



## Article

<https://doi.org/10.1038/s44220-024-00294-2>

# Building machine learning prediction models for well-being using predictors from the exposome and genome in a population cohort

---

Received: 9 October 2023

Accepted: 11 July 2024

Published online: 14 August 2024

Check for updates

---

Dirk H. M. Pelt<sup>1,2</sup>✉, Philippe C. Habets<sup>1</sup>, Christiaan H. Vinkers<sup>1,2,3,4,5</sup>,  
Lannie Ligthart<sup>1</sup>, Catharina E. M. van Beijsterveldt<sup>1</sup>, René Pool<sup>1</sup> & Meike Bartels<sup>1,2</sup>

Effective personalized well-being interventions require the ability to predict who will thrive or not, and the understanding of underlying mechanisms. Here, using longitudinal data of a large population cohort (the Netherlands Twin Register, collected 1991–2022), we aim to build machine learning prediction models for adult well-being from the exposome and genome, and identify the most predictive factors ( $N$  between 702 and 5874). The specific exposome was captured by parent and self-reports of psychosocial factors from childhood to adulthood, the genome was described by polygenic scores, and the general exposome was captured by linkage of participants' postal codes to objective, registry-based exposures. Not the genome ( $R^2 = -0.007 [-0.026-0.010]$ ), but the general exposome ( $R^2 = 0.047 [0.015-0.076]$ ) and especially the specific exposome ( $R^2 = 0.702 [0.637-0.753]$ ) were predictive of well-being in an independent test set. Adding the genome ( $P = 0.334$ ) and general exposome ( $P = 0.695$ ) independently or jointly ( $P = 0.029$ ) beyond the specific exposome did not improve prediction. Risk/protective factors such as optimism, personality, social support and neighborhood housing characteristics were most predictive. Our findings highlight the importance of longitudinal monitoring and promises of different data modalities for well-being prediction.

Well-being is generally defined as a multidimensional concept encompassing not only the absence of negative aspects (such as illness or distress) but also the presence of positive elements, including a sense of fulfillment, productivity and overall flourishing<sup>1</sup>. Different conceptualizations of well-being exist, the most common distinction being made between subjective well-being and psychological well-being, also respectively known as hedonic and eudaimonic well-being<sup>1,3</sup>. The current study focuses, due to data availability, on subjective well-being, defined as the cognitive and affective evaluation of one's life. The cognitive aspect is usually indicated by life satisfaction or quality of life,

the affective aspect by positive affect or happiness and the absence of negative affect<sup>2</sup>. For brevity, we simply use the term well-being.

For the development of personalized well-being interventions, it is important to be able to predict who will thrive and to understand why this is the case. In the current study, we aim to develop optimal risk prediction models for well-being at the individual level and identify the factors driving these predictions<sup>4</sup>. Previous studies have provided possible risk and protective factors from the exposome and genome for well-being. The exposome is defined as all the environmental exposures one is exposed to during lifetime, and is increasingly being linked

---

A full list of affiliations appears at the end of the paper. ✉e-mail: [h.m.pelt@vu.nl](mailto:h.m.pelt@vu.nl)

to health and well-being<sup>5–9</sup>. It consists of three interlinked and partly overlapping categories: the internal, specific external, and general external exposome<sup>5</sup>. In this Article, we define the specific (external) exposome as all lifestyle (for example, substance use, diet and medical information) and psychosocial factors related to (social) stress and support (for example, personality, life events, education and occupation). Related to this domain, we know that adult well-being often has its developmental origins in childhood and adolescence, as indicated by associations with childhood psychopathology<sup>10,11</sup>. In adolescence and adulthood, psychosocial factors such as personality traits<sup>12</sup>, social support<sup>13,14</sup> and health indicators<sup>4</sup> appear to be important. In addition, many environmental factors are associated with mental health, including socio-economic status (SES), childhood maltreatment, substance use and life events<sup>15</sup>. The general (external) exposome is defined here as the built, external environment in which an individual is embedded. From this domain, objective neighborhood characteristics such as urbanicity, air pollution, greenspace availability and SES indicators (for example, income, social security beneficiaries and housing stock value) have recently been associated with well-being and traits like depression<sup>7,16–20</sup>. Further, a recent data-driven, environment-wide association study revealed the importance of neighborhood safety for well-being, even after correcting for SES at the individual and neighborhood level<sup>7</sup>. Since subjective reports of one's environment may be influenced by one's mental health status<sup>21</sup>, the current study focuses on objective exposure measurements. Another reason for this focus relates to well-being policies: identifying modifiable neighborhood characteristics associated with individuals' well-being allows for the formulation of targeted governmental policies. Finally, mental health traits (for example, depression, life satisfaction and positive affect) are partly driven by thousands of genetic variants with many small but relevant effects, many of which are shared across traits<sup>22–25</sup>. This is evidenced by high genetic correlations between depression, anxiety, personality traits and well-being, indicating a shared etiology<sup>26,27</sup>.

