

Conversion (cross-walking) of body-mass index (BMI) metrics

NCD-RisC work on body-mass index (BMI) pools and analyses population-based studies that had measured height and weight in people aged 5 years and older to estimate mean BMI in 200 countries and territories.

Estimating mean BMI from prevalence of different BMI categories

In ~0.5% of our data points for children and adolescents (aged 5-19 years) and ~2% of our data points for adults (aged 20 years and older), mean BMI is not reported, but data are available for the prevalence of one or more BMI categories. These data are mostly extracted from published reports or from a previous pooling analysis.¹ In order to use these data, we developed conversion (or cross-walking) regressions to estimate mean BMI from the available data.

The dependent variable in each regression is mean BMI. The independent variable (predictor) is the prevalence of one or more BMI categories that is reported by at least one study that did not report mean BMI. All regressions include terms for age, sex, year of study, and country's income (natural logarithm of per-capita gross domestic product in 2011 international dollars), as well as regional random intercepts and interactions between predictors and age and sex, based on the Bayesian Information Criterion (BIC).² For regional random intercepts, there are 21 regions used in NCD-RisC, which are based on geography and national income.³ Mean BMI is inversed when used as a dependent variable because the relationship between inverse mean BMI and the probit of prevalence is closer to linear than relationships between other functions of mean BMI. The coefficients of these regressions are estimated using data sources from 1975 onwards, and national studies from the 3 years prior to 1975 used as 1975 studies. We excluded data points with fewer than 25 participants.

Cut-offs used to define underweight, overweight, and obesity for children and adolescents are different from those for adults and vary by age and sex because of the natural growth in

childhood and adolescence. This is taken into account when developing the cross-walking models. The regression coefficients and number of data points we use to estimate the coefficients are shown in Table 1 for children and adolescents and in Table 2 for adults.

When applying the conversion models, all sources of uncertainty, including the sampling uncertainty of the data on the available metrics, the uncertainty of the regression coefficients and random intercepts, and the regression residuals, are carried forward by using repeated draws from their respective distributions. We account for the correlation among the uncertainties of regression coefficients and random intercepts by drawing from their joint posterior distribution.

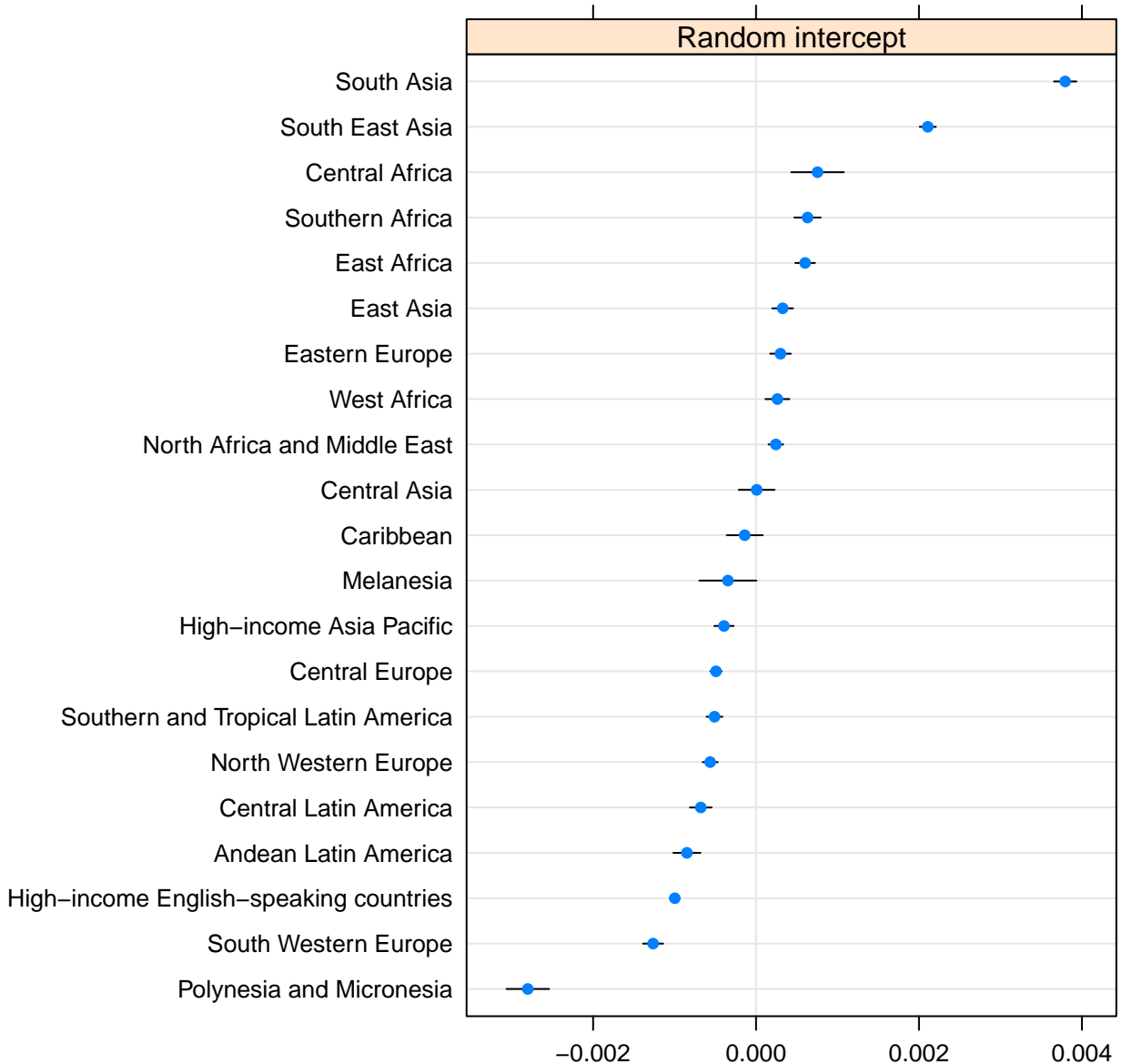
Table 1: Model specifications and regression coefficients to estimate mean BMI from other metrics for children and adolescents (aged 5-19 years).

The dependent variable in all regressions was the inverse of mean BMI, fitted using a linear mixed model. Random intercepts for regions in regression are presented after the table of coefficients.

* denotes statistical interaction. CI: confidence interval.

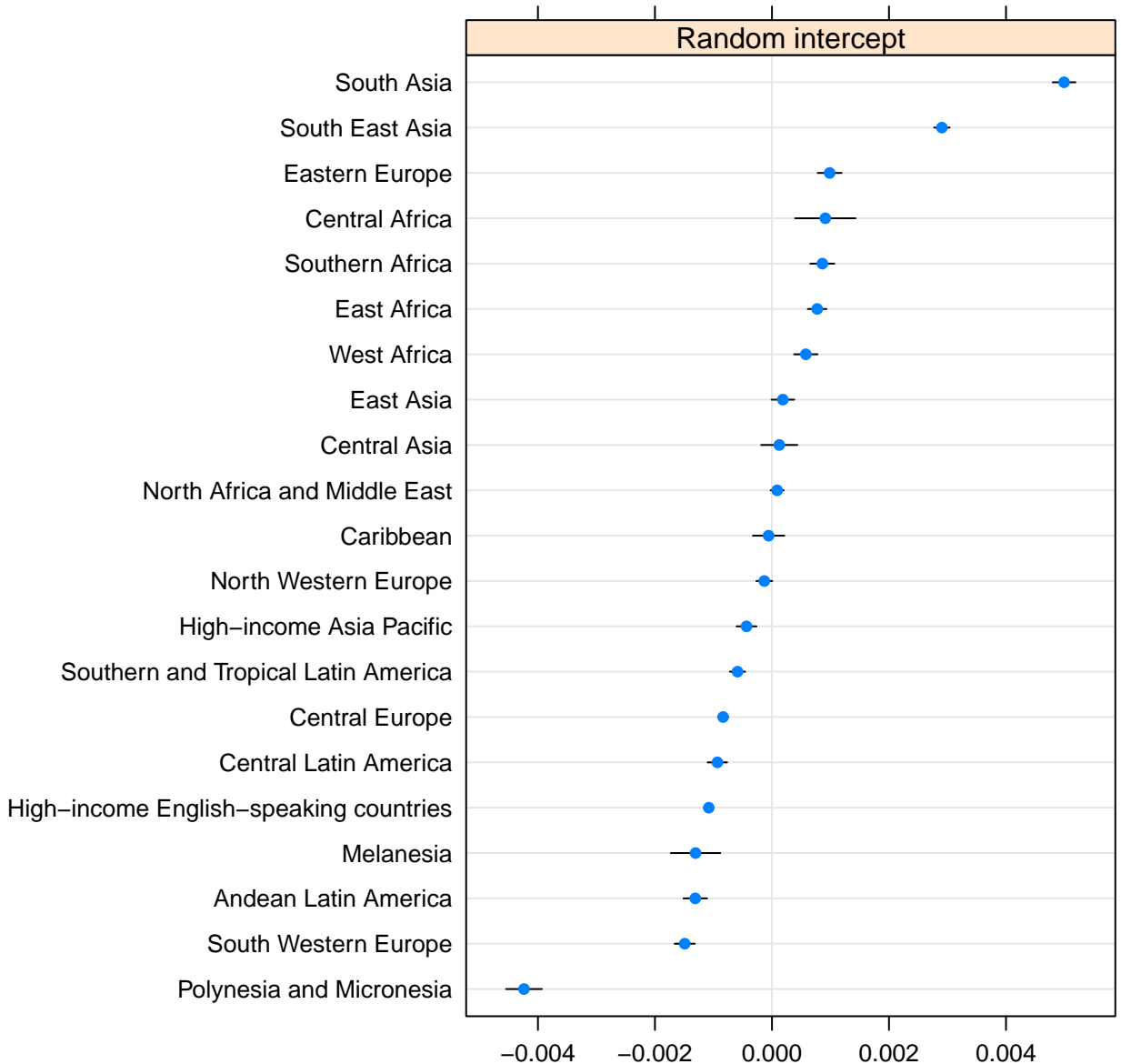
Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI ≥ 25 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.056 (0.056, 0.057)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²)	-0.0069 (-0.0070, -0.0067)
Mid-age of age group	-0.00070 (-0.00072, -0.00068)
Male sex	0.00050 (0.00036, 0.00063)
Study mid-year (per one more recent year since 1975)	-1.6e-05 (-1.9e-05, -1.3e-05)
Natural logarithm of per-capita gross domestic product	-0.00035 (-0.00040, -0.00030)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * mid-age of age group	0.00015 (0.00014, 0.00016)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * male sex	0.00010 (2.1e-05, 0.00018)
Number of data points used to fit the model = 13,773	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.947.



