Acylcarnitines and prediction of renal function decline in type 2 diabetes

Vincenzo Trischitta, Andrea Fontana, Hetal S. Shah, Mario Mastroianno, Cornelia Prehn, Salvatore De Cosmo, Jerzy Adamski, Massimiliano Copetti, Alessandro Doria, Claudia Menzaghi

Online Supplemental Material

Online Supplemental Methods

eGFR (ml/min/and eGFR slope determination

In the aGMS, serum creatinine was measured using the modified kinetic Jaffè reaction (Hitachi 737 Autoanalyzer), calibrated to be traceable to an isotope dilution mass spectrometry (IDMS). GFR was then estimated for each patient using the 2009 Chronic Kidney Disease Epidemiology (CKD-EPI) Collaboration formula derived (1) from serum creatinine values at baseline and during follow-up and was reported as mL/min per 1.73 m^2. We traced the decline of eGFR in patients having at least three eGFR records at intervals of 4 months with at least 1.5 years of follow-up. Since multiple eGFR measurements were collected for each participant, linear mixed-effects models were used to estimate the average of annual change in eGFR across the population. Follow-up time was included as a fixed-effect covariate, while both the intercept and slope for each participant were modeled as random effects.

In the JKS, serum creatinine concentrations at baseline and during follow-up were measured at the Advanced Research and Diagnostic Laboratory (ARDL), University of Minnesota, MN using Roche assays. The 2009 CKD-EPI Collaboration equation (1) was used to estimate creatinine-derived (SCre-eGFR) GFR (2).

The average rate of eGFR decline (eGFR slope) was estimated in the entire sample by linear mixed effect regression with eGFR as the dependent variable, time as fixed effect and participant ID as random effect.

Acylcarnitine quantification and normalization

In the aGMS, forty acylcarnitines (Cx:y), were quantified using baseline fasting serum samples and the Absolute*IDQ*™ p180 Kit (BIOCRATES Life Sciences AG, Innsbruck, Austria), as previously described (3-5). For a full list of all quality-controlled metabolites, see Supplemental Table S1. Mass spectrometric analyses were carried out on an API 4000 triple quadrupole system (SCIEX Deutschland GmbH, Darmstadt, Germany) equipped with a 1260 Series HPLC (Agilent Technologies Deutschland GmbH, Böblingen, Germany) and an HTC-xc PAL auto sampler (CTC Analytics, Zwingen, Switzerland) controlled by the software Analyst 1.6.2. For the LC-part, compounds were identified and quantified based on scheduled multiple reaction monitoring measurements (sMRM), for the FIA-part on MRM. Data evaluation for

quantification of metabolite concentrations and quality assessment was performed with the software MultiQuant 3.0.1 (Sciex) and the MetIDQ™ software package. Metabolite concentrations were calculated using internal standards and reported in μM.

In the JHS, global biochemical profiles were determined by Metabolon as previously described (6). A total of 903 metabolic compounds were identified. Of these, 36 were acylcarnitines. Acylcarnitine values were normalized in terms of raw area counts and then volume normalized. Each acylcarnitine was then rescaled to set the median equal to 1, and missing values were imputed with the minimum. Volume normalized, median-scaled acylcarnitines were transformed into ranked Z-scored for subsequent analyses.

Internal Validation

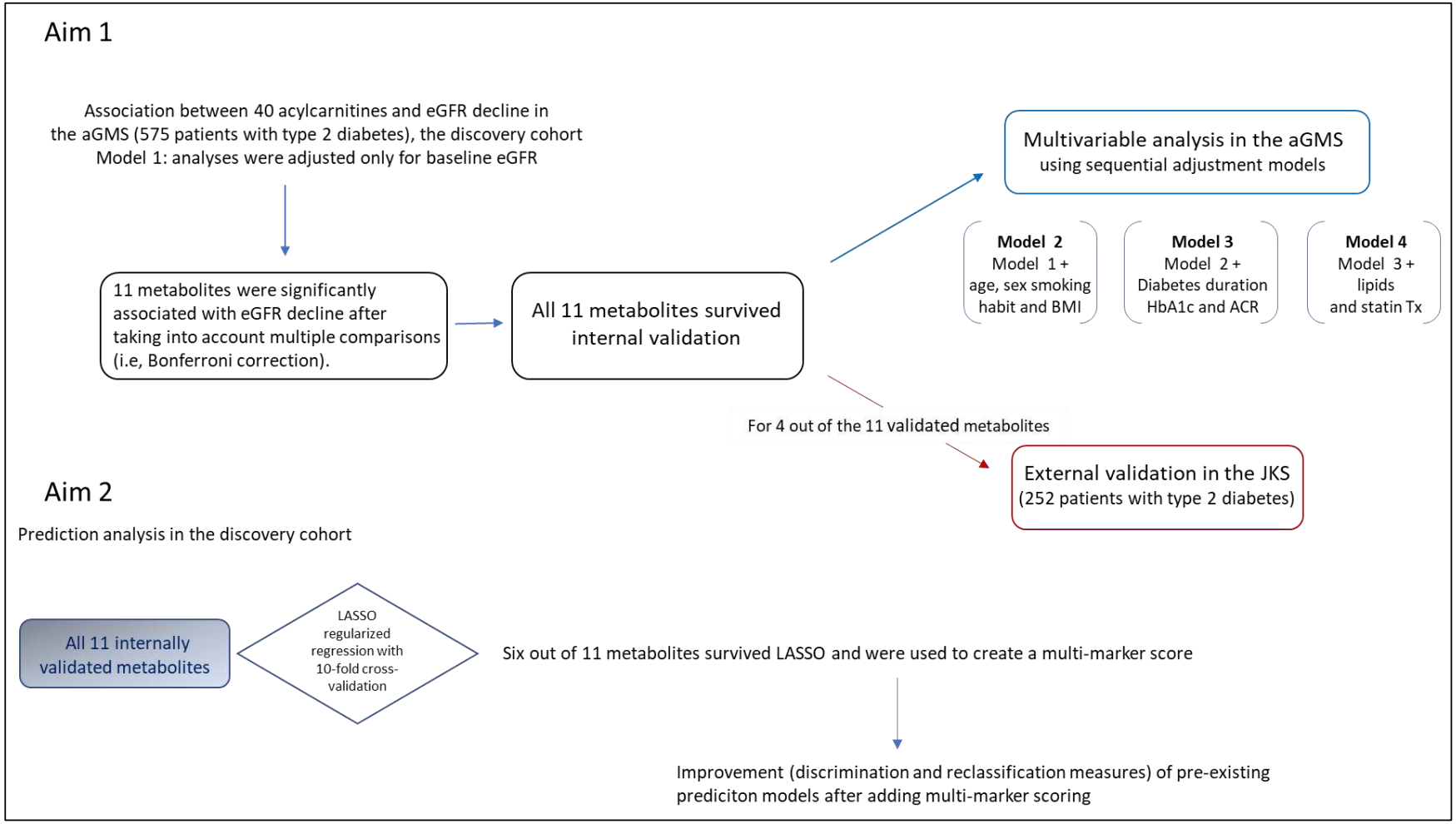
Metabolites were internally validated by including each of them in a linear regression model, incorporating a repeated 3-fold cross-validation procedure with 10,000 iterations (7). This process generated a total of 30,000 random validation sets (i.e., 10,000 iterations × 3 folds). In each iteration, each fold was used as a validation set while the other two folds served as the training set. Each model included both baseline eGFR and the metabolite as fixed covariates.