Aforementioned risk/protective factors are largely identified in cross-sectional studies associating well-being with one or more predictors following a pick-and-choose approach, providing group-level associations. Yet, finding an association with a certain risk/protective factor does not imply that it is useful to predict well-being at the individual level. Prediction, in this context, refers to the process of making forecasts on new, unseen data points based on patterns learned from a set of training data<sup>28</sup>. Associations may not translate to good predictions, for example due to low effect sizes or redundancy with respect to other more predictive variables<sup>29</sup>, but also given individual differences in pathways to well-being. Ultimately, personalized treatment and well-being interventions will depend on the ability to accurately predict at the individual level by pinpointing the risk/protective factors relevant to each individual.

The pick-and-choose approach further ignores well-being's complexity of genetic, childhood, psychosocial and environmental factors interacting with each other<sup>15,30</sup>, while research on, for example, gene-environment interplay is abundant<sup>15,31,32</sup>. Abdellaoui and colleagues<sup>33</sup>, for example, showed that polygenic scores (PGSs) of educational attainment, a trait linked with well-being, were associated with the SES of one's geographic location in the United Kingdom, as well as with social mobility (that is, migrating out of low SES regions). Genetic differences between people are thus correlated with environmental exposures, which may lead to differences in well-being levels.

Given its complexity, optimal well-being prediction will require broad inclusion of possible risk and protective factors from different data modalities; only analyzing them together will lead to the identification of the most relevant factors associated with well-being. In addition, optimal prediction requires the appropriate methods that can deal with the complex origins of well-being. Machine learning methods enable the inclusion of large numbers of variables ('features') from different data modalities, while accounting for their potential

nonlinear interactions, consistent with the consensus that well-being results from complex interactions between developmental, social, psychological, genetic and environmental factors.

Previous studies, using a wide variety of linear (for example, support vector machines (SVM) and linear regression) and nonlinear (for example, extreme gradient boost and random forest (RF)) machine learning algorithms have shown multimodal models to outperform unimodal models in the prediction of resilience and depression-related phenotypes<sup>29,34–38</sup>. However, these studies focused on mental illness rather than on well-being. A recent study on well-being specifically found that expanding the set of features increased prediction considerably<sup>4</sup>, but was based on a single data modality (self-reported psychosocial predictors). Studies including multiple data modalities are often limited to clinical samples (for example, ref. 29), that is, when treatment is already sought, limiting their external/ecological validity and practical usefulness, especially for prevention. Machine learning prediction studies in population samples have largely failed to take an integrative approach using cross-sectional data, while lacking environmental exposures and/or genetic data<sup>4,37–41</sup>. These limitations may explain why predictive accuracies have not reached the standards needed for clinical use<sup>42</sup>.

In the present study, we address some of previous studies' caveats by building prediction models for well-being using longitudinal objective environmental exposures (general exposome) and psychosocial factors (specific exposome), combined with genetic data (genome). We use data from a large population cohort, the Netherlands Twin Register (NTR), collected between 1991 and 2022, seven waves of the Young NTR (YNTR) collected around age 3, 5, 7, 10, 12, 14 and 16, and three adult waves collected in the Adult NTR (ANTR8 (2009–2012), ANTR10 (2013–2015) and ANTR14 (2019–2022); Table 1). We first train three machine learning algorithms, XGBoost (XGB), SVM and RF, and their predictions serve as inputs for a final XGB meta-model. The second aim is to identify the most predictive features of well-being using Shapley Additive Explanation (SHAP) values<sup>43</sup>, a method used to determine feature importance in predictive models. The outcome, well-being in adulthood, is based on multiple ratings of life satisfaction, happiness and quality of life. Our preprocessing and modeling pipeline is presented in Extended Data Figs. 1–3 and Supplementary Fig. 1.

## Results

### Unimodal analyses

The performance of the model including specific exposome predictors in the independent test set was high given conventional standards ( $R^2 = 0.702$ , 95% confidence interval (CI) [0.637–0.753]; Fig. 1). The genome showed not to be predictive of well-being ( $-0.007$  [ $-0.026$ – $0.010$ ]), while our initial general exposome model showed small but significant predictive power as indicated by the confidence interval not including zero (0.036 [ $0.011$ – $0.057$ ]). However, many general exposome features were characterized by distributions with many zeros (for example, the number of swimming pools within a 1 km radius usually is zero). As a sensitivity analysis, we therefore first dichotomized all general exposome features for which the mode was zero and subsequently re-ran our unimodal general exposome model. Given the considerably higher  $R^2$ -value of this model (0.047 [ $0.015$ – $0.076$ ]) and our aim of building optimal models, results based on transformed general exposome features are reported below (nontransformed results are in Supplementary Material 3).

