

Tinnitus risk factors and its evolution over time: a cohort study

Appendix

A1. Developing the predictive models predicting tinnitus presence and severity

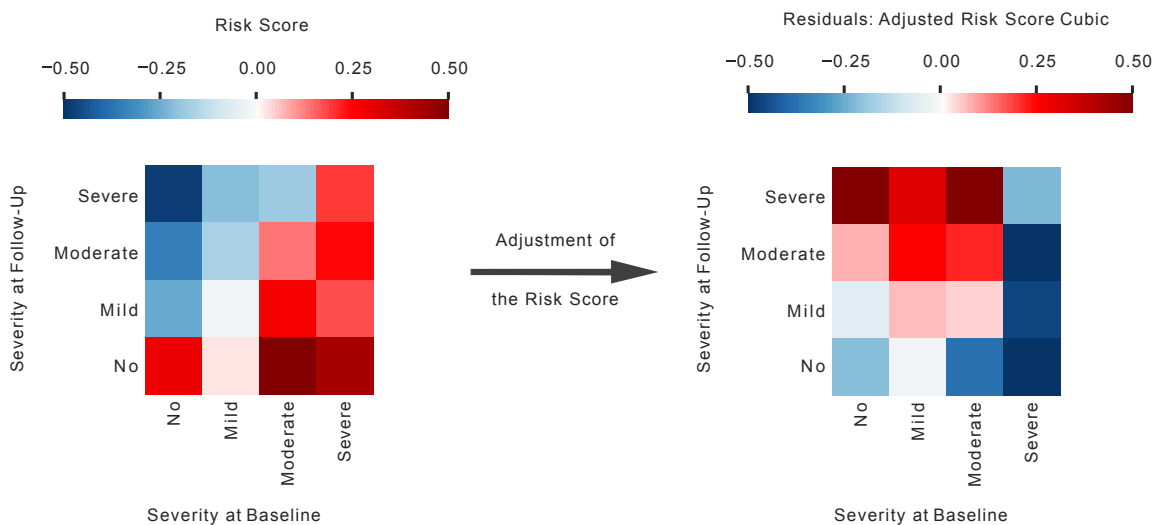
The nonlinear iterative partial least square (NIPALS)²⁷ method was trained within the discovery dataset to identify latent scores (\mathbf{E} and \mathbf{Z}) and loadings (\mathbf{P} and \mathbf{H}) that maximize the covariance between matrix of standardized psychosocial features (X_i) (size (147385, 101) for tinnitus presence, and (43983, 101) for tinnitus severity) and a vector of self-reported tinnitus presence or severity (\mathbf{Y}) (size (147385, 1) for tinnitus presence, and (43983, 1) for tinnitus severity):

1. Compute singular vectors $\boldsymbol{\mu}$, \mathbf{v} (weights) of covariance matrix $C = X^T Y$
2. Obtain the latent scores \mathbf{E} and \mathbf{Z} by projecting X and Y onto singular vectors $\boldsymbol{\mu}$ and \mathbf{v}
3. Compute loadings \mathbf{P} and \mathbf{H} by iteratively regressing X onto \mathbf{E} (power iteration)
4. Deflate X and Y using $X + 1 = X - \mathbf{E}\mathbf{P}^T$ and $Y + 1 = Y - \mathbf{Z}\mathbf{H}^T$, respectively
5. Fit training (discovery) data X using the projection matrix \mathbf{P} to obtain latent space \bar{x} so that $\bar{x} = X\mathbf{P}$
6. Use the latent space to predict left-out data \mathbf{Y}_v (size (20853,1) for tinnitus presence, (4291,1) for tinnitus severity) using the coefficient matrix $\beta \in R^{d \times t}$ such that $\mathbf{Y}_v = X_v \beta$, where X_v denotes the matrix of psychosocial features in the validation set (X_v size (20853,101) for tinnitus presence, (4291,101) for tinnitus severity)

Further information on the implementation can be found at https://scikit-learn.org/stable/modules/cross_decomposition.html#cross-decomposition.