The performance of a cross-validated model was assessed by the Mean Squared Error (MSE) and its derived Coefficient of Determination (R^2). The distribution of MSE and R^2 across the validation sets was summarized by reporting the mean and median values. To compare the performance of the model estimated in the original (full) sample and the mean performance of the models obtained in the validation sets, the Standardized Mean Difference (SMD) was calculated using the observed standard deviation of the distribution of the cross-validated measures as the reference (denominator). An SMD < 0.50 was considered to indicate a small effect size (8). In addition, the beta distribution from the training sets was shown and the difference between each of these coefficients and the one obtained from the original model, performed on the full sample, was calculated (bias).

Supplemental Table S1: List of acylcarnitine measured with the Absolute*IDQ*® p180 Kit GAC, Helmholtz Zentrum, München.

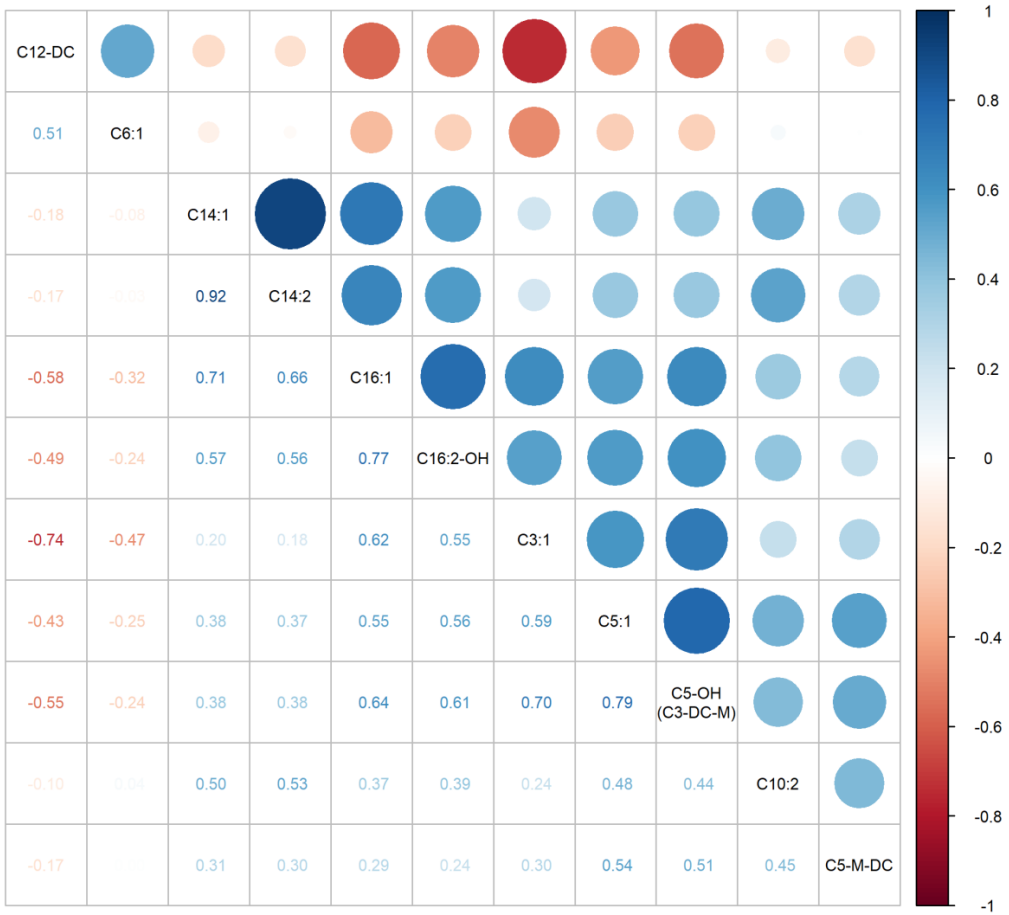
Acylcarnitine (n=40)	
C0	Carnitine
C2	Acetylcarnitine
C3	Propionylcarnitine
C3:1	Propenoylcarnitine
C3-OH	Hydroxypropionylcarnitine
C4	Butyrylcarnitine
C4:1	Butenoylcarnitine
C4-OH (C3-DC)	Hydroxybutyrylcarnitine
C5	Valerylcarnitine
C5:1	Tiglylcarnitine
C5:1-DC	Glutaconylcarnitine
C5-DC (C6-OH)	Glutarylcarnitine (Hydroxyhexanoylcarnitine)
C5-M-DC	Methylglutarylcarnitine
C5-OH (C3-DC-M)	Hydroxyvalerylcarnitine (Methylmalonylcarnitine)
C6 (C4:1-DC)	Hexanoylcarnitine (Fumarylcarnitine)
C6:1	Hexenoylcarnitine
C7-DC	Pimelylcarnitine
C8	Octanoylcarnitine
C9	Nonanoylcarnitine
C10	Decanoylcarnitine
C10:1	Decenoylcarnitine
C10:2	Decadienylcarnitine
C12	Dodecanoylcarnitine
C12:1	Dodecenoylcarnitine
C12-DC	Dodecanedioylcarnitine
C14	Tetradecanoylcarnitine
C14:1	Tetradecenoylcarnitine
C14:1-OH	Hydroxytetradecenoylcarnitine
C14:2	Tetradecadienylcarnitine
C14:2-OH	Hydroxytetradecadienylcarnitine
C16	Hexadecanoylcarnitine
C16:1	Hexadecenoylcarnitine
C16:1-OH	Hydroxyhexadecenoylcarnitine
C16:2	Hexadecadienylcarnitine
C16:2-OH	Hydroxyhexadecadienylcarnitine
C16-OH	Hydroxyhexadecanoylcarnitine
C18	Octadecanoylcarnitine
C18:1	Octadecenoylcarnitine
C18:1-OH	Hydroxyoctadecenoylcarnitine
C18:2	Octadecadienylcarnitine

Supplemental Figure S1



Study design. aGMS = aggregate Gargano Mortality Study, JKS = Joslin Diabetes Center, eGFR = estimates Glomerular Filtration Rate, BMI = Body Mass Index, HbA1c = glycated hemoglobin, LASSO = Least Absolute Shrinkage and Selection Operator.

Supplemental Figure S2. Correlation plot showing Pearson correlation coefficients between the 11 associated-acylcarnitine