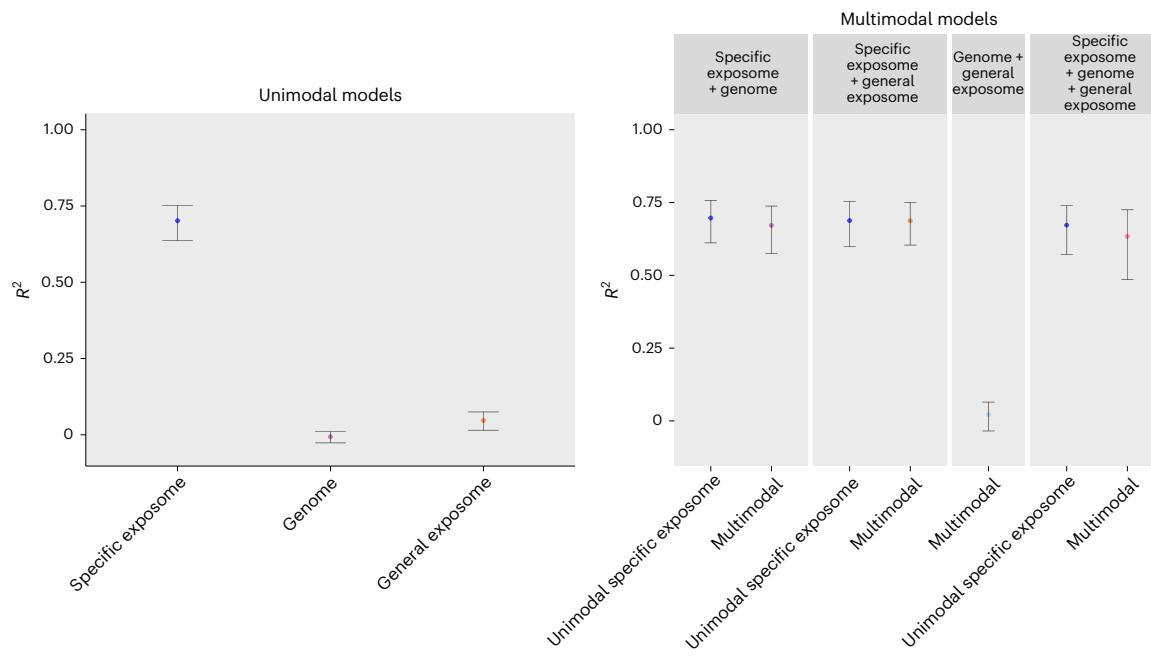
### Multimodal analyses

As reported in Fig. 1, the multimodal model including genetic and specific exposome predictors showed an  $R^2$  value of 0.671 [0.574–0.738], while the model including general and specific exposome predictors reached an  $R^2$  value of 0.688 [0.606–0.750]. In line with the unimodal results for the general exposome and genome, the combination of the environmental and genetic features showed little predictive power for well-being (0.022 [ $-0.034$ – $0.066$ ]). When all three data modalities

