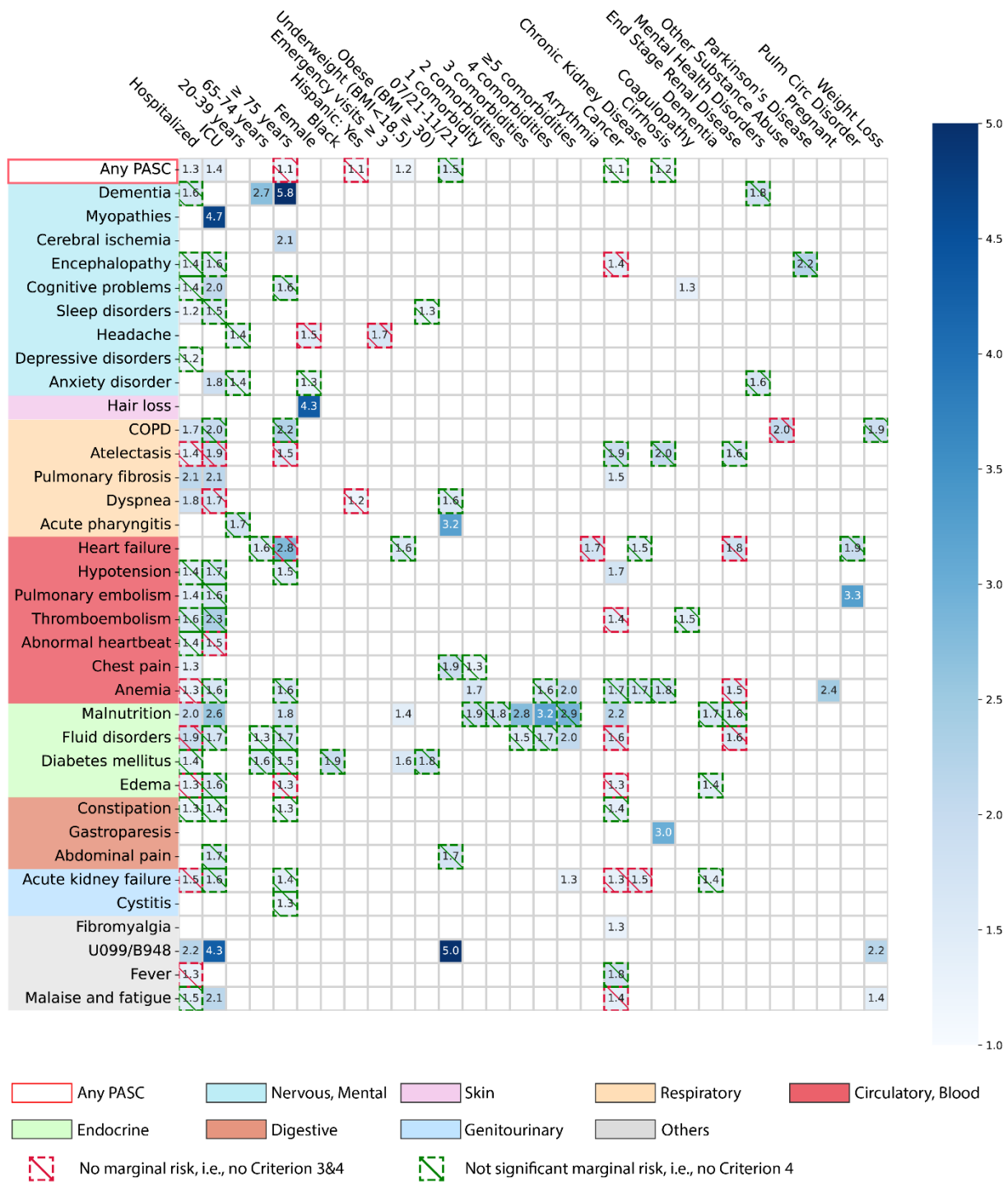
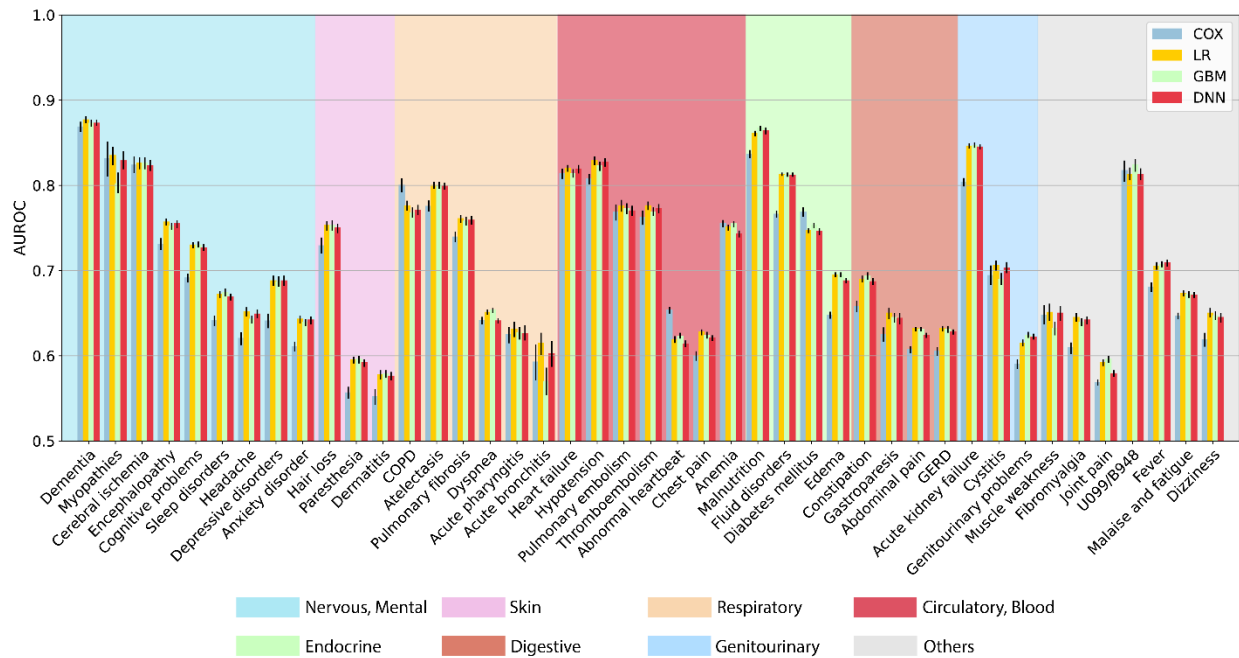


Supplementary Information

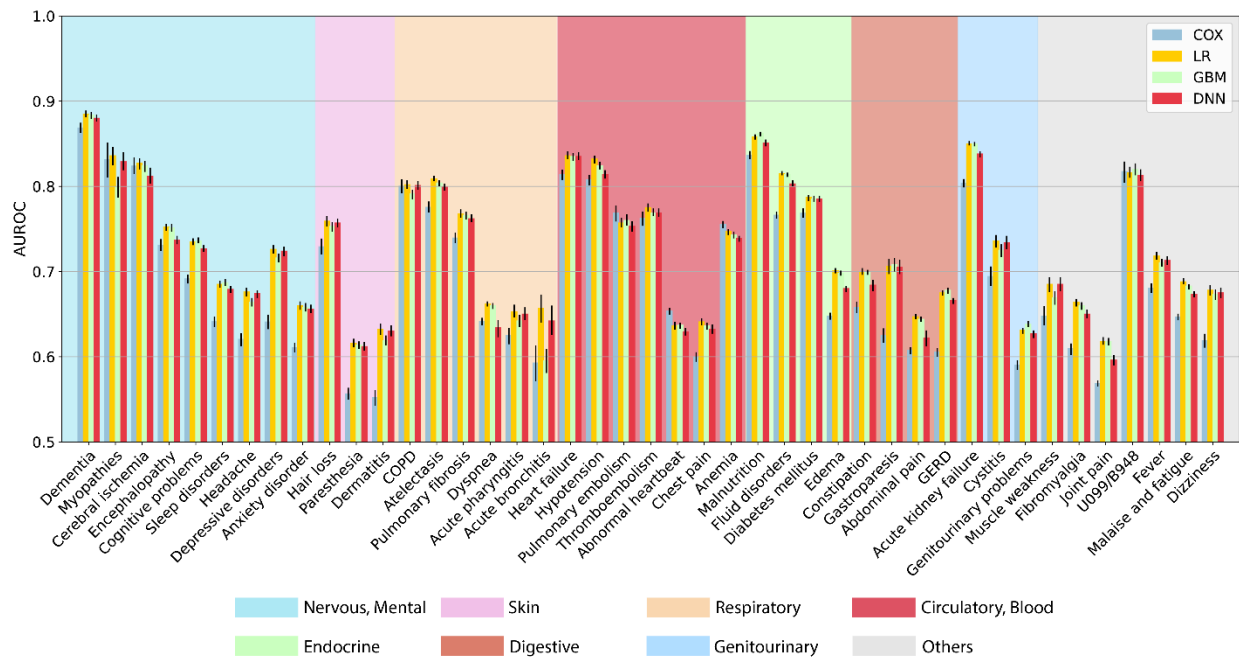
We acknowledge the following RECOVER-EHR Consortium Members: PCORnet Core Contributors-Louisiana Public Health Institute, Tom Carton, PI, Anna Legrand, Elizabeth Nauman; PCORnet Core Contributors-Weill Cornell Medicine, Rainu Kaushal, PI, Sajjad Abedian, Dominique Brown, Christopher Cameron, Thomas Campion, Andrea Cohen, Marietou Dione, Rosie Ferris, Wilson Jacobs, Michael Koropsak, Alex LaMar, Colby V. Lewis, Dmitry Morozyuk, Peter Morrissey, Duncan Orlander, Jyotishman Pathak, Mahfuza Sabiha, Edward J. Schenck, Stephenson Strobel, Zoe Verzani, Fei Wang, Mark Weiner, Zhenxing Xu, Chengxi Zang, Yongkang Zhang; PCORnet Data Contributors, Albert Einstein College of Medicine Parsa Mirhaji, PI | Columbia University Soumitra Sengupta, PI, | Emory University Rishi Kamaleswaran, PI | Icahn School of Medicine at Mount Sinai Carol R. Horowitz, PI | New York University Langone Health Saul Blecker, PI | University of Florida Mei Liu, PI, | University of Miami Nick Tsinoremas, PI | University of South Florida | Nicklaus Children's Hospital, Sandy Gonzalez, PI | Weill Cornell Medicine Rainu Kaushal, PI



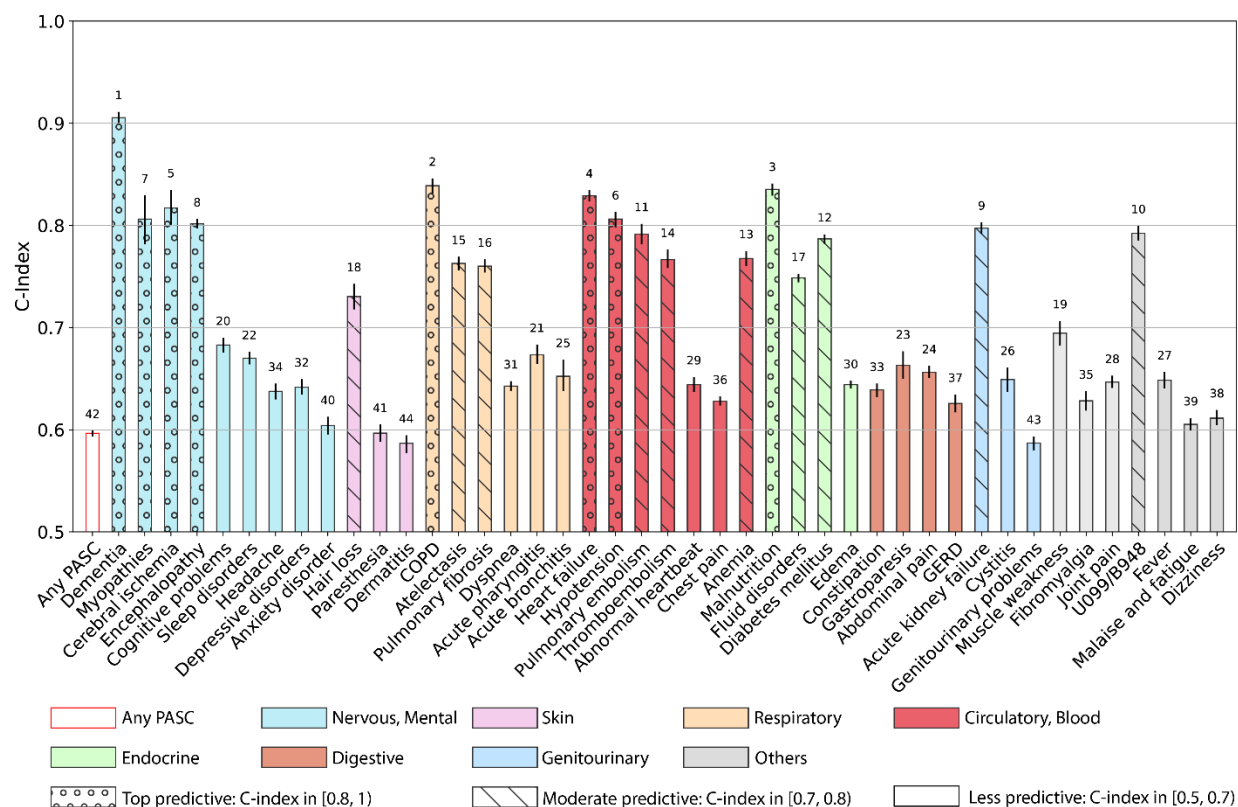