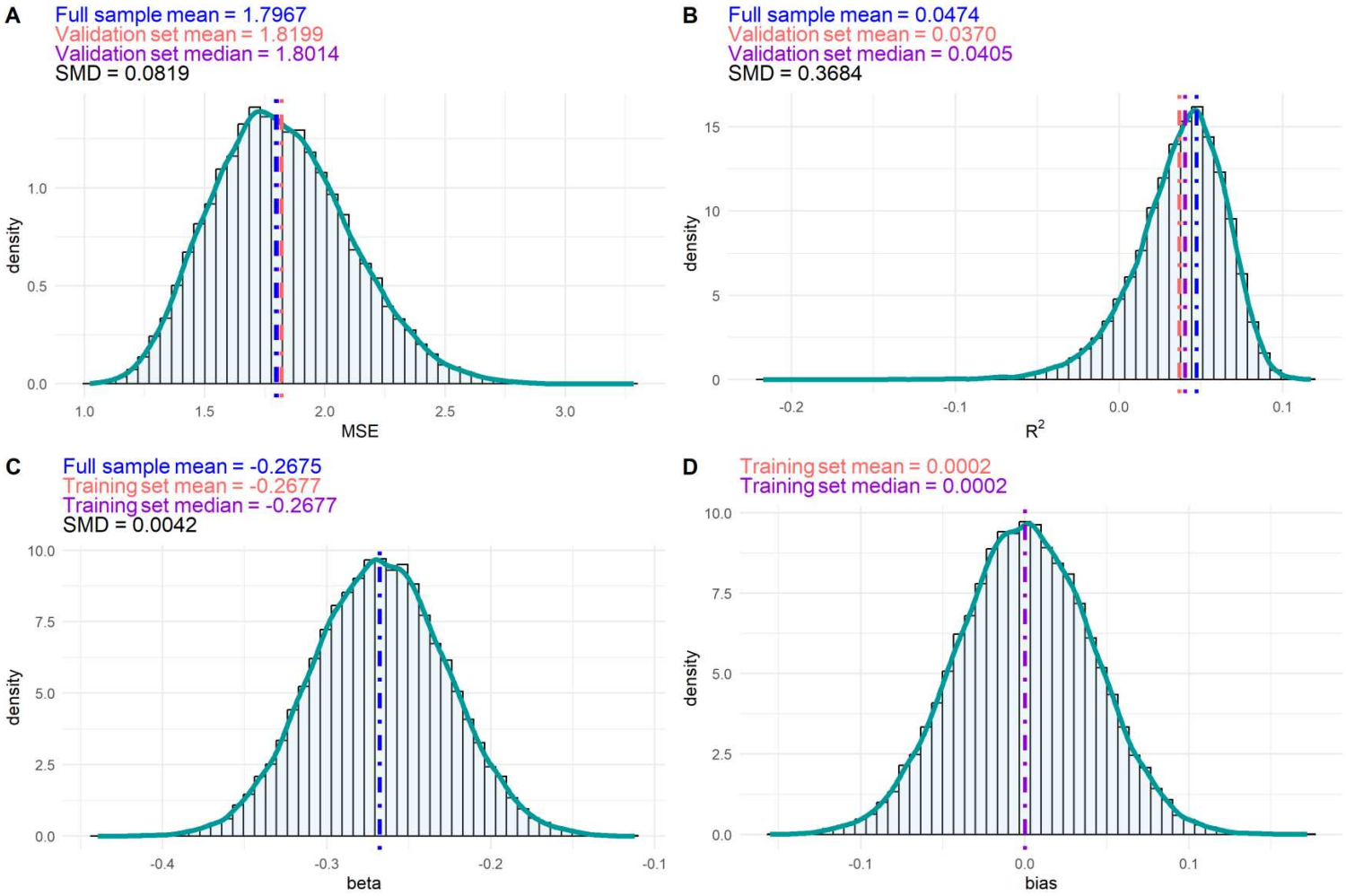


Correlation plot of active variables showing Pearson correlation coefficients, with colours ranging from dark red to dark blue indicating the maximum negative and positive correlations, respectively

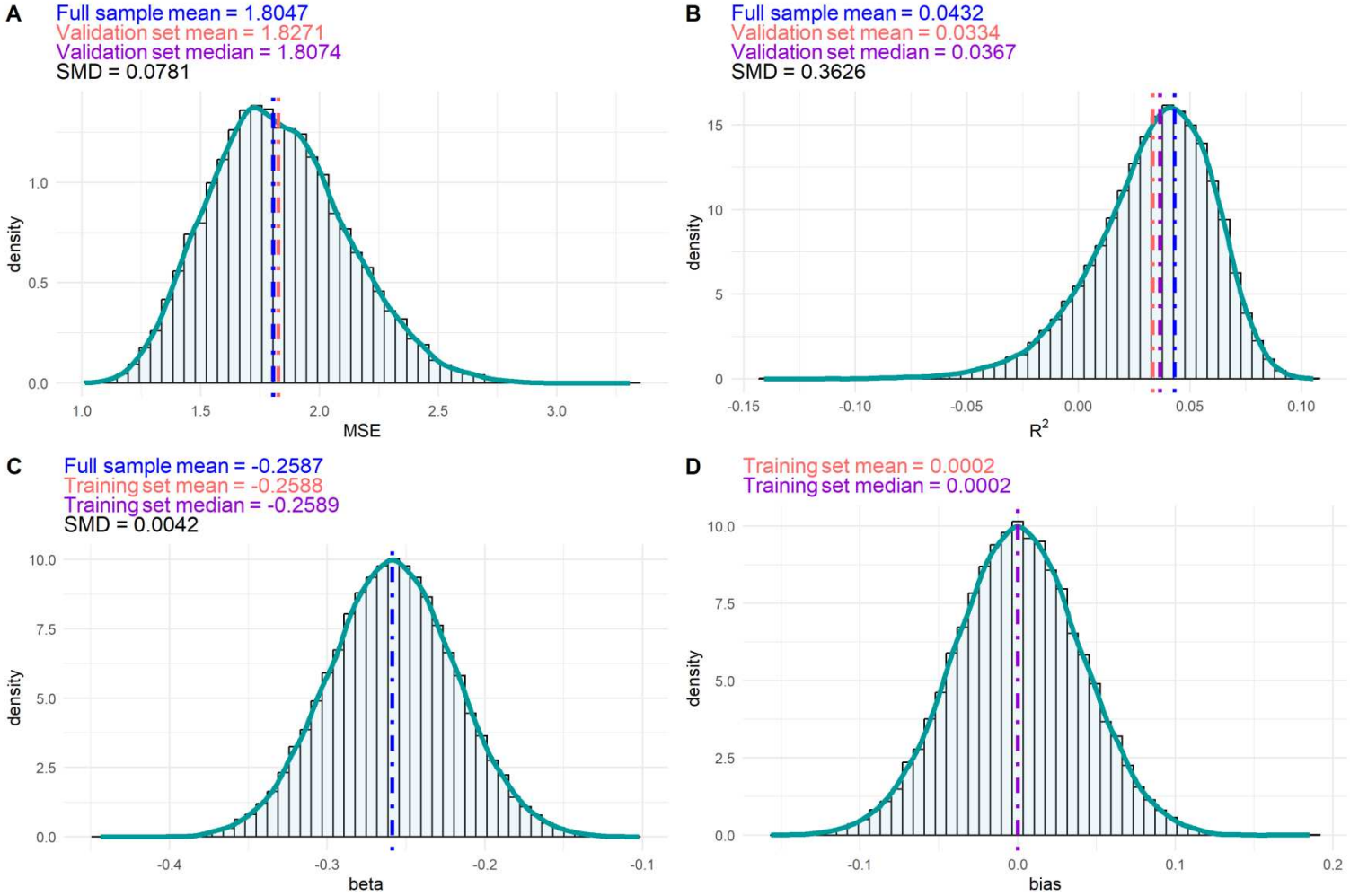
Cross-validation of the observed associations with eGFR decline

The MSE and R^2 values obtained from each multiple regression model (one per metabolite) estimated on the full sample closely match those derived from the validation samples. For example, the model incorporating baseline eGFR and propenoylcarnitine (Supplementary Figure S2) achieved an MSE of 1.797 and an R^2 of 0.047 in the full dataset, compared to a mean MSE of 1.820 and mean R^2 of 0.037 in the validation samples. This resulted in a negligible SMD of 0.08 for MSE and a small SMD of 0.37 for R^2, indicating successful validation. Similar results were observed across all models (Supplementary Figures S3-S12).