Table 1 | Demographic information

Dataset	N	%W	N <sub>train</sub>	N <sub>test</sub>	Gender	YNTR3	YNTR5	YNTR7	YNTR10	YNTR12	YNTR14	YNTR16	YNTR18	YNTR10	ANTR14	ANTR10	Age	Background <sup>a</sup>	
																		Dutch (%)	Non-Dutch (%)
Specific exposome	1,404	71 <sup>b</sup>	1,128	276	W	3.25 (0.22)	5.28 (0.31)	7.26 (0.24)	9.95 (0.3)	12.15 (0.37)	14.92 (0.59)	17.09 (0.57)	18.53 (0.74)	21.13 (2.11)	27.83 (2.3)	95	5	5	1,001
Genome	5,874	66 <sup>b</sup>	4,755	1119	M	[2.9-4.3]	[4.7-7]	[6.8-8.8]	[9.4-12.3]	[11.5-13.5]	[13.2-18.4]	[15.7-18.7]	[18-22]	[18-26]	[23-33]	95	5	5	
General exposome	3,859	66 <sup>c</sup>	3,055	804	W	3.29 (0.25)	5.61 (0.77)	7.43 (0.43)	10.05 (0.36)	12.14 (0.38)	15.07 (1.17)	17.24 (1.21)	21.57 (14.45)	34.17 (14.98)	38.27 (13.2)	95	5	5	
Specific exposome + genome	897	73	701	186	W	3.29 (0.23)	5.53 (0.68)	7.37 (0.39)	10.09 (0.36)	12.17 (0.4)	15.25 (1.41)	17.23 (1.18)	21.81 (9.19)	23.56 (7.97)	31.54 (9.37)	85	6	6	2,974
Specific + general exposome	1,080	71 <sup>b</sup>	868	212	W	3.25 (0.21)	5.29 (0.33)	7.26 (0.23)	9.95 (0.28)	12.15 (0.37)	14.91 (0.59)	17.09 (0.58)	18.54 (0.76)	21.32 (2.18)	27.97 (2.33)	95	5	5	2,569
Genome + general exposome	1,801	68	1,443	358	M	[2.9-4.3]	[4.7-7]	[6.8-8.8]	[9.4-11.1]	[11.5-13.5]	[13.2-18.4]	[15.7-18.7]	[18-22]	[18-26]	[23-33]	97	3	3	625
Specific exposome	3,28 (0.24)	5.56 (0.73)	7.4 (0.41)	10.05 (0.35)	12.13 (0.36)	15.11 (1.14)	17.19 (0.97)	19.98 (5.18)	22.8 (5.09)	29.64 (5.52)							94	5	818
Genome + general exposome	3,29 (0.24)	5.63 (0.8)	7.4 (0.44)	10.09 (0.36)	12.15 (0.35)	15.02 (0.85)	17.06 (0.8)	20.88 (7.39)	23.25 (6.88)	30.63 (7.93)	5	89	1,180						
Specific exposome + genome	3,25 (0.21)	5.29 (0.34)	7.27 (0.23)	9.95 (0.28)	12.14 (0.36)	14.9 (0.6)	17.12 (0.58)	18.52 (0.75)	21.36 (2.18)	28.21 (2.17)							96	3	522
General exposome	3,27 (0.21)	5.27 (0.27)	7.25 (0.22)	9.97 (0.3)	12.13 (0.35)	14.93 (0.68)	17.02 (0.54)	18.38 (0.65)	20.76 (2.18)	27.9 (2.24)	94	6							
Genome	3,29 (0.39)	[4.8-6.7]	[6.8-8.1]	[9.4-11.1]	[11.5-13]	[13.2-16.2]	[16.1-18.8]	[18-20]	[18-26]	[24-32]									
Specific exposome	3,25 (0.21)	5.29 (0.4)	7.27 (0.23)	9.95 (0.28)	12.14 (0.36)	14.9 (0.6)	17.12 (0.58)	18.52 (0.75)	21.36 (2.18)	28.21 (2.17)									
General exposome	3,27 (0.21)	5.27 (0.27)	7.25 (0.22)	9.97 (0.3)	12.13 (0.35)	14.93 (0.68)	17.02 (0.54)	18.38 (0.65)	20.76 (2.18)	27.9 (2.24)	94	6							
Genome	3,29 (0.39)	[4.8-6.7]	[6.8-8.1]	[9.4-11.1]	[11.5-13]	[13.2-16.2]	[16.1-18.8]	[18-20]	[18-26]	[24-32]									

Fam., families; W, women; M, men. The migration background of the participants was determined following the official definition of the Dutch statistics bureau: if at least one parent is born outside of the Netherlands, the background of the participant is 'non-Dutch'. <sup>a</sup>In case columns do not add to 100, remaining percentages are unknown. <sup>b</sup>One observation gender unknown.



**Fig. 1 | Model performance ( $R^2$ ) of unimodal and multimodal analyses.** Error bars represent 95% CIs based on the distribution of 10,000 family-wise bootstrap samples (lower bar given by the 2.5th percentile, upper bar given by the 97.5th percentile), with the  $R^2$  value obtained in the independent test set shown in the center.

were included, an  $R^2$  of 0.634 [0.490–0.728] was reached. We tested whether adding the genetic and general exposome beyond the specific exposome improved prediction by re-estimating unimodal specific exposome models in the smaller multimodal samples (Methods and Fig. 1) and comparing mean squared errors across models. Neither the addition of the genome ( $Z = 0.972, P = 0.334, \Delta_{MSE}$  (difference in mean squared errors) = 0.118 [−0.120, −0.356]), nor the general exposome ( $Z = 0.392, P = 0.695, \Delta_{MSE} = 0.045$  [−0.180, 0.270]), nor both ( $Z = 2.104, P = 0.029$ , not significant compared with our 0.005 threshold,  $\Delta_{MSE} = 0.289$  [0.020–0.559]), significantly increased model performance. In fact, model performance decreased somewhat when adding the other two data modalities beyond the specific exposome. However, in the model including the specific and general exposome, two general exposome features were in the top 15 most predictive features: the number of newly built social rent houses around age 12 was the third most predictive feature (SHAP), and household income of housing of benefit receivers was the 14th most predictive feature (permutation importance). Similarly, in the model based on all three data modalities, the number of newly built social rent houses around age 12 was the fifth most predictive feature (SHAP).