A2. Adjustment of the risk scores for the longitudinal evaluations

To examine the prognostic value of the presence risk score, we regressed out the squared values of the tinnitus presence level (0: No, 1: Some of the time, 2: A lot of the time, 3: All of the time) at baseline from presence the risk score calculated at baseline. Making the score orthogonal to the Tinnitus presence level at baseline allowed us to interpret interindividual deviations in this adjusted score as risk of recovery or worsening of tinnitus presence at the follow-up visit. Similar analysis was performed on the severity risk score with the levels 0: No, 1: Mild, 2: Moderate and 3: Severe). The process is depicted in the following figure.



A3. Reduced risk score replication: TRI Database

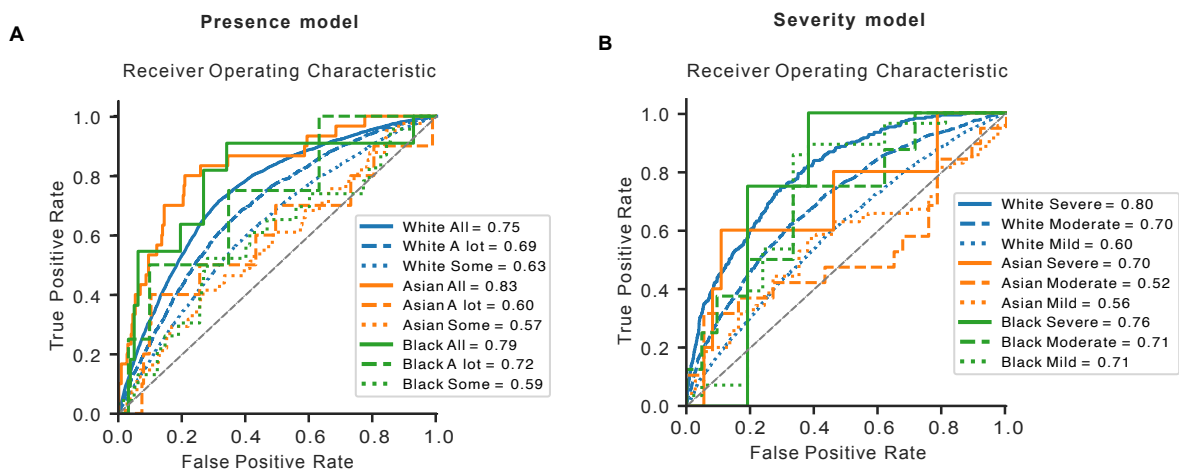
The reduced severity risk score was derived using five binarized items:

1. **Do you think you have a hearing problem?** 0: No / 1: Yes
2. **Do you wear hearing aids** 0: No / 1: Yes
3. **How much of the time have you had trouble sleeping at night?** 0: At no time, 0: Some of the time, 0: Slightly less than half the time / 1 = Slightly more than half the time, 1: Most of the time, 1: All the time
4. **How much of the time have you felt lacking in energy and strength?** 0: At no time, 0: Some of the time, 0: Slightly less than half the time / 1 = Slightly more than half the time, 1: Most of the time, 1: All the time
5. **How much of the time have you had a bad conscience or feelings of guilt?** 0: At no time, 0: Some of the time, 0: Slightly less than half the time / 1 = Slightly more than half the time, 1: Most of the time, 1: All the time

The TRI database did not contain information on speech-in-noise hearing difficulties, which could be a proper equivalent of the original item “Do you find it difficult to follow a conversation if there is background noise (such as TV, radio, children playing)?”. We replaced it by the hearing health item having the larger weight in the severity model, which is “Do you wear hearing aids?”.