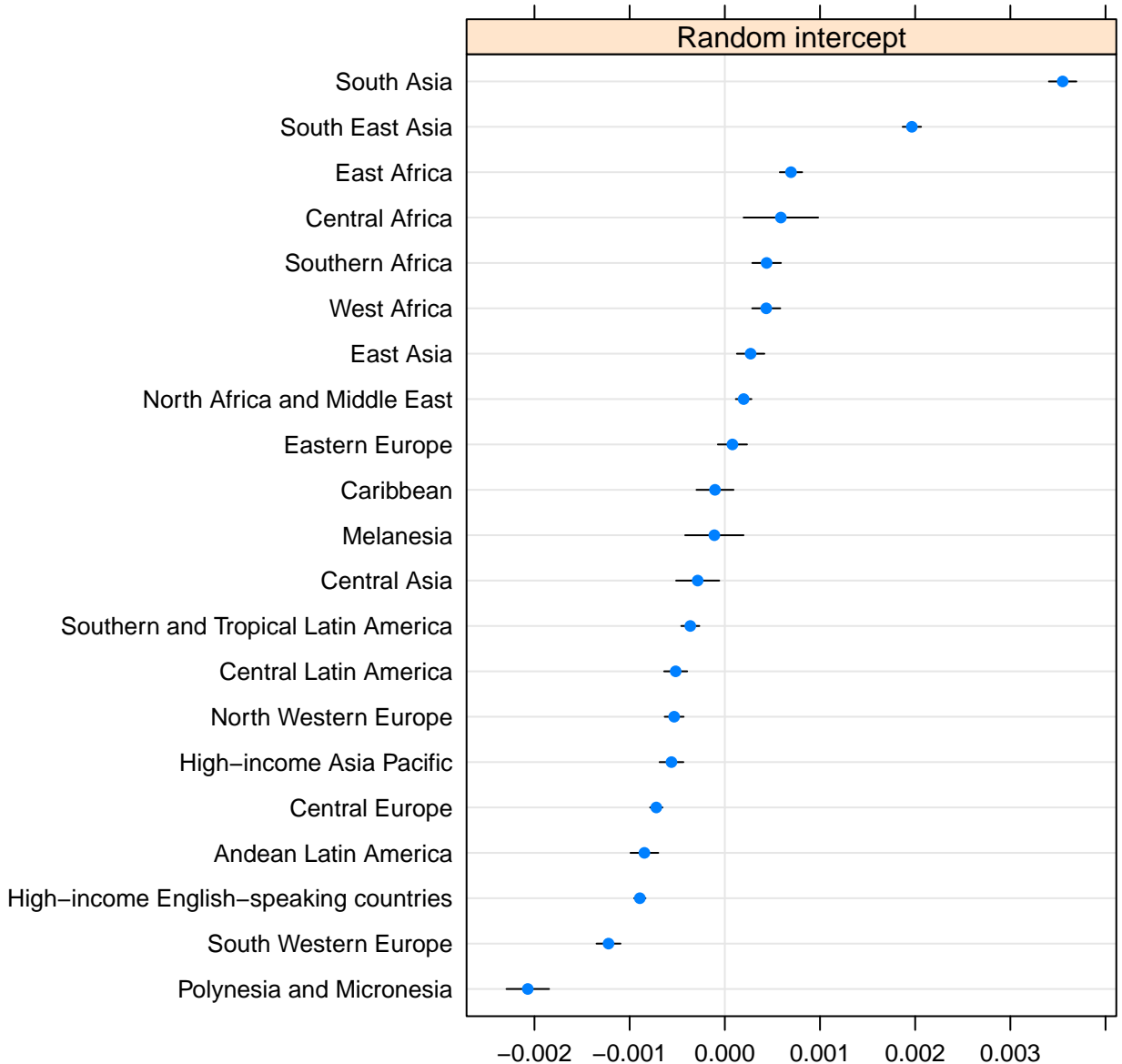
Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI ≥ 30 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.058 (0.057, 0.060)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²)	-0.0062 (-0.0065, -0.0060)
Mid-age of age group	-0.00071 (-0.00074, -0.00067)
Male sex	0.00069 (0.00043, 0.00095)
Study mid-year (per one more recent year since 1975)	-3.0e-05 (-3.4e-05, -2.6e-05)
Natural logarithm of per-capita gross domestic product	-0.00062 (-0.00069, -0.00055)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * mid-age of age group	0.00019 (0.00017, 0.00020)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * male sex	7.7e-05 (-4.2e-05, 0.00020)
Number of data points used to fit the model = 10,321	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.919.



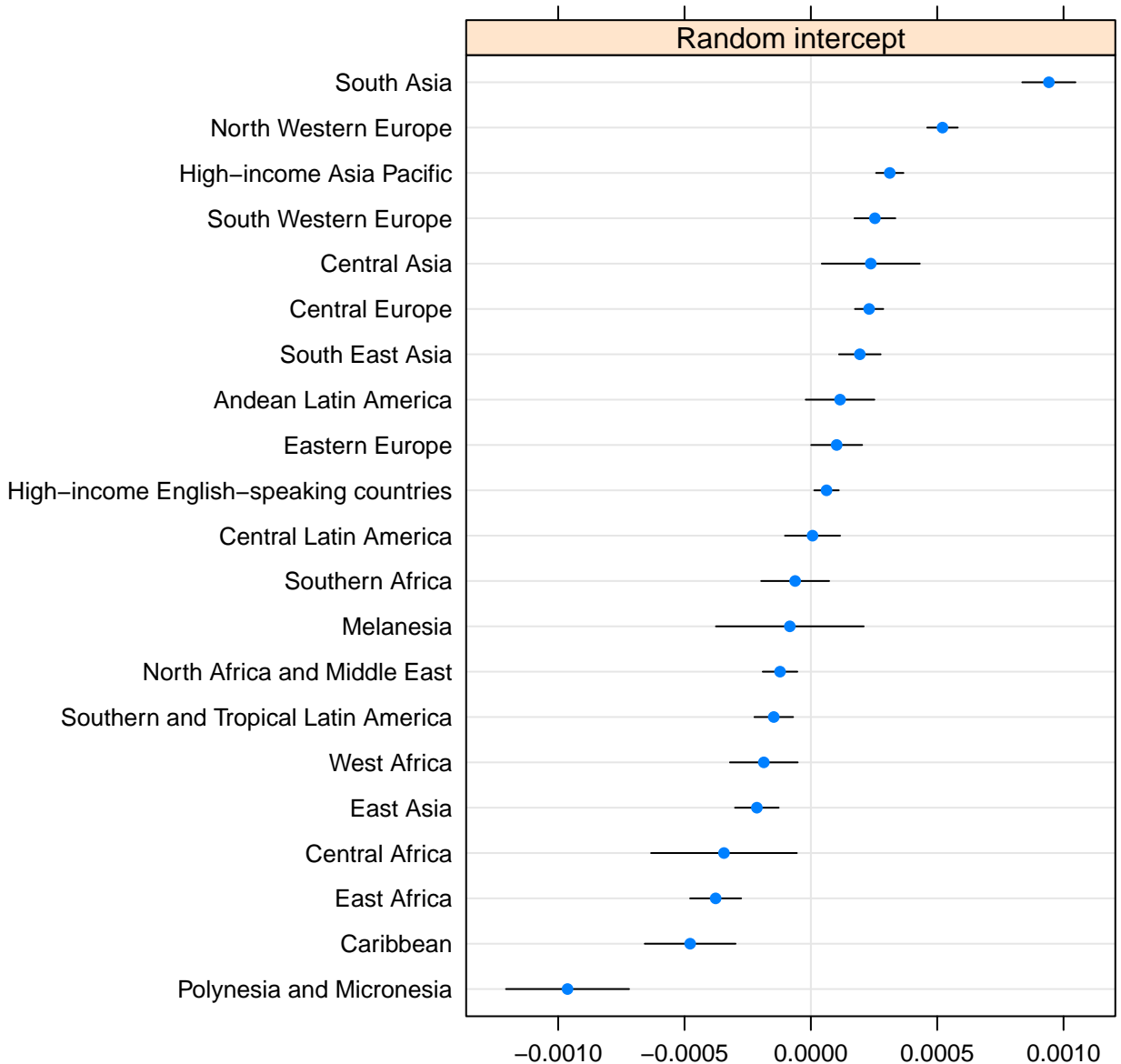
Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI ≥ 25 kg/m²) and prevalence (BMI ≥ 30 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.052 (0.051, 0.053)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²)	-0.011 (-0.012, -0.011)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²)	0.0028 (0.0025, 0.0032)
Mid-age of age group	-0.00070 (-0.00074, -0.00067)
Male sex	0.00012 (-9.9e-05, 0.00033)
Study mid-year (per one more recent year since 1975)	-3.7e-06 (-6.9e-06, -5.5e-07)
Natural logarithm of per-capita gross domestic product	-7.3e-05 (-0.00012, -2.0e-05)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * mid-age of age group	0.00041 (0.00038, 0.00044)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * mid-age of age group	-0.00023 (-0.00025, -0.00020)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * male sex	0.00027 (6.5e-05, 0.00047)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * male sex	-0.00029 (-0.00049, -8.9e-05)
Number of data points used to fit the model = 10,139	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.955.