### Feature importance

**Unimodal specific exposome.** In the unimodal specific exposome model, the top 15 items from adulthood related to optimism, loneliness, personality (neuroticism and extraversion), subjective health, mental health traits (emptiness and worthlessness) and social relations and support (Fig. 2). One item was from an attention deficit hyperactivity disorder measure ('underachiever'). For both SHAP values and permutation importances, one top 15 item was measured in childhood, that is, parental exercise behavior around age 10. The remaining top items were measured in adulthood.

The bottom panel of Fig. 2 shows the SHAP values of the top 15 features in relation to their feature values, based on the optimal XGB model. While most features show a relatively linear association with well-being, others show nonlinear associations. An illustrative example is bor5\_10 ('feel empty'; a borderline personality scale item). Although high scores on this item are always predictive of lower well-being levels, there is a subgroup of people with high items scores for which the

influence is relatively small (SHAP values of approximately −0.07) and similar to those with intermediate scores on this item. For others, the influence is larger (SHAP values between −0.10 and −0.15). Similarly, for four individuals, agreeing with asr35\_14 ('worthless') has a very strong negative influence on well-being, while for others this relation is less strong. Some symptoms may thus contribute to well-being in different ways across individuals.

**Unimodal general exposome.** Figure 3 shows the top 15 (dichotomized when the mode was zero) general exposome features (top) and the top 15 SHAP values based on the optimal extreme gradient boosting model (bottom). Notably, most features are related to housing stock and the large share of top 15 features from childhood/adolescence (10 (67%) for SHAP values, and 7 (47%) for permutation importances).

The bottom panel of Fig. 3 clearly shows nonlinear patterns in the prediction of well-being for a number of general exposome features. For example, a low number of rented businesses premises in the area around age 16 (vh\_bedrijfsp\_y\_ntr16) has a strong positive contribution to well-being for some individuals, while for others the contribution is strong and negative. As another example, a higher percentage of public housing in the area measured at wave ANTR8 (p\_wcorpw\_antr8) generally has a very small negative effect on well-being, but for some individuals the influence is strong and positive.

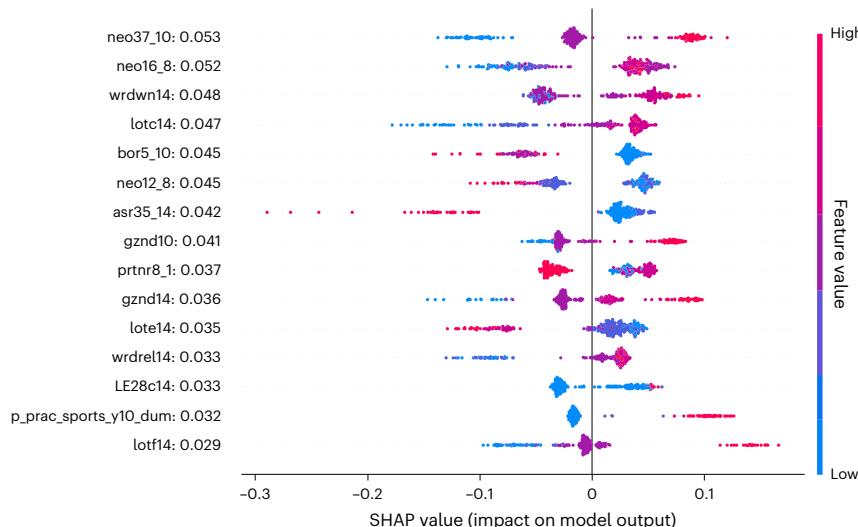
We refrain from reporting feature importance values for the unimodal genome model because of its nonsignificant predictive performance (see Supplementary Material 1).

### Longitudinal prediction based on the specific exposome

Single waves of specific exposome data from early childhood showed nonsignificant  $R^2$  values at age 3 (0.028 [−0.015–0.060]) and age 5 (0.045 [−0.018–0.094]), as shown in Fig. 4. Model performance at age 7 (0.081 [0.028–0.124]), age 10 (0.073 [−0.009–0.141]) and age 12 (0.082 [−0.006–0.154]) were highly similar. In adolescence, model performance increases steadily with time of measurement being closer to adulthood (YNTR14: 0.156 [0.069–0.228] and YNTR16: 0.229 [0.133–0.315]). Performance of single waves in adulthood ranged from 0.275 [0.169–0.369] for ANTR8, to 0.491 [0.413–0.556] for ANTR10, with ANTR14 in between (0.463 [0.340–0.562]). Across the ten single