Supplementary Fig. 1. Sensitivity analyses of associations ruled out by screening criteria 3 or 4, the INSIGHT cohort, March 2020 to November 2021. The associations ruled out by Criterion 3, namely marginal risk than the control group, were highlighted in red squares, and the associations ruled out by Criterion 4, namely the marginal risk than the control group should be significant, were highlighted in green squares. See details of screen criteria in the Method-association analysis section. The adjusted hazard ratios of different factors were reported. The colors represent different risk levels. Any PASC represents having at least one of the conditions below. The color panels represent different organ systems, including (from top to bottom): the nervous system or mental disorders, skin, respiratory system, circulatory system, endocrine and metabolic, digestive system, genitourinary system, and other signs. ICD-10 codes B948 (sequelae of other specified infectious and parasitic diseases) and U099 (post-COVID-19 condition, unspecified) were used to capture general PASC diagnoses.



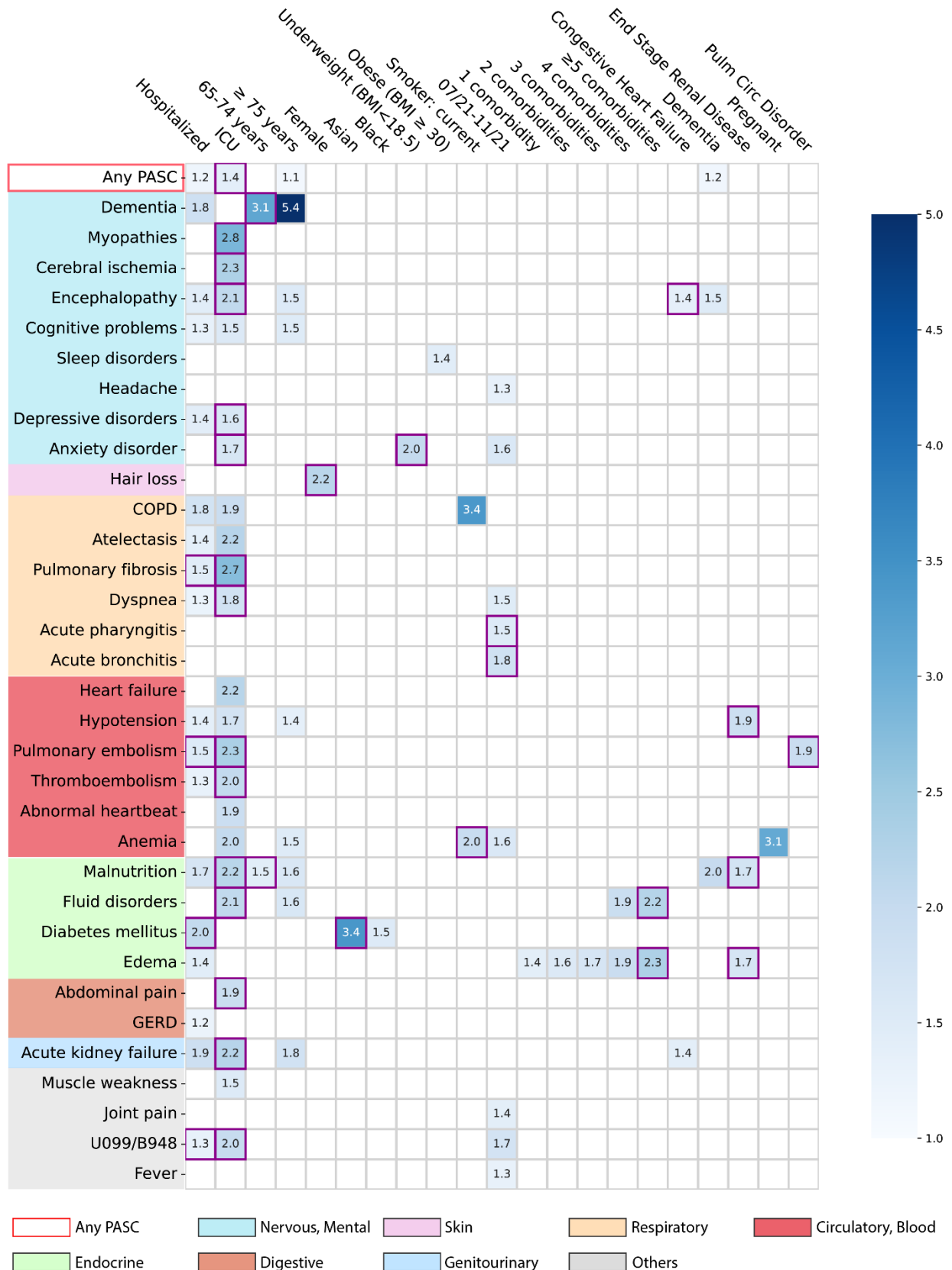
