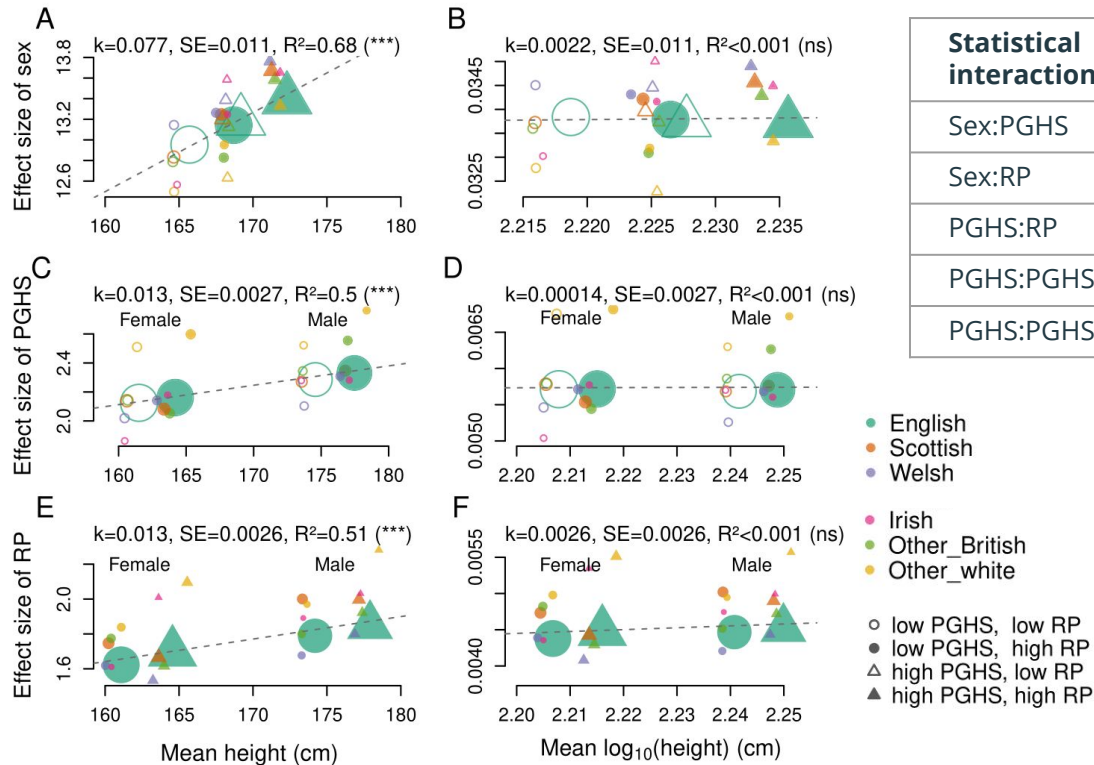
Results: individual-level data

Model	Adjusted R^2
Height ~ Sex	0.508
Height ~ Sex + PGHS	0.568
Height ~ Sex + PGHS + RP	0.603



Relation of standard deviation to mean of height (A) and log-height (B) for six groups of British individuals of white descent from UK biobank, defined based on place of birth and split by sex, median polygenic score, and median residual predictor (48 groups in total).

Results: individual-level data



Relation between the estimate of the effect size of sex (A, B), genotype (C, D; PGHS), other factors (E, F; RP) and mean height (A, C, E) and log-height (B, D, F) for groups of British individuals

Statistical interaction	P-value (Height)	P-value (log-height)
Sex:PGHS	1e-16	0.55
Sex:RP	1e-13	0.68
PGHS:RP	1e-05	0.07
PGHS:PGHS	1.32e-02	0.76
PGHS:PGHS (strong)	1e-06	0.48

Conclusion

The log-normal approximation may be a useful alternative to the normal approximation in analysis of big (hundreds of thousands of individuals) height data, analysis of heterogeneous populations, and analysis of the extremes of height.

Acknowledgements

We are grateful to Marianna Bevova and Pavel Borodin for fruitful discussion and help with conceptualising our findings.