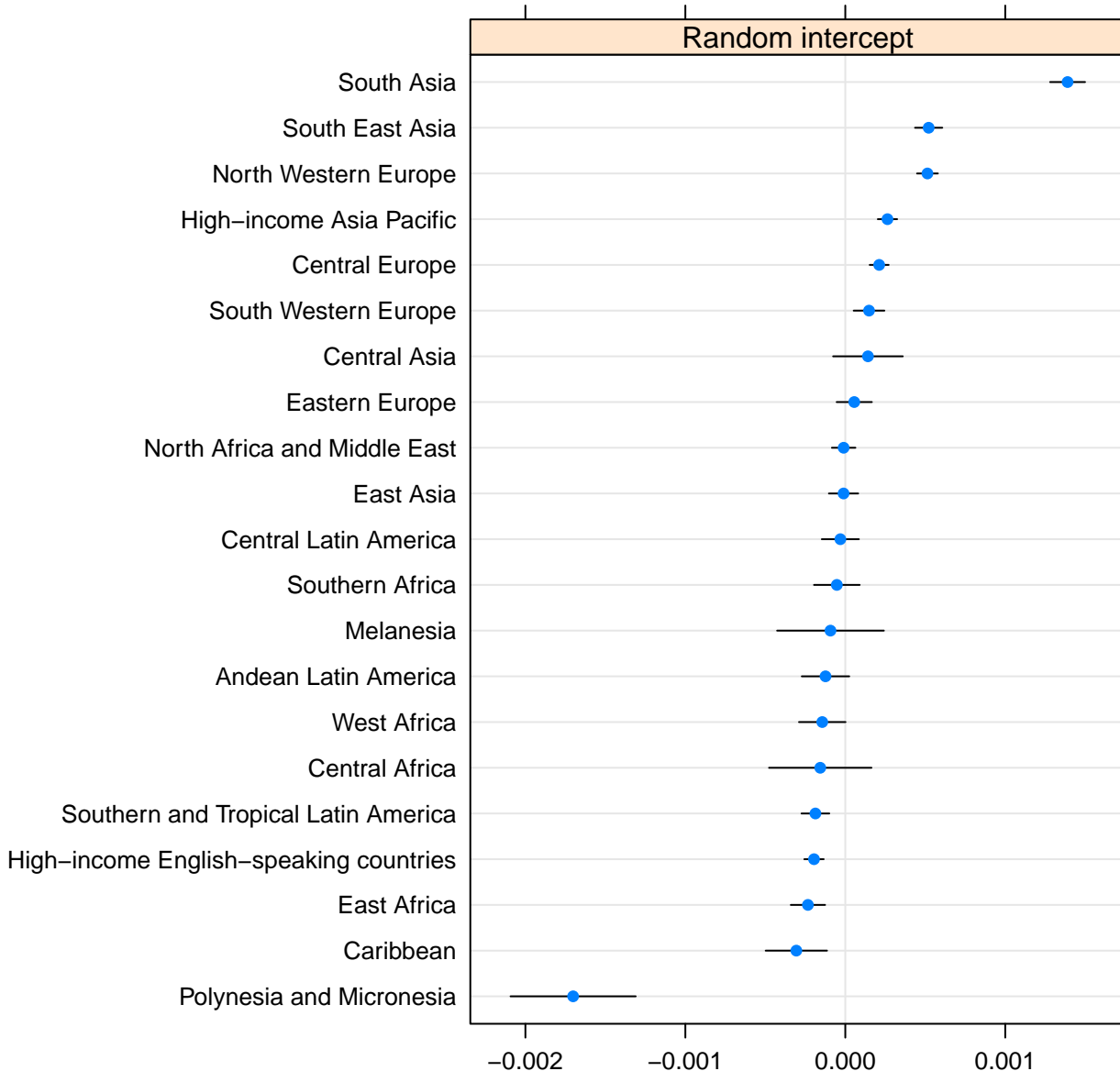
Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI > +1SD), prevalence (BMI > +2SD) and prevalence (BMI < -1SD)	
Variables	Coefficients (95% CI)
Intercept	0.074 (0.074, 0.075)
Probit-transformed prevalence (BMI > +1SD)	-0.0042 (-0.0045, -0.0038)
Probit-transformed prevalence (BMI > +2SD)	-0.0012 (-0.0015, -0.00085)
Probit-transformed prevalence (BMI < -1SD)	0.0031 (0.0029, 0.0033)
Mid-age of age group	-0.0017 (-0.0018, -0.0017)
Male sex	0.00041 (0.00020, 0.00063)
Study mid-year (per one more recent year since 1975)	2.2e-06 (-1.6e-08, 4.3e-06)
Natural logarithm of per-capita gross domestic product	-0.00034 (-0.00038, -0.00030)
Probit-transformed prevalence (BMI > +1SD) * mid-age of age group	-2.4e-05 (-5.0e-05, 2.0e-06)
Probit-transformed prevalence (BMI > +2SD) * mid-age of age group	3.2e-05 (1.1e-05, 5.3e-05)
Probit-transformed prevalence (BMI < -1SD) * mid-age of age group	-8.8e-05 (-0.00010, -7.3e-05)
Probit-transformed prevalence (BMI > +1SD) * male sex	0.0010 (0.00080, 0.0012)
Probit-transformed prevalence (BMI > +2SD) * male sex	-0.00080 (-0.00097, -0.00063)
Probit-transformed prevalence (BMI < -1SD) * male sex	0.00031 (0.00019, 0.00043)
Number of data points used to fit the model = 17,142	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.968.



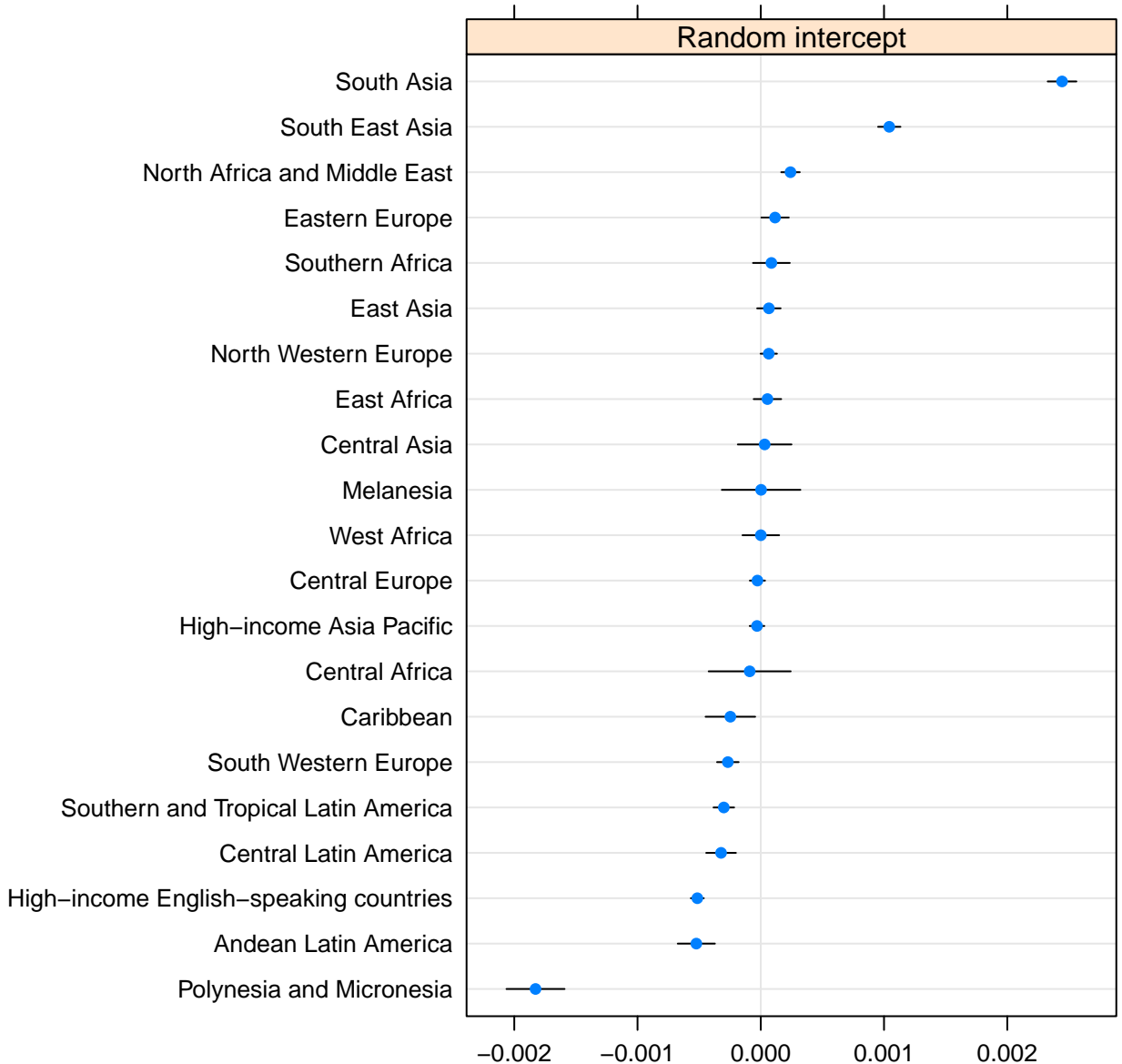
Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI > +1SD), prevalence (BMI > +2SD) and prevalence (BMI < -2SD)	
Variables	Coefficients (95% CI)
Intercept	0.074 (0.073, 0.074)
Probit-transformed prevalence (BMI > +1SD)	-0.0054 (-0.0058, -0.0050)
Probit-transformed prevalence (BMI > +2SD)	-0.00085 (-0.0012, -0.00049)
Probit-transformed prevalence (BMI < -2SD)	0.0020 (0.0018, 0.0022)
Mid-age of age group	-0.0018 (-0.0018, -0.0018)
Male sex	0.00092 (0.00065, 0.0012)
Study mid-year (per one more recent year since 1975)	3.4e-06 (9.8e-07, 5.8e-06)
Natural logarithm of per-capita gross domestic product	-0.00028 (-0.00032, -0.00023)
Probit-transformed prevalence (BMI > +1SD) * mid-age of age group	-1.2e-05 (-4.0e-05, 1.7e-05)
Probit-transformed prevalence (BMI > +2SD) * mid-age of age group	4.2e-05 (1.7e-05, 6.7e-05)
Probit-transformed prevalence (BMI < -2SD) * mid-age of age group	-8.3e-05 (-9.6e-05, -7.0e-05)
Probit-transformed prevalence (BMI > +1SD * male sex	0.0013 (0.0011, 0.0016)
Probit-transformed prevalence (BMI > +2SD) * male sex	-0.0011 (-0.0013, -0.00090)
Probit-transformed prevalence (BMI < -2SD) * male sex	0.00053 (0.00043, 0.00064)
Number of data points used to fit the model = 14,847	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.966.



Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI > +1SD) and prevalence (BMI > +2SD)	
Variables	Coefficients (95% CI)
Intercept	0.069 (0.068, 0.069)
Probit-transformed prevalence (BMI > +1SD)	-0.0064 (-0.0067, -0.0060)
Probit-transformed prevalence (BMI > +2SD)	-0.0010 (-0.0013, -0.00068)
Mid-age of age group	-0.0016 (-0.0016, -0.0015)
Male sex	0.00011 (-5.6e-05, 0.00028)
Study mid-year (per one more recent year since 1975)	4.3e-06 (1.9e-06, 6.7e-06)
Natural logarithm of per-capita gross domestic product	-0.00029 (-0.00033, -0.00024)
Probit-transformed prevalence (BMI > +1SD) * mid-age of age group	6.5e-05 (4.2e-05, 8.8e-05)
Probit-transformed prevalence (BMI > +2SD) * mid-age of age group	3.9e-05 (1.8e-05, 6.1e-05)
Probit-transformed prevalence (BMI > +1SD) * male sex	0.00059 (0.00040, 0.00078)
Probit-transformed prevalence (BMI > +2SD) * male sex	-0.00071 (-0.00089, -0.00053)
Number of data points used to fit the model = 17,335	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.962.



Dependent variable: Inverse mean BMI	
Age range: 5-19 years	
Independent variable: Prevalence (BMI > +1SD) and prevalence (BMI < -2SD)	
Variables	Coefficients (95% CI)
Intercept	0.075 (0.074, 0.075)
Probit-transformed prevalence (BMI > +1SD)	-0.0063 (-0.0064, -0.0061)
Probit-transformed prevalence (BMI < -2SD)	0.0021 (0.0019, 0.0022)
Mid-age of age group	-0.0018 (-0.0018, -0.0017)
Male sex	0.0015 (0.0012, 0.0017)
Study mid-year (per one more recent year since 1975)	-2.4e-06 (-4.8e-06, 4.0e-09)
Natural logarithm of per-capita gross domestic product	-0.00032 (-0.00036, -0.00028)
Probit-transformed prevalence (BMI > +1SD) * mid-age of age group	5.7e-05 (4.4e-05, 7.0e-05)
Probit-transformed prevalence (BMI < -2SD) * mid-age of age group	-7.1e-05 (-8.4e-05, -5.9e-05)
Probit-transformed prevalence (BMI > +1SD) * male sex	0.00029 (0.00019, 0.00039)
Probit-transformed prevalence (BMI < -2SD) * male sex	0.00044 (0.00034, 0.00055)
Number of data points used to fit the model = 15,923	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.962.