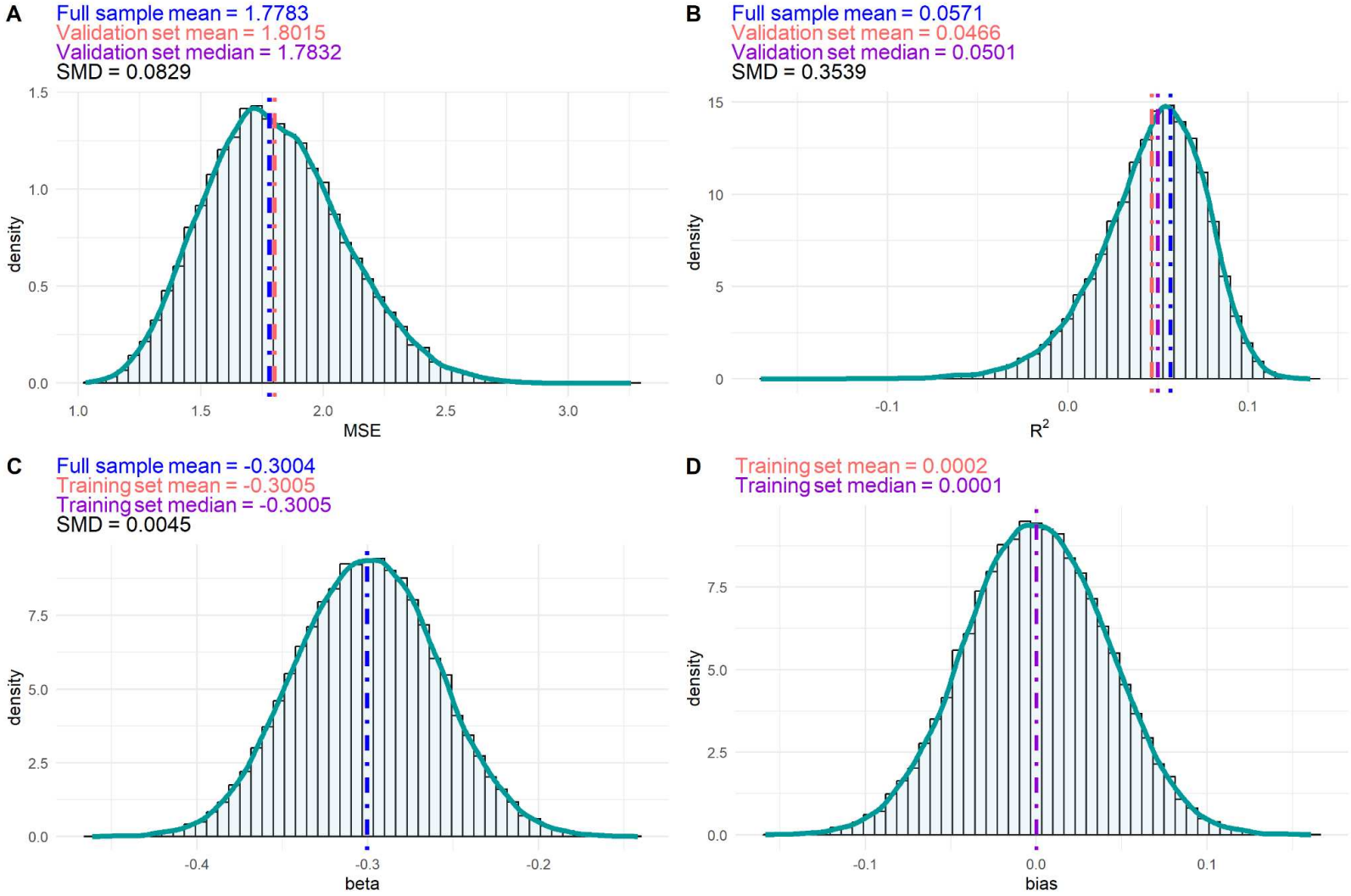
Supplemental Figure S3: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Propenoylcarnitine (C3:1)



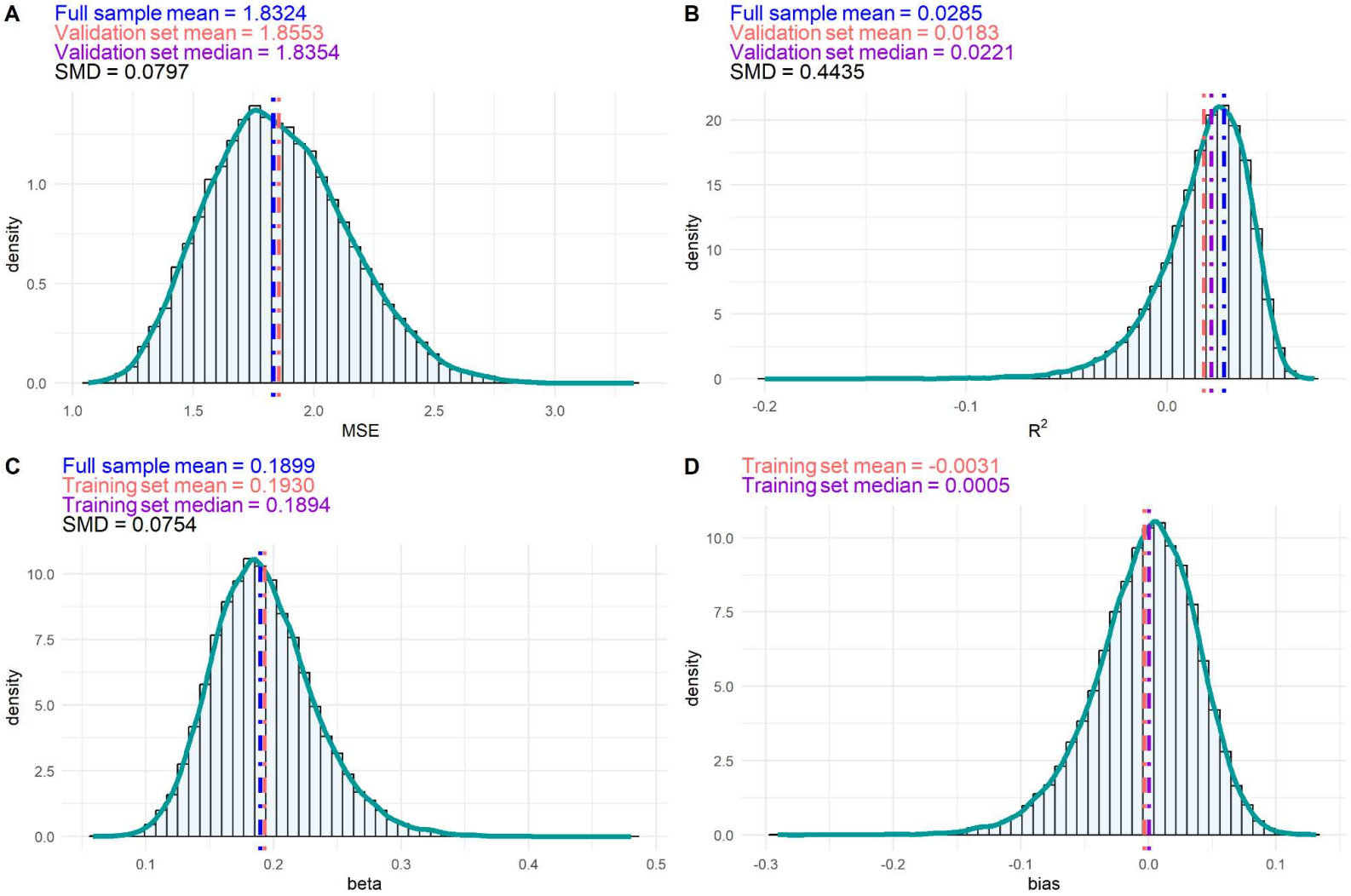
Supplemental Figure S4: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Tiglylcarnitine (C5:1)