Supplementary Fig. 2. Predictive performance of incident post-acute sequelae of SARS-CoV-2 infection (PASC) by different machine learning models, based on the pre-defined baseline covariates and the severity of acute infection, INSIGHT, March 2020 to November 2021. The pre-defined baseline covariates were the same ones used in the risk factor analysis. The COX model was measured by C-index. Regularized logistic regression (LR), gradient boosting machine (GBM), and deep forward neural network (DNN) were measured by the area under the receiver operating characteristic curve (AUROC). The color panels represent different organ systems, including (from top to bottom): the nervous system or mental disorders, skin, respiratory system, circulatory system, endocrine and metabolic, digestive system, genitourinary system, and other signs. The error bars are 95% confidence interval estimated by 1000-times bootstrapping performance on the testing dataset in repeated cross-validation.



Supplementary Fig. 3. Predictive performance of incident post-acute sequelae of SARS-CoV-2 infection (PASC) based on building baseline covariates in a data-driven way, by different machine learning models, INSIGHT, March 2020 to November 2021. The data-driven baseline covariates were selected based on the first 3 digits of ICD-10 codes and prescribed medications at the ingredient level in a data-driven way. The COX model was measured by C-index. Regularized logistic regression (LR), gradient boosting machine (GBM), and deep forward neural network (DNN) were measured by the area under the receiver operating characteristic curve (AUROC). The color panels represent different organ systems, including (from top to bottom): the nervous system or mental disorders, skin, respiratory system, circulatory system, endocrine and metabolic, digestive system, genitourinary system, and other signs. The error bars are 95% confidence interval estimated by 1000-times bootstrapping performance on the testing dataset in repeated cross-validation.



Supplementary Fig. 4. Prediction performance of incident potential PASC conditions from baseline characteristics and the severity in the acute phase with the OneFlorida+ cohort from March 2020 to November 2021. The C-index with a 95% confidence interval was reported. Any PASC represents having at least one of the conditions below. The bars in different colors were organized by their organ systems including (from left to right): the nervous system or mental disorders, skin, respiratory system, circulatory system, endocrine and metabolic, digestive system, genitourinary system, and other signs. The conditions with a C-index in [0.8, 1) were highlighted with “o” texture, and those with a C-index in [0.7, 0.8) were highlighted with “\” texture. The numbers at the top of the bars denote the rank of the predictability quantified by the C-index among all the bars with 1 representing the top predictive condition. The error bars are 95% confidence interval estimated by 1000-times bootstrapping performance on the testing dataset in repeated cross-validation.



Supplementary Fig. 5. Identified risk factors associated with incident PASC conditions from the OneFlorida+ cohort, March 2020 to November 2021. The adjusted hazard ratios of different factors were reported. The colors represent different risk levels. The associations whose marginal increased risks than non-infected control patients were also significant were highlighted in purple squares. Any PASC represents having at least one of the conditions below. The color panels represent different organ systems, including (from top to bottom): the nervous system or mental disorders, skin, respiratory system, circulatory system, endocrine and metabolic, digestive system, genitourinary system, and other signs. ICD-10 codes B948 (sequelae of other specified infectious and parasitic diseases) and U099 (post-COVID-19 condition, unspecified) were used to capture general PASC diagnoses.