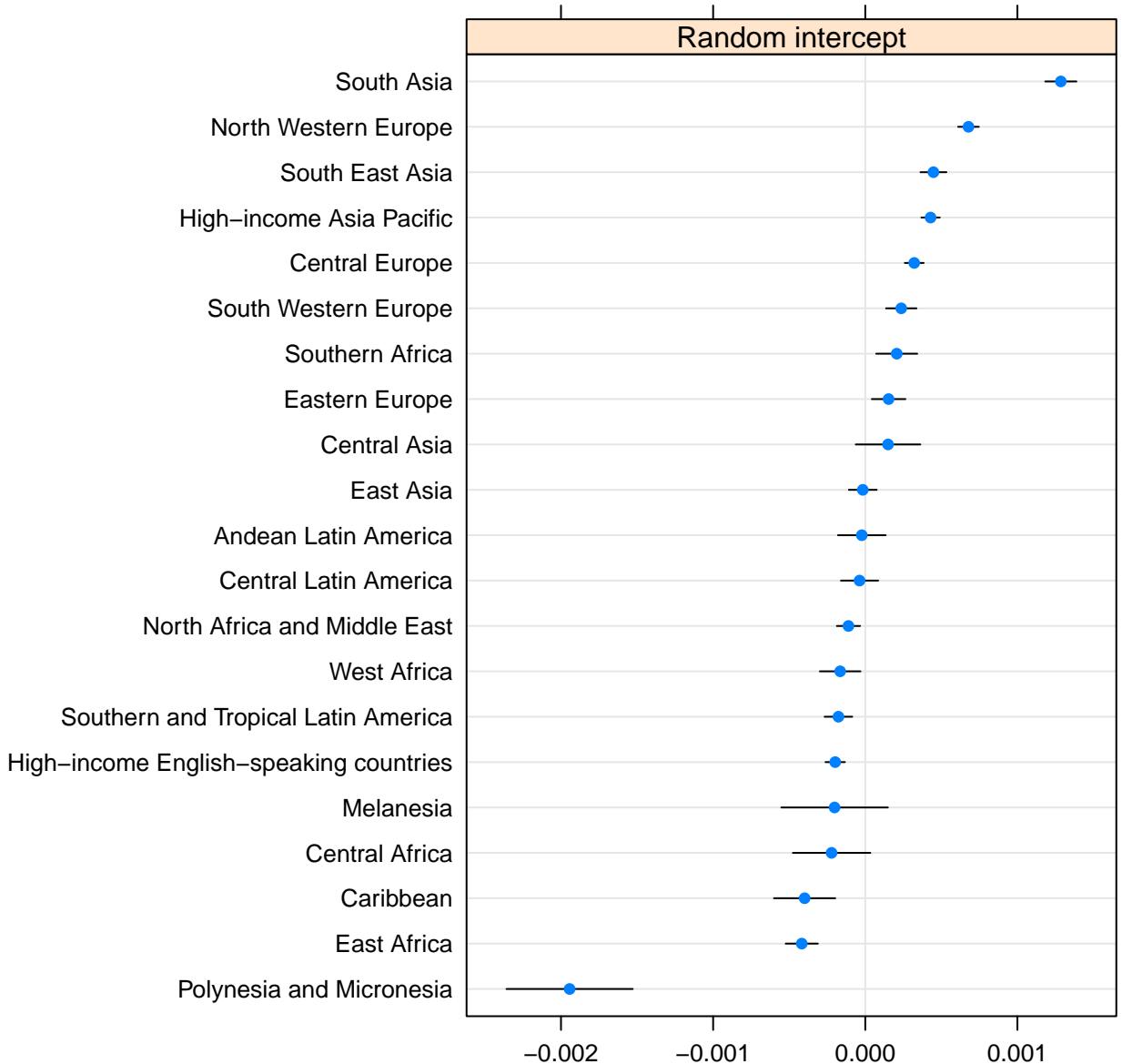


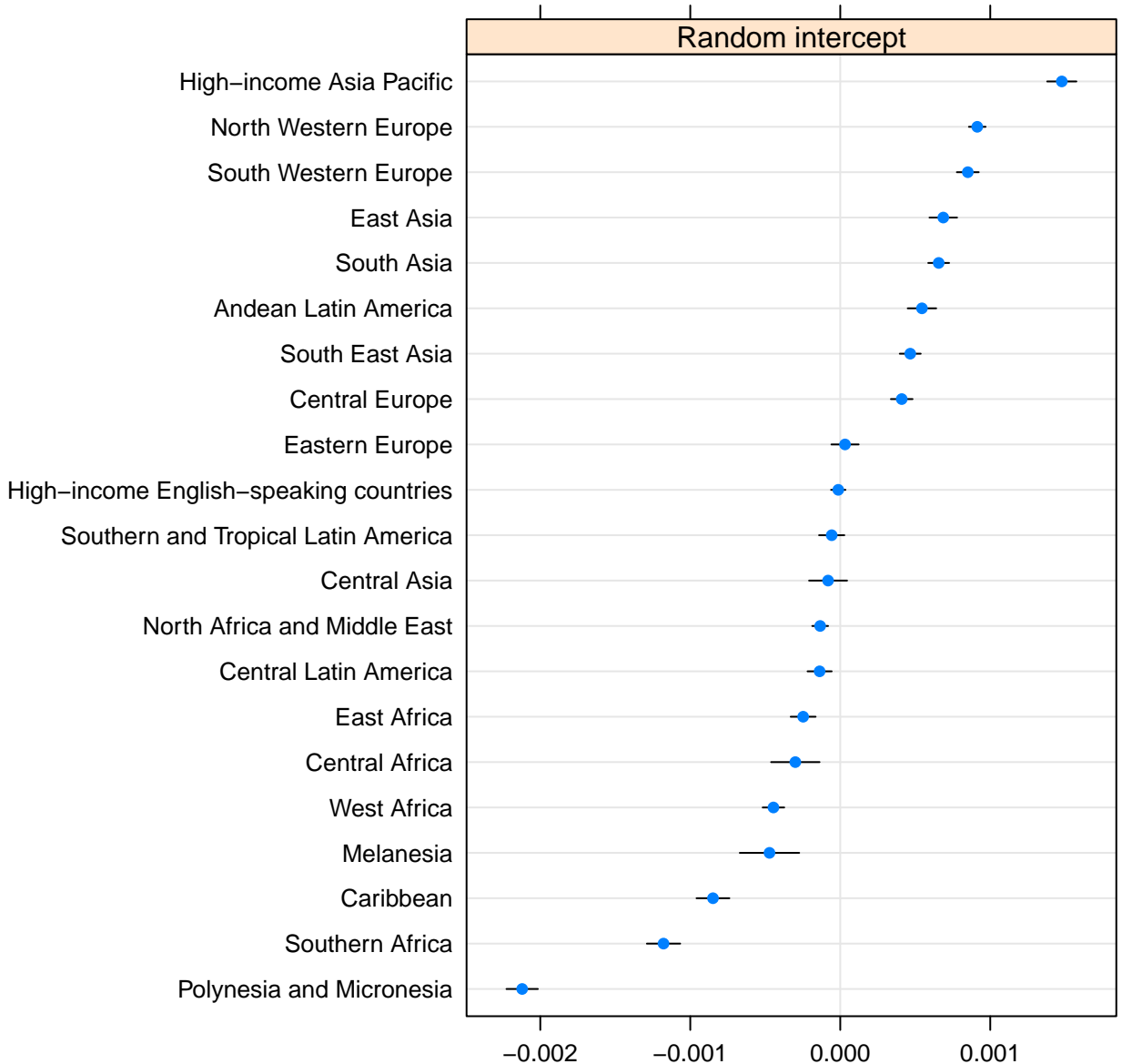
Table 2: Model specifications and regression coefficients to estimate mean BMI for adults (aged 18 years and older).

The dependent variable in all regressions was the inverse of mean BMI, fitted using a linear (mixed) model. Random intercepts for regions in regression are presented after the table of coefficients.

* denotes statistical interaction. CI: confidence interval.

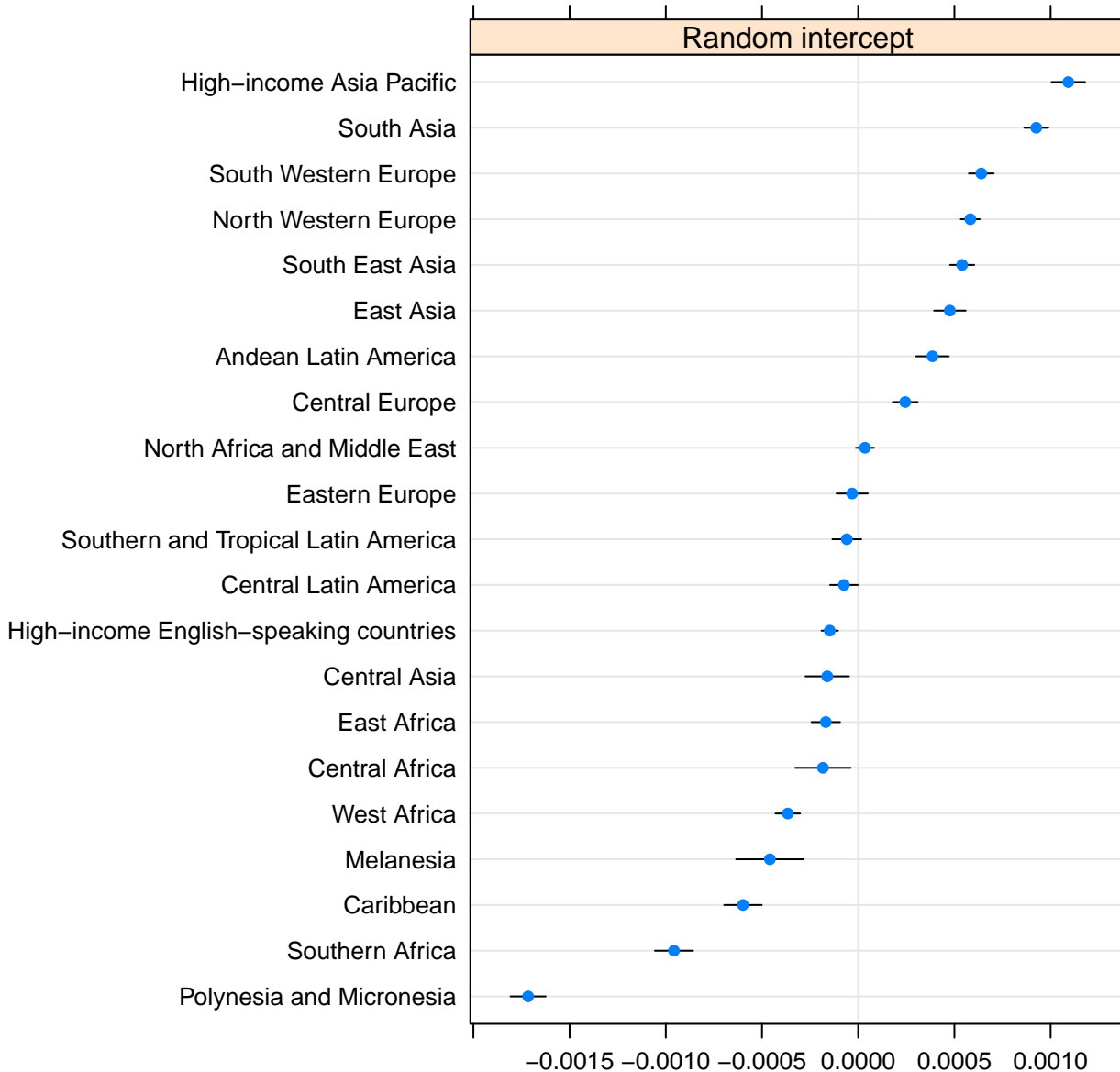
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 22 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.048 (0.047, 0.048)
Probit-transformed prevalence (BMI ≥ 22 kg/m ²)	-0.0062 (-0.0063, -0.0061)
Mid-age of age group	1.5e-06 (2.6e-07, 2.7e-06)
Male sex	0.00051 (0.00047, 0.00056)
Study mid-year (per one more recent year since 1975)	-2.3e-05 (-2.5e-05, -2.1e-05)
Natural logarithm of per-capita gross domestic product	-0.00038 (-0.00041, -0.00035)
Probit-transformed prevalence (BMI ≥ 22 kg/m ²) * mid-age of age group	1.8e-06 (5.7e-07, 3.1e-06)
Probit-transformed prevalence (BMI ≥ 22 kg/m ²) * male sex	0.00091 (0.00086, 0.00095)
Number of data points used to fit the model = 15,027	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.948.