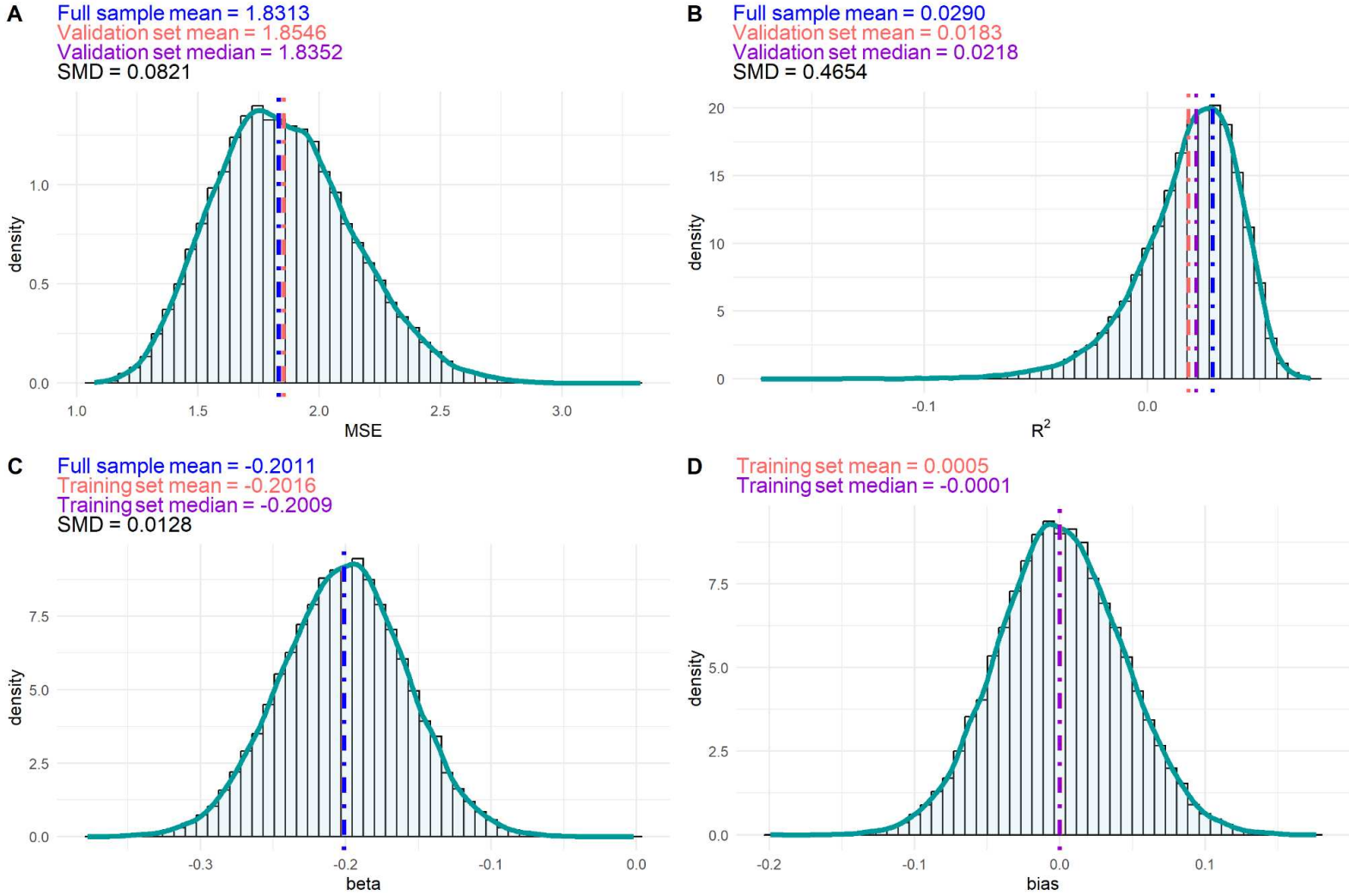
Supplemental Figure S5: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Hydroxyvalerylcarnitine C5-OH (C3-DC-M)



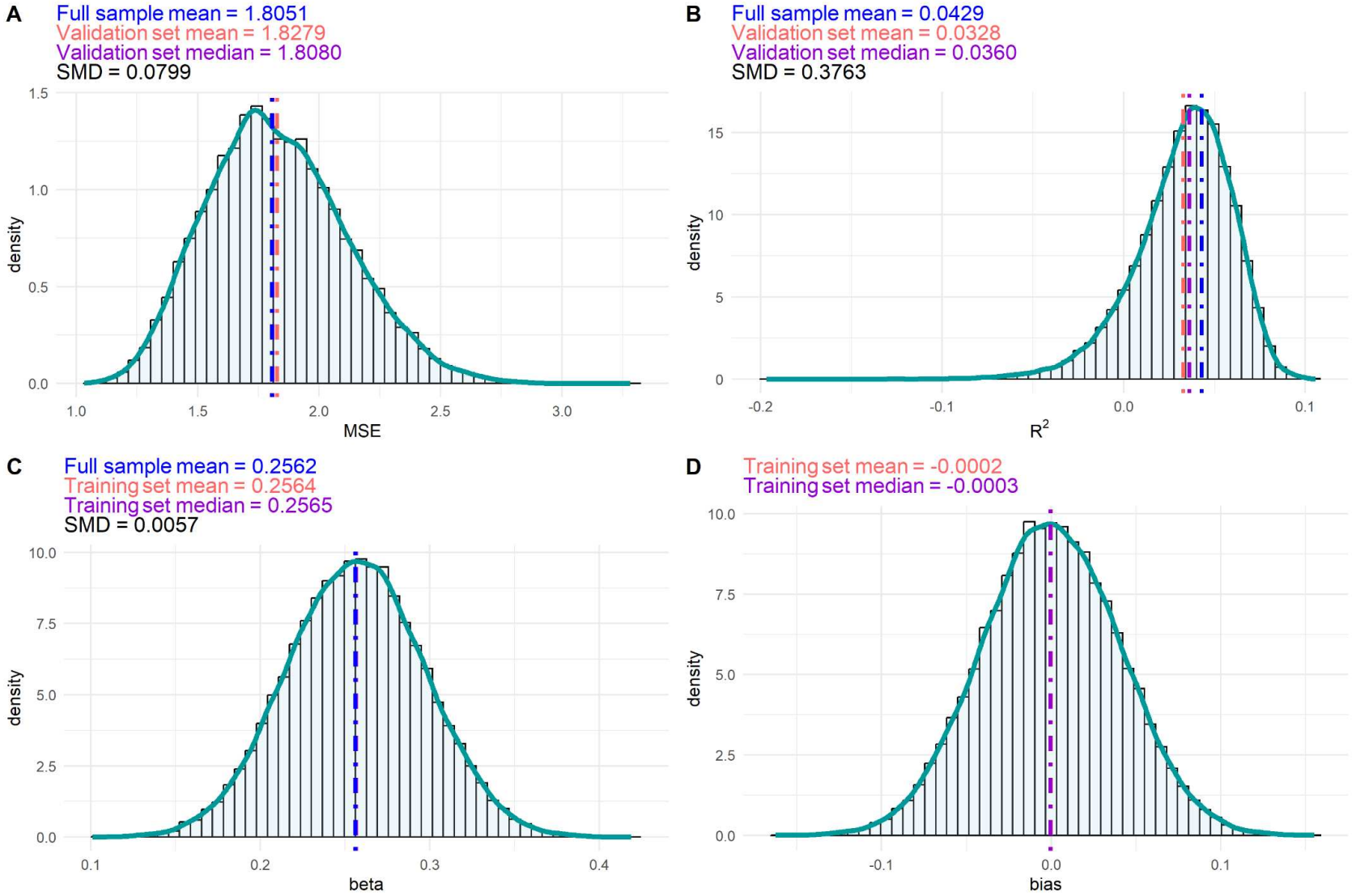
Supplemental Figure S6: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Hexenoylcarnitine (C6:1)