A4: Validation of the presence and severity risk scores over ethnicities

To ensure the validity of the presence and severity models on different ethnicities, we tested the performances of the models separately on groups of Asian, Black and White individuals. We tested participants which were not included in the rest of the analysis, using data of participants who did not had a full hearing evaluation at baseline visit but had one in the follow-up visit. We tested 466 Asian, 203 Black and 41,969 White individuals with the presence model, and 81 Asian, 38 Black and 10,115 White individuals with the severity model. The presence model (**A**) had good to excellent performances to classify individuals perceiving tinnitus a lot or all the time (except for the Asian “a lot of the time” group). The severity model (**B**) had good to excellent performances to classify individuals with moderate or severe distress (except for the Asian “moderate” group). Overall, the models had good performances for the extreme levels (all the time, or severe distress) for all ethnicities. The performances for the intermediate levels (A lot of the time, moderate distress) were good for Black and White individuals, but low for Asian. Those results should be replicated on larger groups of individuals.



A5. Examining the severity at follow-up depending on the reduced risk score at baseline

This figure depicts the odd ratios of experiencing a severe, moderate, mild or no distress associated with tinnitus at the follow-up visit, depending on the Reduced risk score at baseline. Based on those odd ratios, we can conclude that 0 and 1 risk scores are associated with a low risk of experiencing moderate or severe tinnitus over time. Scores 2 and 3 are associated with a moderate risk, and score 4 and 5 are associated with a large risk of experiencing moderate or severe tinnitus distress over time.

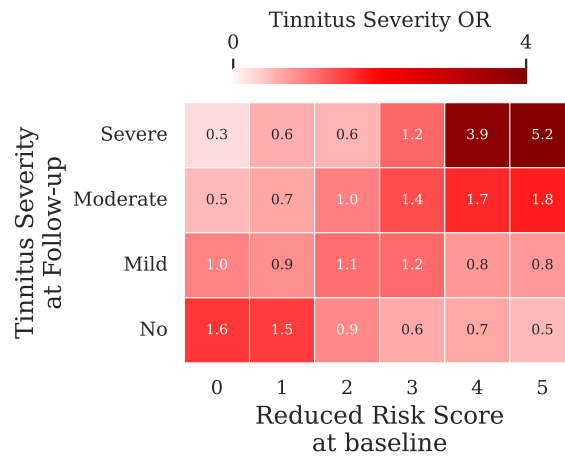


Table A1: Variables extracted from the UK Biobank and their associated data field.

Type of measure	Details	UK Biobank Data Field
Tinnitus evaluation		
Tinnitus		<u>4803</u>
Tinnitus severity/nuisance		<u>4814</u>
Hearing		
Hearing difficulties self-reported		<u>2247</u>
Hearing difficulties self-reported with background noise		<u>2257</u>
Hearing test right ear		<u>20021</u>
Hearing test left ear		<u>20019</u>
Hearing aid		<u>3393</u>
Cochlear implants		<u>4792</u>
Noisy workplace		<u>4825</u>
Loud music exposure		<u>4836</u>
Mood		
Frequency of depressed mood in last 2 weeks		<u>2050</u>
Frequency of unenthusiasm/disinterest in last 2 weeks		<u>2060</u>
Frequency of tenseness/restlessness in last 2 weeks		<u>2070</u>
Frequency of tiredness/lethargy in last 2 weeks		<u>2080</u>
Seen doctor (GP) for nerves, anxiety, tension or depression		<u>2090</u>
Seen psychiatrist for nerves, anxiety, tension or depression		<u>2100</u>
Risk taking	Whether the participant describes themselves as someone who takes risks	<u>2040</u>
Neuroticism		
Neurotic behaviors	Mood swings	<u>1920</u>
	Miserableness	<u>1930</u>
	Irritability	<u>1940</u>
	Sensitivity/hurt feelings	<u>1950</u>
	Fed-up feelings	<u>1960</u>
	Nervous feelings	<u>1970</u>
	Worrier/anxious feelings	<u>1980</u>
	Tens/highly strung	<u>1990</u>
	Worry too long after embarrassment	<u>2000</u>
	Suffer from nerves	<u>2010</u>
	Loneliness	<u>2020</u>
	Guilty feeling	<u>2030</u>
Total neuroticism score	Derived from the 12 neurotic behaviors, as a sum score of the total “yes” answers to these questions	<u>20127</u>
Life stressors		
Serious illness, injury, bereavement in the last 2 years	Serious trauma to self	<u>6145</u>
	Serious trauma to close relative	
	Death of close relative	
	Death of spouse or partner	
	Marital separation/ divorce	
	Financial difficulties	
Life stressors within last 2 years	Derived from the life stressors 6145, as a sum score of the total “yes” answers to these questions	-
Sleep		
Sleep duration	Hours of sleep in every 24 hours	<u>1160</u>