Feature	Label	Domain
lotc14	Optimistic about future, ANTR14	Optism
neo37_10	Cheerful/vivacious, ANTR10	Psychological, personality
p_prac_sports_y10_dum	Practice any sports no/yes (parent), YNTR10	Exercise
neo16_8	Lonely/blue, ANTR8	Psychological, personality
neo2_8	Not cheerful/light-hearted, ANTR8	Psychological, personality
asr35_14	Worthless, ANTR14	Mental health, psychopathology
lote14	Rarely count on good things, ANTR14	Optism
wrdwn14	Rating living situation (0–10), ANTR14	Living situation, SES
lon1_10	Lack companionship, ANTR10	Loneliness
bor5_10	Feel empty, ANTR10	Mental health, borderline
gznd10	How would you rate your general health?, ANTR10	Health
gznd14	How would you rate your general health?, ANTR14	Health
lotf14	Expect more good than bad things, ANTR14	Optism
asr12_10	Feel lonely, ANTR10	Mental health, psychopathology
prtnr8	Steady relationship (yes/no), ANTR8	Social relations

Feature	Label	Domain
p_prac_sports_y10_dum	Practive any sports no/yes (parent), YNTR10	Exercise
lotc14	Optimistic about future, ANTR14	Optism
neo37_10	Cheerful/vivacious, ANTR10	Psychological, personality
neo12_8	Not cheerful/light-hearted, ANTR8	Psychological, personality
asr35_14	Worthless, ANTR14	Mental health, psychopathology
gznd14	How would you rate your general health?, ANTR14	Health
lote14	Rarely count on good things, ANTR14	Optism
neo16_8	Lonely/blue, ANTR8	Psychological, personality
lon1_10	Lack companionship, ANTR10	Loneliness
asr12_10	Feel lonely, ANTR10	Mental health, psychopathology
bor5_10	Feel empty, ANTR10	Mental health, borderline
lotf14	Expect more good than bad things, ANTR14	Optism
crs15_8	Underachiever, ANTR8	ADHD
prtnr8	Steady relationship (yes/no), ANTR8	Social relations
wrdrel14	Rating relationship status (0–10), ANTR14	Social relations



**Fig. 2 | Feature importance of the specific exposome.** Top: SHAP (top left) and permutation importance values (top right) of the top 15 specific exposome features across all three machine learning models (extreme gradient boosting, SVM and RF). Bottom: SHAP values of top 15 phenotypic specific exposome

features based on optimal extreme gradient boosting model. LE28c14, life event: had a child (1–5 years ago). Values shown on y axis are mean absolute SHAP values per feature.

wave models, model performance was positively associated with the total number of available features for each wave before ( $r(8) = 0.771$ ,  $P = 0.009$  [0.276–0.943]) and after feature selection ( $r(8) = 0.732$ ,  $P = 0.016$  [0.191–0.932]). Sensitivity analyses showed, however, that the increased prediction by more proximal features was not solely due to the sheer number of features (Supplementary Material 2).

$R^2$  values of models based on all features from childhood/adolescence and adulthood separately, and their combination, are presented in Fig. 5. The model based on only childhood and adolescence features reached an  $R^2$  of 0.268 [0.193–0.333], while the performance of the model including only (all) adulthood waves (0.629 [0.517–0.712]) was significantly higher ( $Z = -6.219$ ,  $P < 0.001$ ,  $\Delta_{MSE} = -0.745$  [−0.980–−0.510]). The difference in performance between the model based on solely adulthood features and the full model based on all childhood/adolescence and adulthood features combined (0.702) was not significant ( $Z = 0.892$ ,  $P = 0.372$ ,  $\Delta_{MSE} = -0.089$  [−0.283–0.106]).