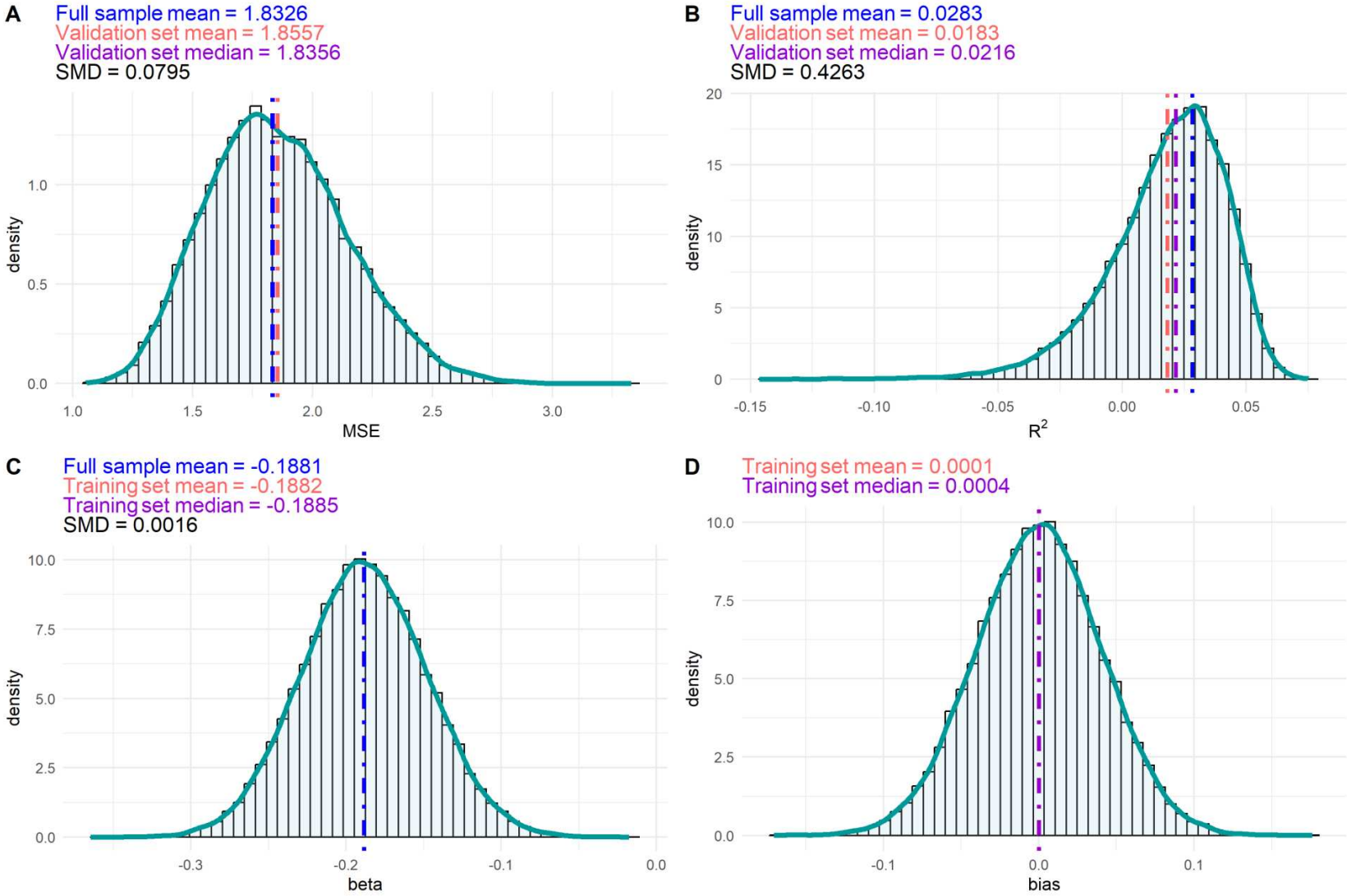
Supplemental Figure S7: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Decadienylcarnitine (C10:2)



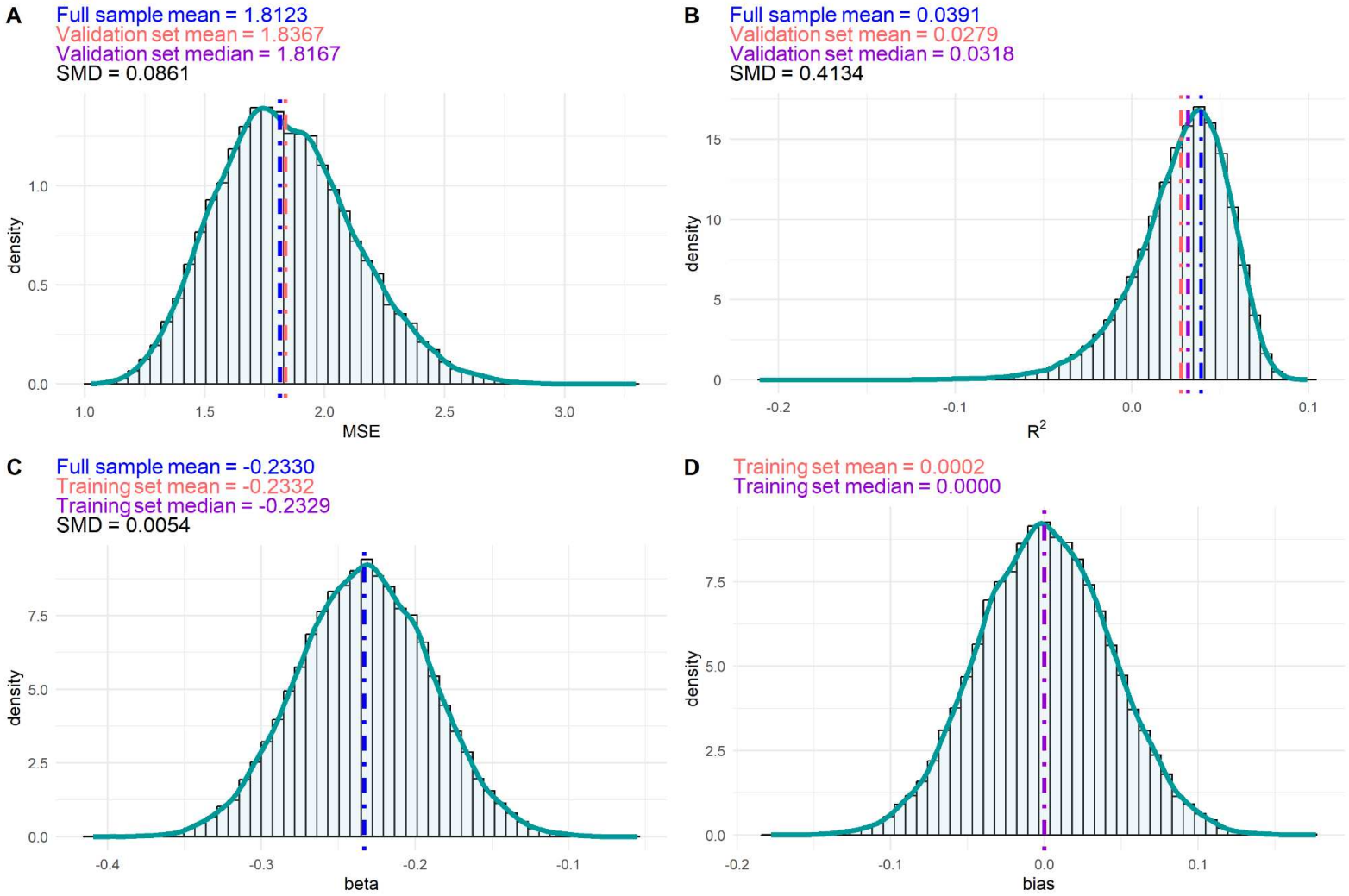
Supplemental Figure S8: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Dodecanedioylcarnitine (C12-DC)