Getting up in the morning	Difficulty getting up	<u>1170</u>
Sleeplessness/insomnia	Trouble falling asleep at night or waking up in the middle of the night	<u>1200</u>
Nap during the day		<u>1190</u>
Chronotype	Late chronotype (evening person) Early chronotype (morning person)	<u>1180</u>
Daytime dozing (narcolepsy)	Likely to doze off or fall asleep during the daytime	<u>1220</u>
Physical activity		
Hand grip strength	Average between left- and right-hand grip strength (units of measurement: Kg).	<u>46 & 47</u>
IPAQ activity group	Low Moderate High	<u>22032</u>
MET minutes per week for walking	Units of measurement: minutes/week	<u>22037</u>
MET minutes per week for moderate activity	Units of measurement: minutes/week	<u>22038</u>
Above moderate/vigorous recommendation	Indicates if the participant met the 2017 UK Physical activity guidelines of 150 minutes of moderate activity per week or 75 minutes of vigorous activity	<u>22035</u>
Substance Use		
Smoking status	Previous smoker	<u>20116</u>
Current tobacco smoking	Daily smoker Occasional smoker	<u>1239</u>
Past tobacco smoking	How often the participant smoked tobacco in the past	<u>1249</u>
Ever smoked	Derived from current and past tobacco smoking status (data fields 1239 & 1249)	<u>20160</u>
Smokers in household	If anyone smokes in the participant's household	<u>1259</u>
Hours of exposure to tobacco at home	Hours per week	<u>1269</u>
Alcohol drinker status	Never drank Previous drinker	<u>20117</u>
Alcohol intake frequency		<u>1558</u>
Alcohol intake vs. 10 years previously	Indicates the change in the participant's alcohol use compared to 10 years ago	<u>1628</u>
Anthropometric		
Body mass index (BMI)	Constructed from height and weight measured during the in-person assessment visit	<u>21001</u>
Weight change compared with one year ago	Gained weight Lost weight	<u>2306</u>
Weight	Unit of measurement: kg	<u>21002</u>
Systolic blood pressure	Units of measurement: mmHg	<u>4080</u>
Diastolic blood pressure	Units of measurement: mmHg	<u>4079</u>
Pulse rate	Pulse rate measured during the automated blood pressure readings (units of measurement: bpm).	<u>102</u>
Fractured/broken bones in last 5 years		<u>2463</u>
Occupational		
Job involves heavy manual or physical work		<u>816</u>
Job involves walking or standing		<u>806</u>
Current employment status	Looking after home or family Paid employment/self-employed Retired Unemployed Unable to work or disable	<u>6142</u>
Education (qualifications)	College or university Advanced level	<u>6138</u>

	Ordinary level Certificate secondary education Practical career diploma Other professional qualifications	
Demographics		
Ethnic background	White Asian Black Mixed Other	<u>21000</u>
Sex	Female / Male	<u>31</u>
Age	Age in years	<u>21003</u>
Socioeconomics		
Number in household	Number of people living together in participant's household including themselves	<u>709</u>
Relationship with people living in household	Living with partner Living with children Living with siblings Living with parents Living with grandparents Living with grandchildren Living with related relatives Living with unrelated relatives	<u>6141</u>
Frequency of friends and family visits		<u>1031</u>
Able to confide	How often the participant has been able to confide to someone close to them	<u>2110</u>
Number of vehicles in household		<u>728</u>
Average total household income		<u>738</u>