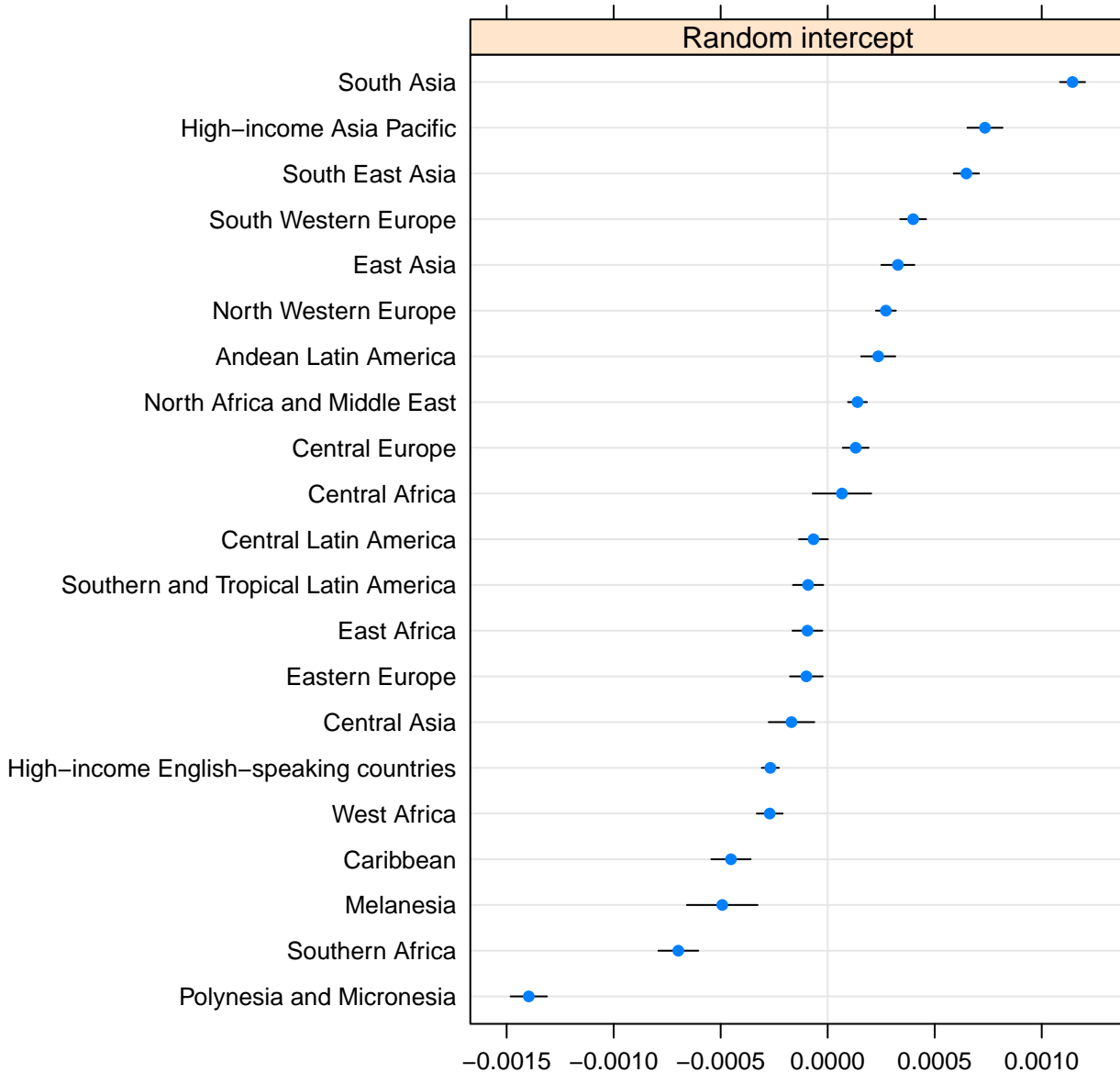
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 23 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.045 (0.044, 0.045)
Probit-transformed prevalence (BMI ≥ 23 kg/m ²)	-0.0061 (-0.0062, -0.0061)
Mid-age of age group	8.5e-06 (7.5e-06, 9.5e-06)
Male sex	0.00062 (0.00058, 0.00065)
Study mid-year (per one more recent year since 1975)	-1.8e-05 (-1.9e-05, -1.6e-05)
Natural logarithm of per-capita gross domestic product	-0.00027 (-0.00030, -0.00025)
Probit-transformed prevalence (BMI ≥ 23 kg/m ²) * mid-age of age group	-6.2e-07 (-1.7e-06, 5.0e-07)
Probit-transformed prevalence (BMI ≥ 23 kg/m ²) * male sex	0.00087 (0.00083, 0.00091)
Number of data points used to fit the model = 15,076	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.958.



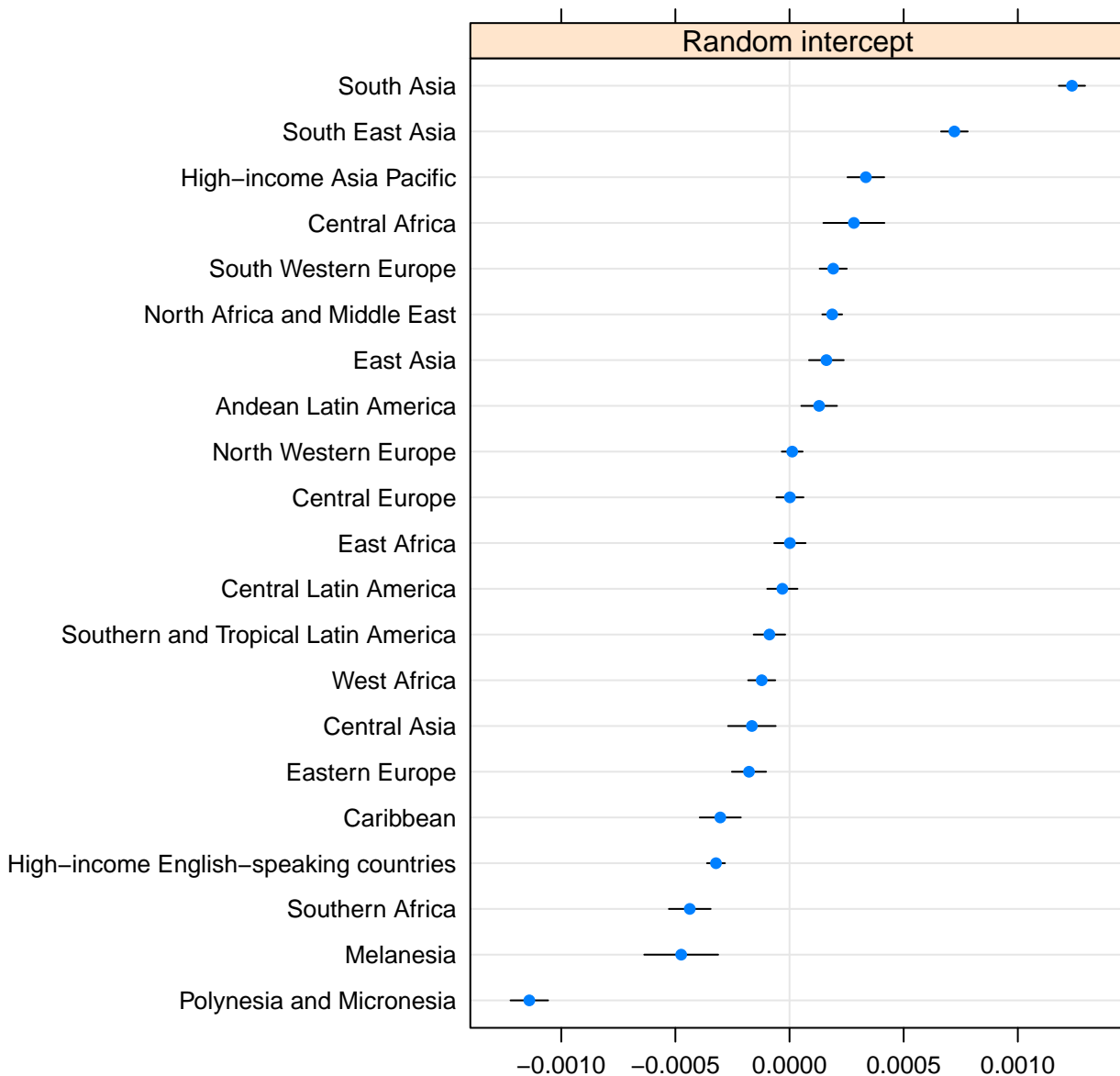
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 24 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.042 (0.042, 0.042)
Probit-transformed prevalence (BMI ≥ 24 kg/m ²)	-0.0059 (-0.0060, -0.0059)
Mid-age of age group	1.1e-05 (1.0e-05, 1.2e-05)
Male sex	0.00064 (0.00061, 0.00067)
Study mid-year (per one more recent year since 1975)	-1.3e-05 (-1.5e-05, -1.2e-05)
Natural logarithm of per-capita gross domestic product	-0.00019 (-0.00022, -0.00016)
Probit-transformed prevalence (BMI ≥ 24 kg/m ²) * mid-age of age group	-4.8e-06 (-5.8e-06, -3.7e-06)
Probit-transformed prevalence (BMI ≥ 24 kg/m ²) * male sex	0.00080 (0.00076, 0.00084)
Number of data points used to fit the model = 15,085	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.962.