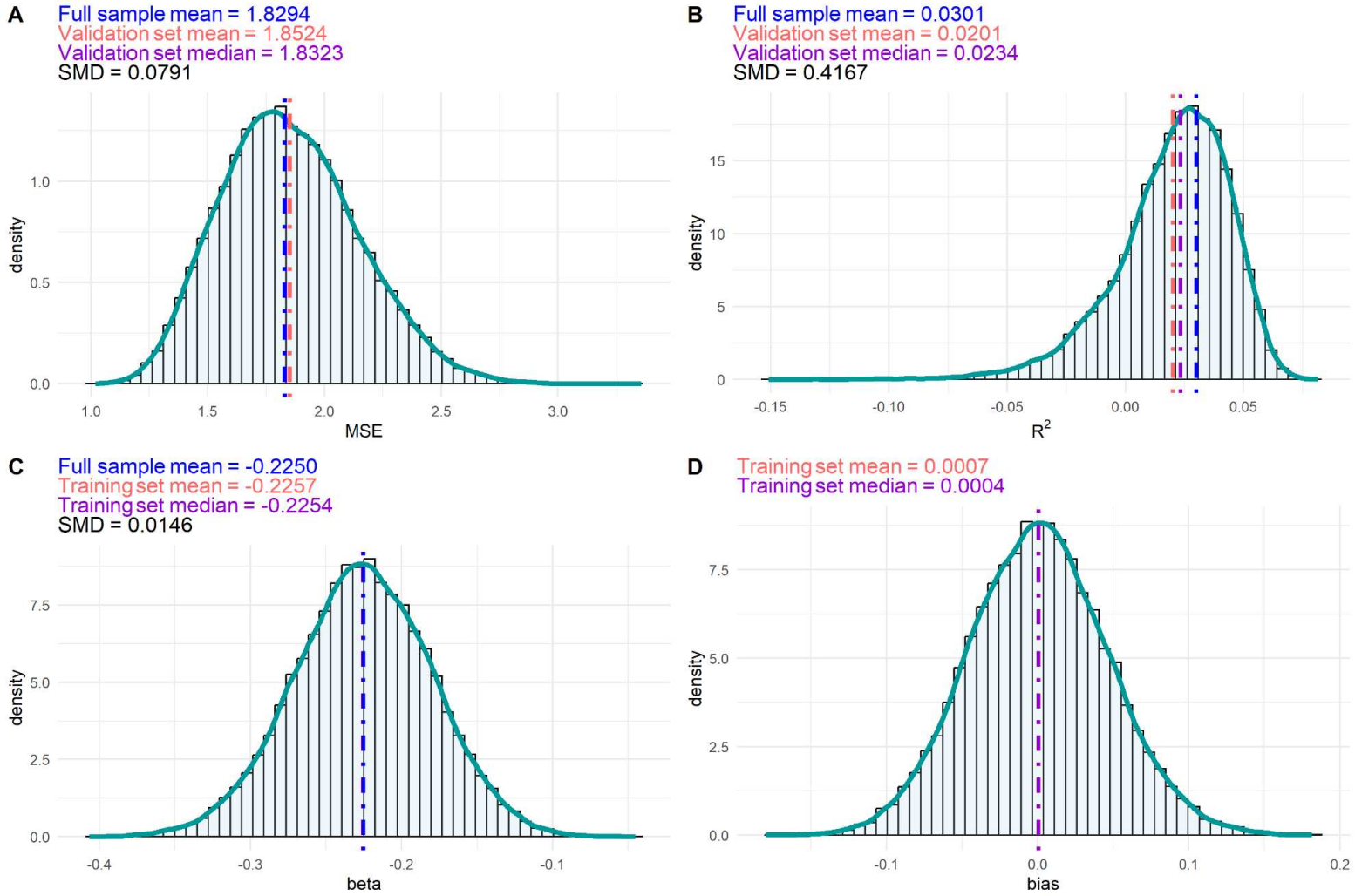
Supplemental Figure S9: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Etradecadienylcarnitine (C14:2)



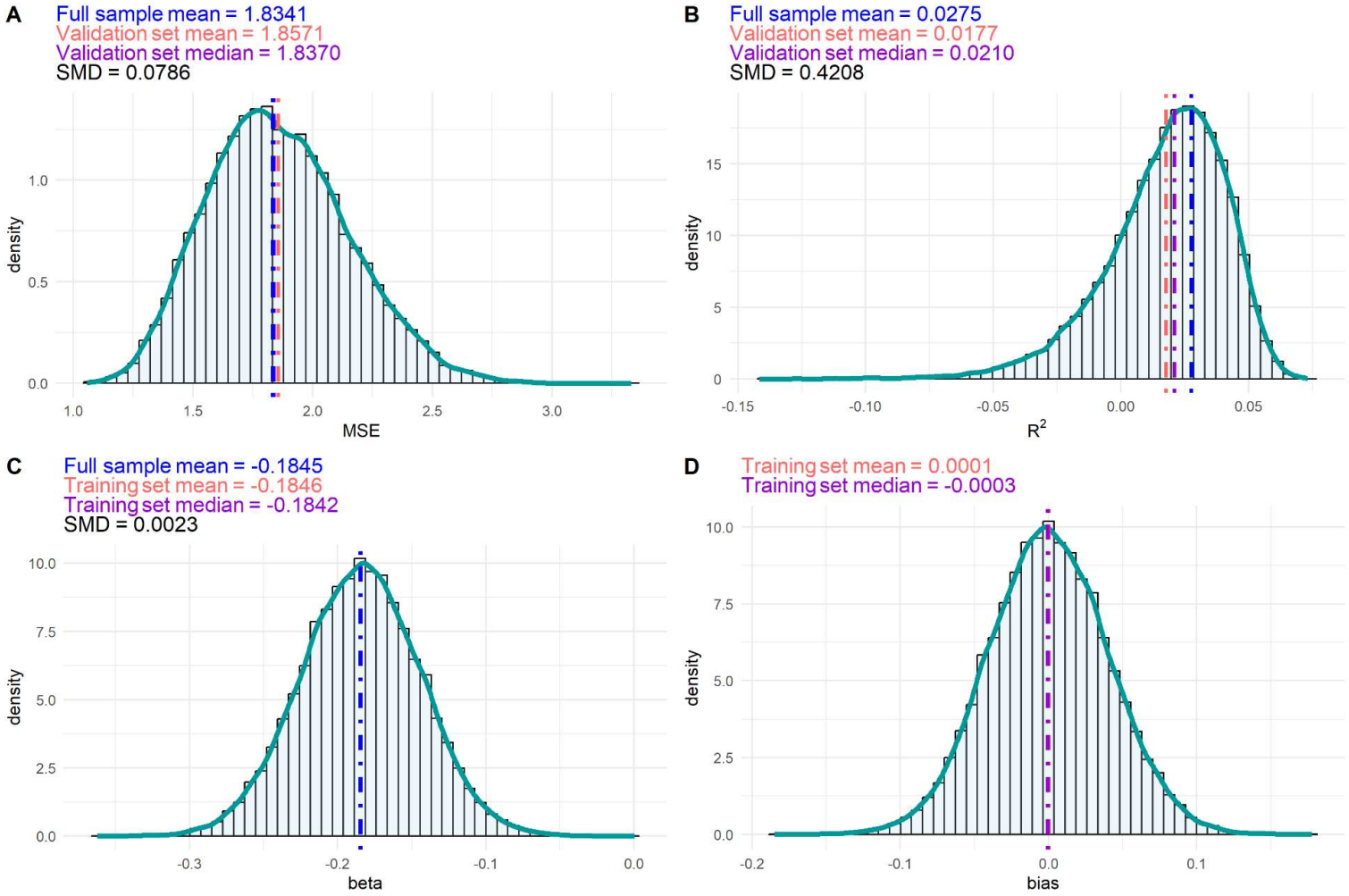
Supplemental Figure S10: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Hydroxyhexadecadienylcarnitine (C16:2-OH)