Figure A1: Features weights of the Presence Risk Score model

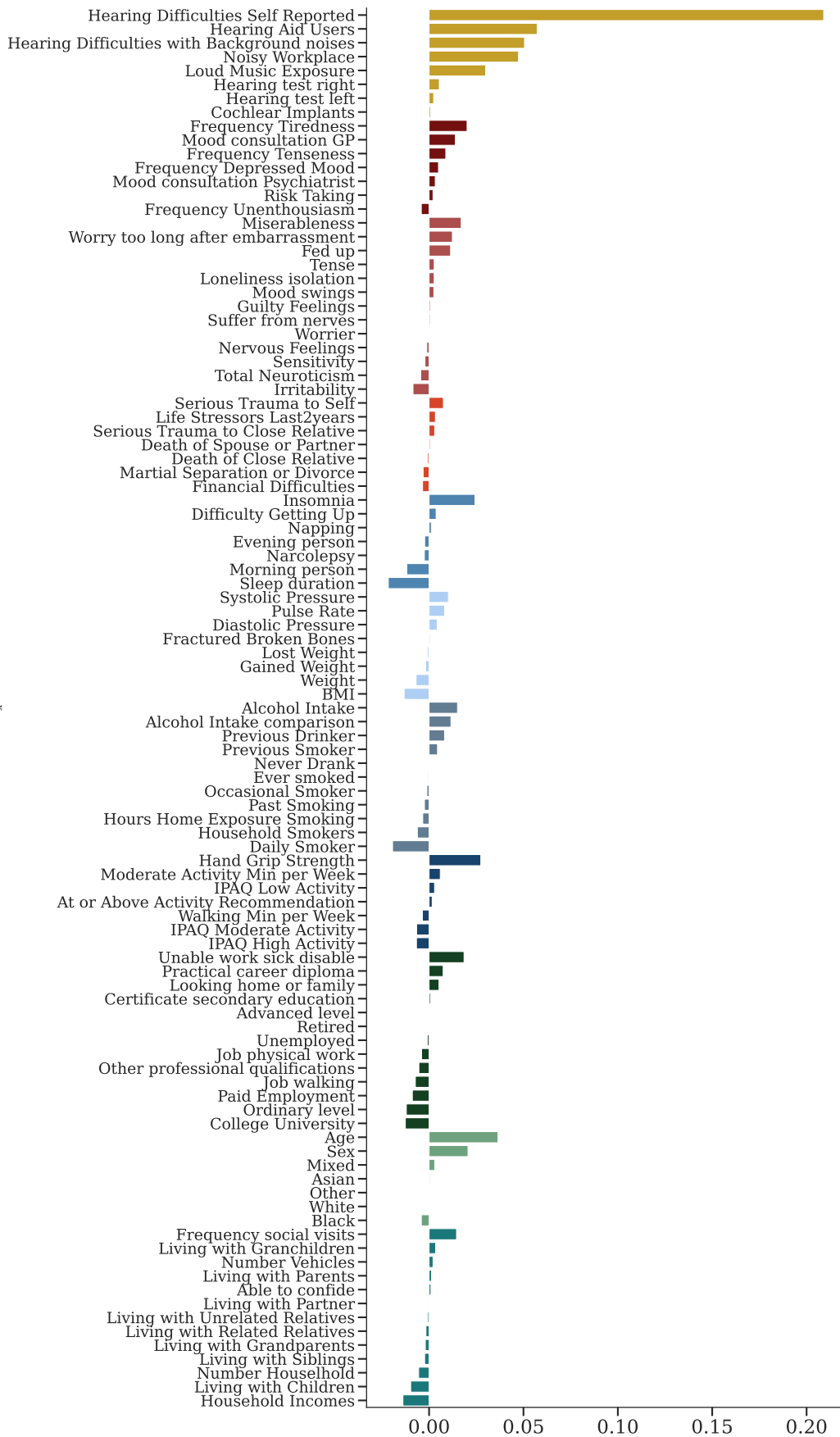


Figure A2: Features loads (Pearson's r correlation coefficient between features and the models' scores) of the Presence Risk Score model