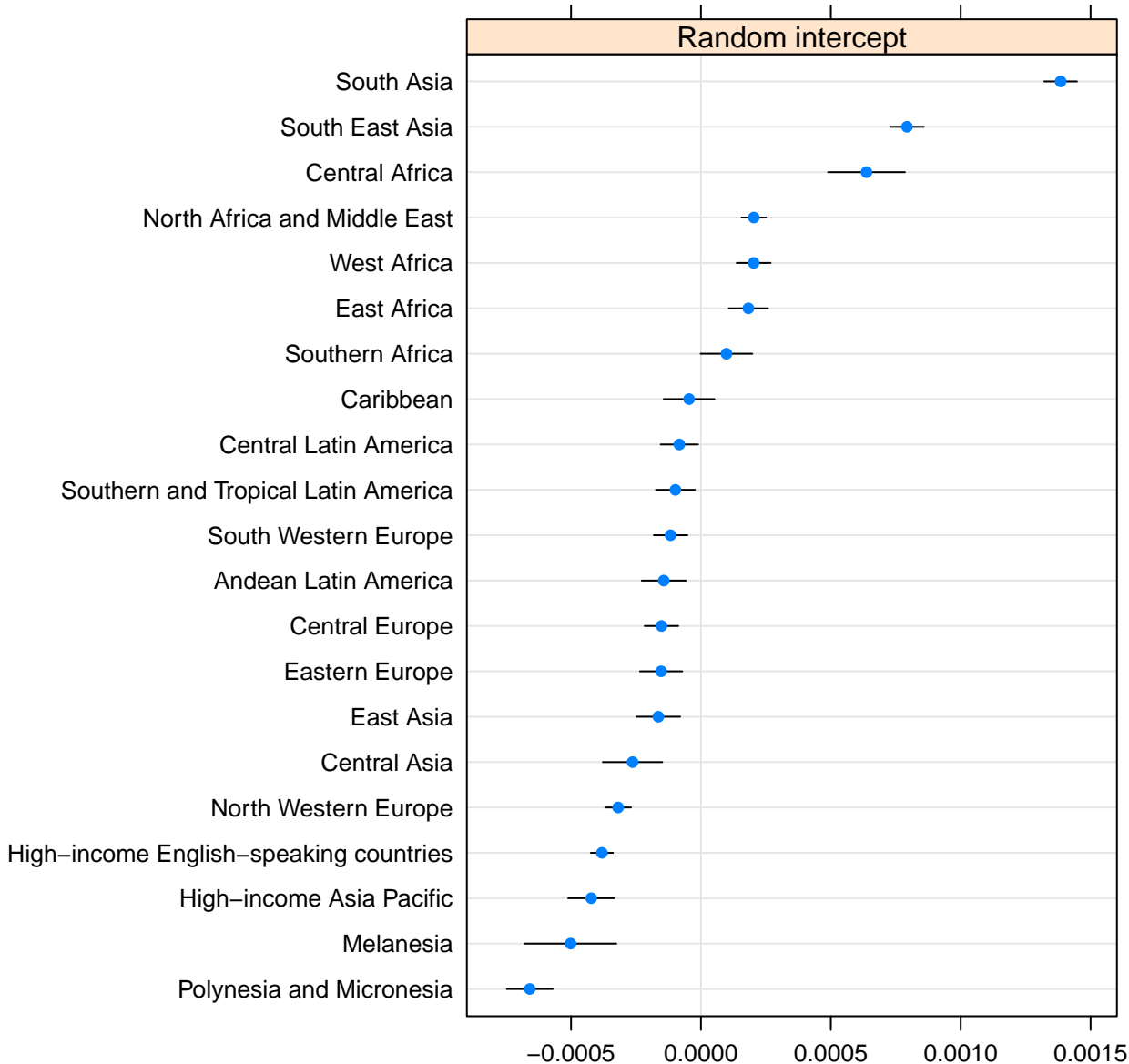
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 25 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.040 (0.040, 0.040)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²)	-0.0058 (-0.0059, -0.0058)
Mid-age of age group	9.8e-06 (8.9e-06, 1.1e-05)
Male sex	0.00059 (0.00056, 0.00062)
Study mid-year (per one more recent year since 1975)	-8.5e-06 (-1.0e-05, -6.8e-06)
Natural logarithm of per-capita gross domestic product	-0.00012 (-0.00014, -0.000091)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * mid-age of age group	-7.5e-06 (-8.5e-06, -6.4e-06)
Probit-transformed prevalence (BMI ≥ 25 kg/m ²) * male sex	0.00076 (0.00072, 0.00079)
Number of data points used to fit the model = 15,061	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.963.



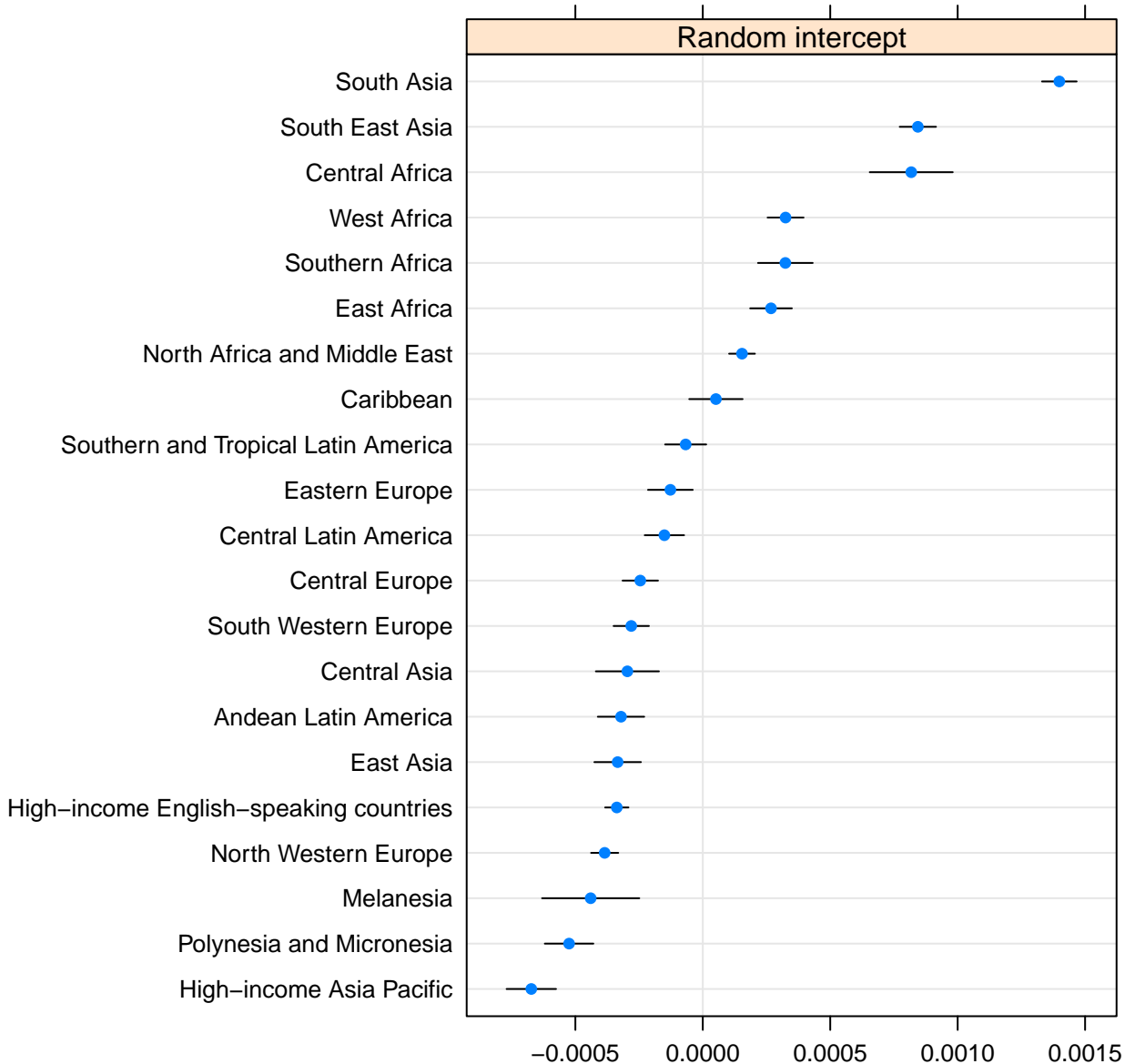
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 27 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.037 (0.037, 0.037)
Probit-transformed prevalence (BMI ≥ 27 kg/m ²)	-0.0058 (-0.0058, -0.0057)
Mid-age of age group	1.2e-07 (-1.1e-06, 1.4e-06)
Male sex	0.00033 (0.00029, 0.00037)
Study mid-year (per one more recent year since 1975)	2.5e-06 (7.1e-07, 4.4e-06)
Natural logarithm of per-capita gross domestic product	-4.9e-05 (-7.8e-05, -2.0e-05)
Probit-transformed prevalence (BMI ≥ 27 kg/m ²) * mid-age of age group	-9.2e-06 (-1.0e-05, -8.0e-06)
Probit-transformed prevalence (BMI ≥ 27 kg/m ²) * male sex	0.00059 (0.00055, 0.00063)
Number of data points used to fit the model = 14,901	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.955.