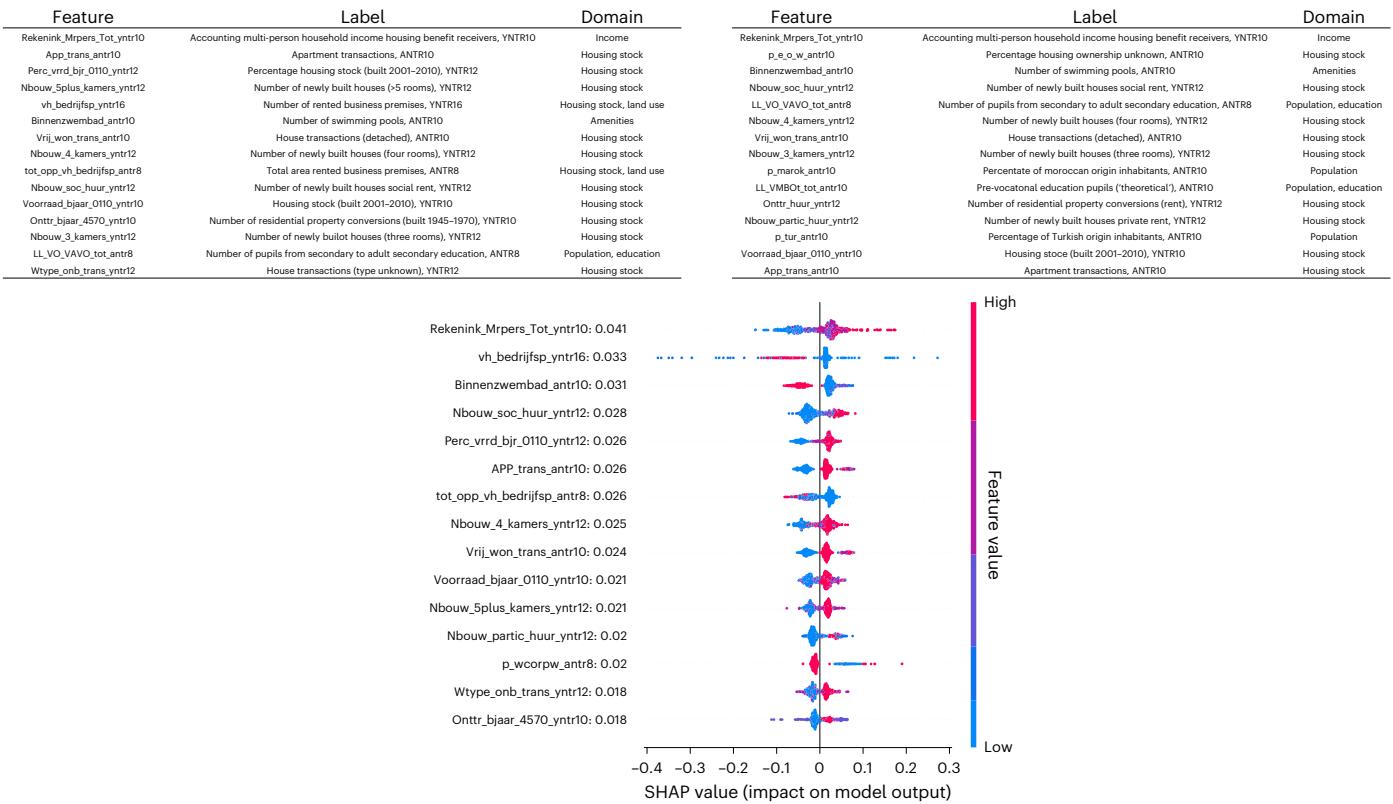
Note that model performance of all childhood and adolescence waves (0.268) was only marginally higher than the model based on YNTR16 alone (0.229;  $Z = -1.936$ ,  $P = 0.051$ ,  $\Delta_{MSE} = -0.194$  [−0.390–0.002]). In addition, in the model based on all childhood and adolescent wave features, 7 (SHAP) and 8 (permutation importance) out of 15 top features were from YNTR16, suggesting that this prediction was dominated by YNTR16 features. Thus, the predictive power of the ‘all childhood/adolescence waves features’ model was largely due to YNTR16 features.

## Discussion

Using ten data waves from a large population cohort collected between 1991 and 2022, machine learning models were trained on the basis of an

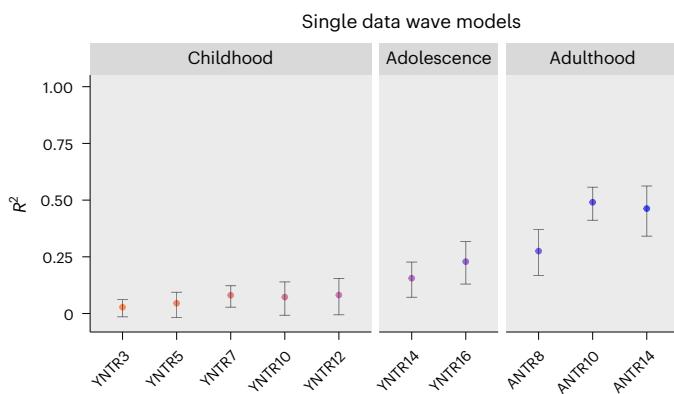
extensive set of predictors from the genome (PGSs), the specific exposome (longitudinal psychosocial predictors) and general exposome (longitudinal objectively measured environmental exposures) to predict well-being in adulthood. In our unimodal specific exposome models, the large amount of predictors from childhood to adulthood resulted in prediction of well-being with high accuracy ( $R^2 = 0.702$ ). The most important features were optimism, loneliness, personality traits, other mental health traits and social relations and support. The general exposome showed a modest predictive value for well-being ( $R^2 = 0.047$ ) on its own, but adding its features to the specific exposome did not increase prediction. PGSs for traits from different domains were not significantly predictive of well-being, neither independently nor incrementally predictive beyond the specific exposome. Results were robust to several sensitivity analyses with respect to the number of outcome measurements, outliers, number of features and feature-to-sample ratio (Supplementary Material 2). Importantly, the predictiveness of the specific exposome compared with the genome and general exposome was not solely due to differences in feature numbers (2615, 60 and 732, respectively). Relatedly, feature-to-sample ratios may have been partly responsible for the fact that adding the genome and general exposome beyond the specific exposome decreased performance somewhat, but sensitivity analyses showed this not to be the sole reason for this finding (Supplementary Material 2).