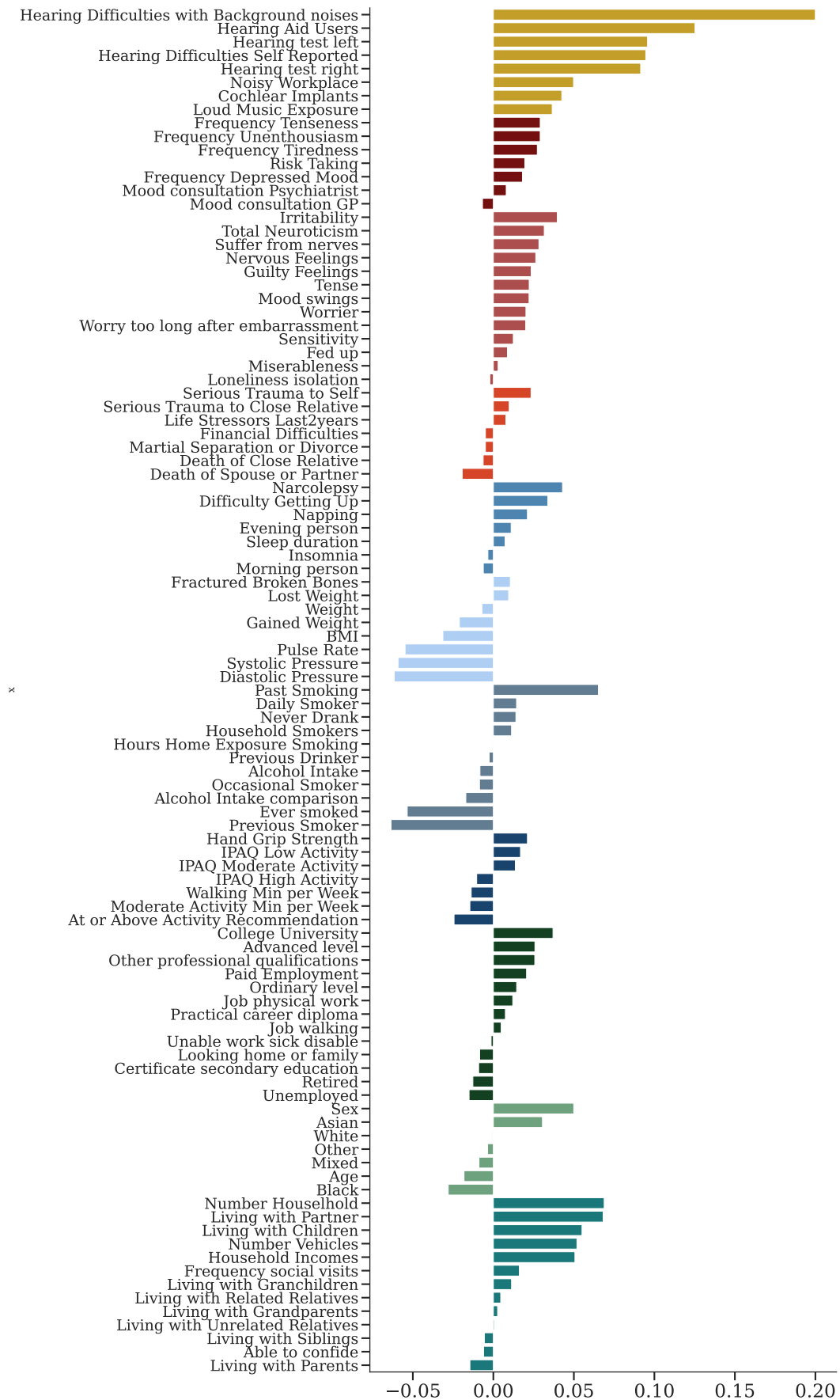


Figure A3: Features weights of the Severity Risk Score model

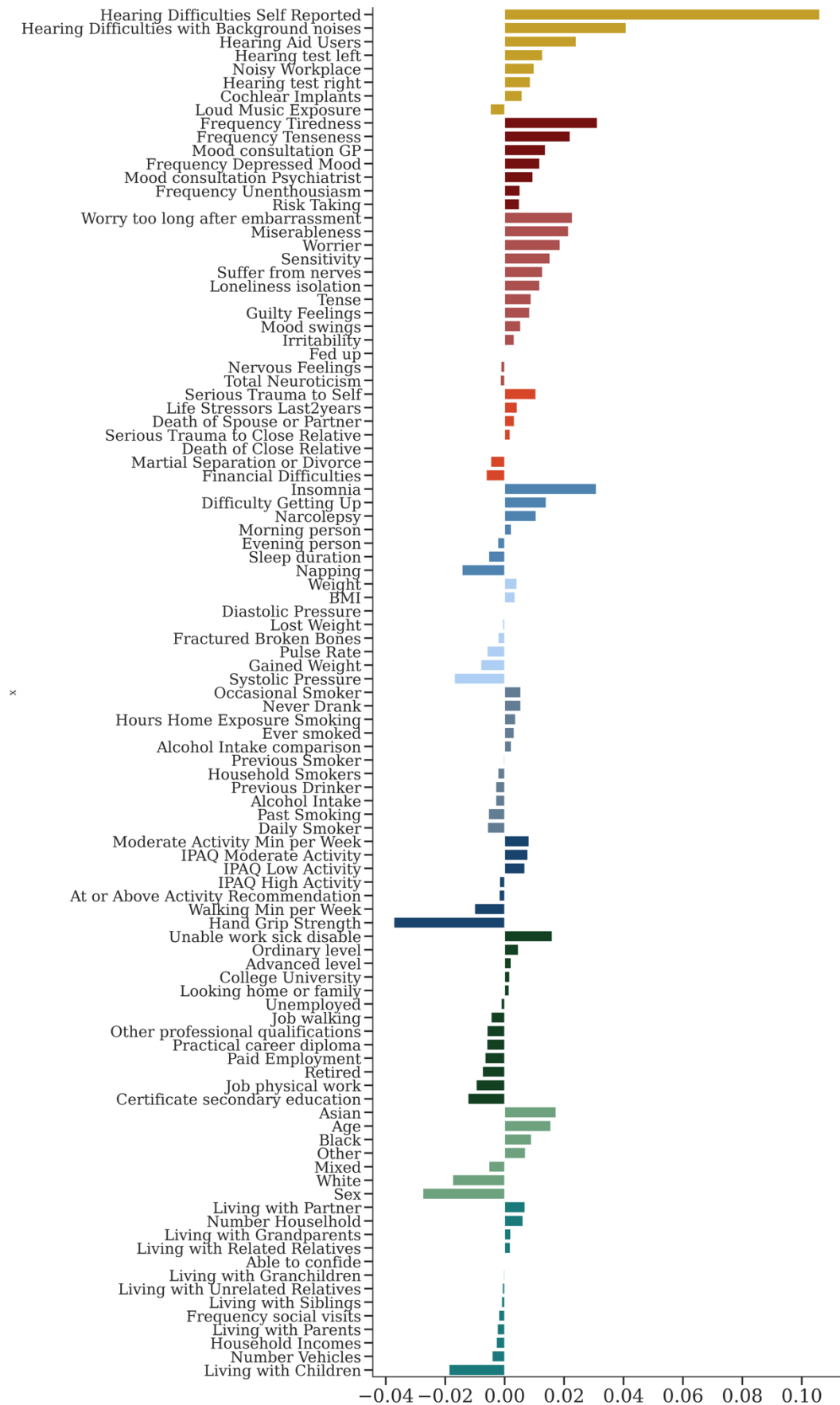


Figure A4: Features loads (Pearson's r correlation coefficient between features and the models' scores) of the Severity Risk Score model