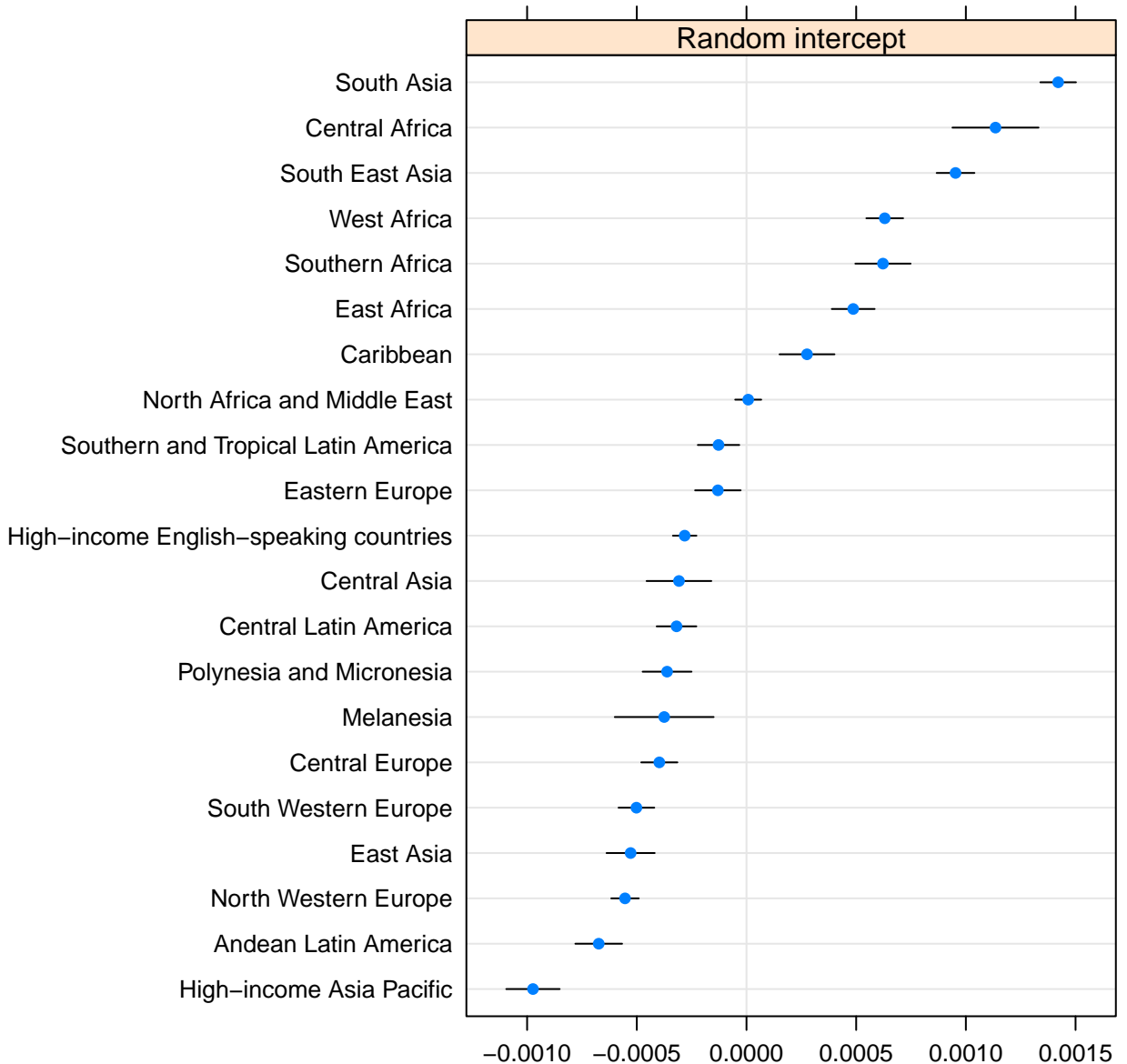
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI \geq28 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.036 (0.035, 0.036)
Probit-transformed prevalence (BMI \geq 28 kg/m ²)	-0.0057 (-0.0058, -0.0057)
Mid-age of age group	-6.6e-06 (-8.1e-06, -5.0e-06)
Male sex	0.00012 (0.00007, 0.00017)
Study mid-year (per one more recent year since 1975)	6.9e-06 (4.9e-06, 8.9e-06)
Natural logarithm of per-capita gross domestic product	-3.2e-05 (-6.3e-05, -2.3e-07)
Probit-transformed prevalence (BMI \geq 28 kg/m ²) * mid-age of age group	-9.7e-06 (-1.1e-05, -8.4e-06)
Probit-transformed prevalence (BMI \geq 28 kg/m ²) * male sex	0.00048 (0.00043, 0.00053)
Number of data points used to fit the model = 14,769	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.947.



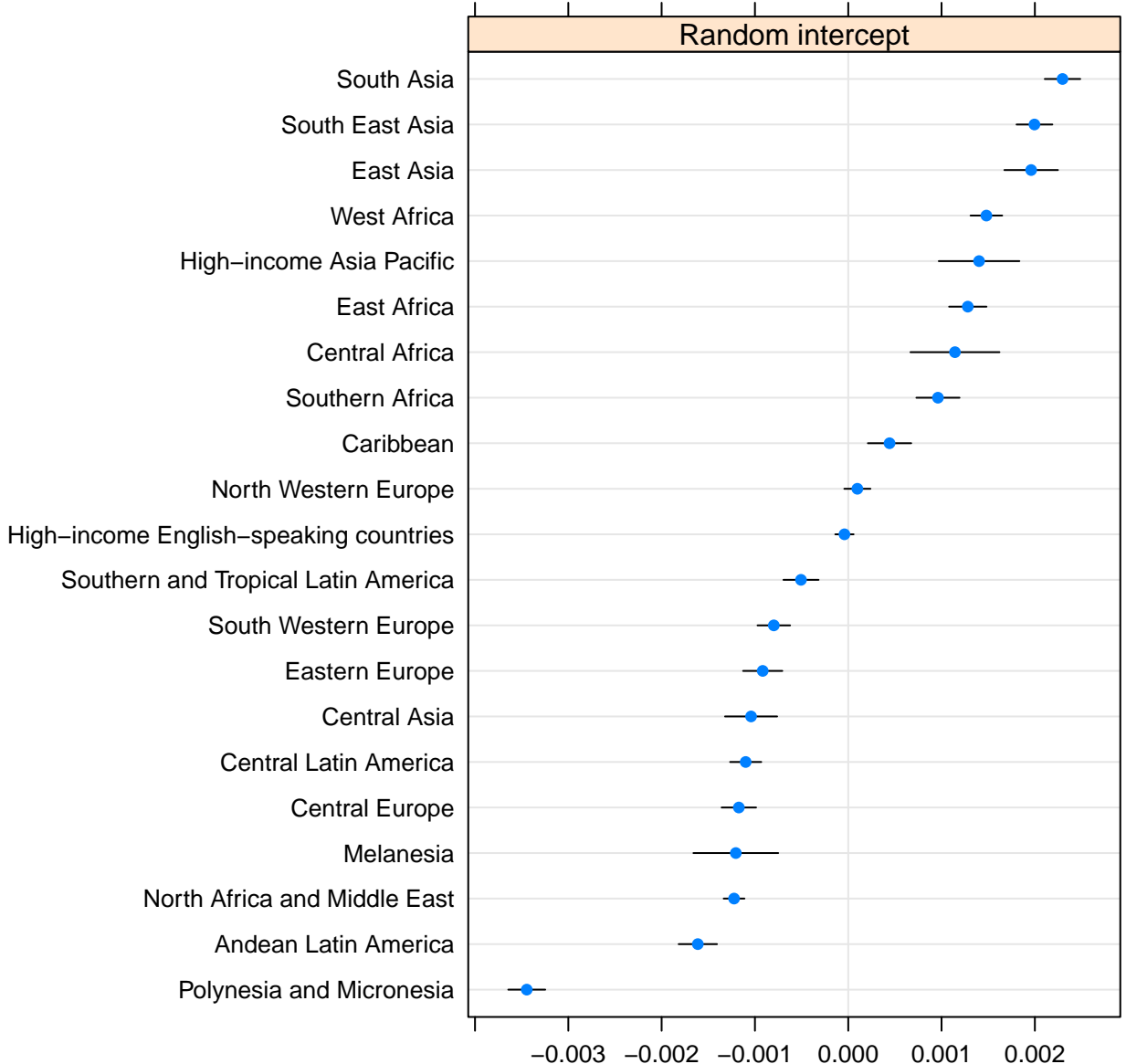
Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 30 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.034 (0.034, 0.035)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²)	-0.0058 (-0.0059, -0.0057)
Mid-age of age group	-2.2e-05 (-2.4e-05, -1.9e-05)
Male sex	-0.00022 (-0.00030, -0.00014)
Study mid-year (per one more recent year since 1975)	1.1e-05 (8.5e-06, 1.3e-05)
Natural logarithm of per-capita gross domestic product	-2.2e-05 (-6.0e-05, 1.5e-05)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * mid-age of age group	-9.3e-06 (-1.1e-05, -7.6e-06)
Probit-transformed prevalence (BMI ≥ 30 kg/m ²) * male sex	0.00041 (0.00035, 0.00047)
Number of data points used to fit the model = 14,400	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.925.



Dependent variable: Inverse mean BMI	
Age range: 18 years and older	
Independent variable: Prevalence (BMI ≥ 40 kg/m²)	
Variables	Coefficients (95% CI)
Intercept	0.036 (0.034, 0.037)
Probit-transformed prevalence (BMI ≥ 40 kg/m ²)	-0.0051 (-0.0053, -0.0049)
Mid-age of age group	-5.2e-05 (6.2e-05, -4.1e-05)
Male sex	0.0020 (0.0017, 0.0024)
Study mid-year (per one more recent year since 1975)	-5.1e-06 (-1.0e-05, -1.6e-07)
Natural logarithm of per-capita gross domestic product	-0.00051 (-0.00059, -0.00044)
Probit-transformed prevalence (BMI ≥ 40 kg/m ²) * mid-age of age group	5.4e-06 (8.0e-07, 9.9e-06)
Probit-transformed prevalence (BMI ≥ 40 kg/m ²) * male sex	0.0011 (0.0008, 0.0012)
Number of data points used to fit the model = 9,376	

Traditional R^2 is not clearly defined for mixed-effect models. The conditional R^2 for the model, which describes the proportion of variance explained by both fixed and random factors,⁴ was 0.730.



References

- 1 NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet* 2016; 387(10026): 1377-1396.
- 2 Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978; 6(2): 461-464.
- 3 NCD Risk Factor Collaboration Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet* 2021; 398, 957-980.
- 4 Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 2013; 4(2): 133-142.