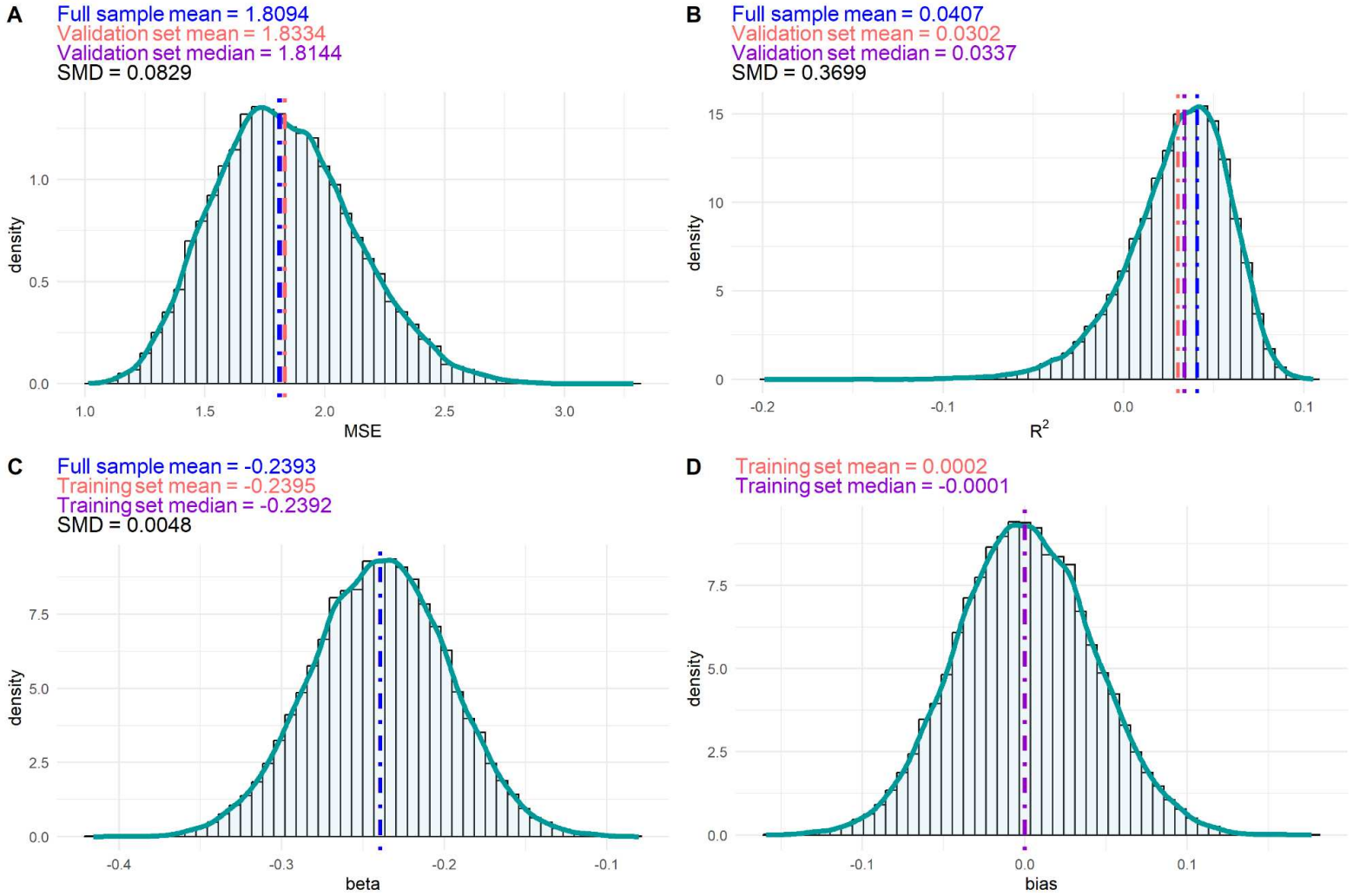
Supplemental Figure S11: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Methylglutarylcarnitine (C5-M-DC)



Supplemental Figure S12: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Myristoleoylcarnitine (C14:1)



Supplemental Figure S13: Distribution of the mean squared error MSE (**panel A**) and coefficient of determination R^2 (**panel B**) in the randomly generated validation sets, the regression coefficient for the acylcarnitine in the model (**panel C**) and its bias (**panel D**) from cross-validated multiple regression models on eGFR slope with respect to baseline eGFR and Palmitoleoylcarnitine (C16:1)



Rerferences

1. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J, Collaboration) C-ECKDE: A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604-612
2. Nowak N, Skupien J, Smiles AM, Yamanouchi M, Niewczas MA, Galecki AT, Duffin KL, Breyer MD, Pullen N, Bonventre JV, Krolewski AS: Markers of early progressive renal decline in type 2 diabetes suggest different implications for etiological studies and prognostic tests development. Kidney Int 2018;93:1198-1206
3. Scarale MG, Mastroianno M, Prehn C, Copetti M, Salvemini L, Adamski J, De Cosmo S, Trischitta V, Menzaghi C: Circulating Metabolites Associate With and Improve the Prediction of All-Cause Mortality in Type 2 Diabetes. Diabetes 2022;71:1363-1370
4. Trischitta V, Mastroianno M, Scarale MG, Prehn C, Salvemini L, Fontana A, Adamski J, Schena FP, Cosmo S, Copetti M, Menzaghi C: Circulating metabolites improve the prediction of renal impairment in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2023;11
5. Zukunft S, Prehn C, Röhring C, Möller G, Hrabě de Angelis M, Adamski J, Tokarz J: High-throughput extraction and quantification method for targeted metabolomics in murine tissues. Metabolomics 2018;14:18
6. Shah HS, Moreno LO, Morieri ML, Tang Y, Mendonca C, Jobe JM, Thacker JB, Mitri J, Monti S, Niewczas MA, Pennathur S, Doria A: Serum Orotidine: A Novel Biomarker of Increased CVD Risk in Type 2 Diabetes Discovered Through Metabolomics Studies. Diabetes Care 2022;45:1882-1892
7. Lumumba VW, Kiprotich D, Mpaine ML, Makena NG, MD. K: Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models. 2024, p. 127-137
8. Sawilowsky, SS: New effect size rules of thumb. Journal of Modern Applied Statistical Methods 2009;8:2