Our specific exposome model performance was highly similar to a recent study reporting  $R^2$  values between 0.70 and 0.80 across models for the prediction of well-being over one year<sup>41</sup>. Yet, it is important to highlight how our study differed from Oparina et al. (2022), who obtained much lower  $R^2$  values (about half) while following a similar data-driven approach. First, we included a much larger number of

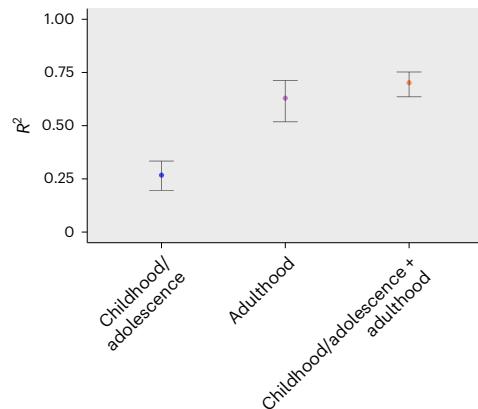


**Fig. 3 | Feature importance of the general exposome.** Top: SHAP (left) and permutation importance values (right) of top 15 general exposome features across all three machine learning models (extreme gradient boosting, SVM and RF). Bottom: SHAP values of top 15 phenotypic specific exposome features based

on an optimal extreme gradient boosting model. p\_wcorpw\_antr8, percentage public housing in ANTR8. Values shown on y axis are mean absolute SHAP values per feature.



**Fig. 4 | Model performance ( $R^2$ ) based on single waves of phenotypic (specific exposome) data.** Error bars represent 95% CIs based on the distribution of 10,000 family-wise bootstrap samples (lower bar given by the 2.5th percentile, upper bar by the 97.5th percentile), with the  $R^2$  value obtained in the independent test set shown in the center.



**Fig. 5 | Model performance ( $R^2$ ) based on the type of longitudinal phenotypic (specific exposome) data.** Error bars represent 95% CIs based on the distribution of 10,000 family-wise bootstrap samples (lower bar given by the 2.5th percentile, upper bar by the 97.5th percentile), with the  $R^2$  value obtained in the independent test set shown in the center.

potential predictors, including longitudinal predictors from childhood and adolescence, and from multiple adult waves. Longitudinal analyses showed that although proximal features were more predictive, the addition of distal features from childhood and adolescence increased prediction. In line with this, about half of the predictors (43%) selected by the feature selection algorithm were from childhood and adolescent waves. Second, our feature set consisted of high-quality features including personality and other mental health conditions (for example, anxiety and depression) while Oparina et al. (2022) largely included

economic predictors. Yet, our models excluding mental health features (as in Oparina et al. 2022) also performed well (Supplementary Material 2), owing to the quantity and quality of the remaining features. A final important distinction is our use of a well-being factor score rather than a single item measure, reducing measurement error and improving model performance. Notably, when comparing results based on single waves of adult data, our findings were in line with Oparina et al. (2022), suggesting robustness of our results.

The environmental exposures showed modest predictive power on their own. In our optimal model—in which features with a mode of zero were dichotomized—half of the top features were measured in childhood/adolescence and predominantly related to housing stock. In the Netherlands, the composition of houses in a neighborhood (for example, public/social versus private) is an important indicator of neighborhood SES<sup>44,45</sup>, with public housing being more concentrated in cities<sup>46</sup>. Our results are in line with studies finding associations between (childhood) housing characteristics and important life outcomes, including well-being<sup>47–49</sup>. In our models including nondichotomized features, feature domains were more diverse (for example, income, population composition, education and amenities), although most indicators were (in)directly related to neighborhood SES and/or urbanization. Some of them were identified as risk/protective factors of mental health (for example, depression, anxiety and psychological distress) in previous studies<sup>16,17</sup>. Previously advocated risk/protective factors for well-being such as air pollution and greenspace<sup>20,50</sup>, although included in our feature set, did not appear to be predictive. Our results underline important differences in what it means to predict (versus associate); these exposures typically show small effect sizes when studied in isolation, and when many other exposures are included their effects may be further depreciated, limiting their collective predictive power.

Drawbacks related to objective environmental exposures linked to participants' postal codes may have reduced their predictive power. For some exposures, the spatial resolution may not be fine-grained enough; for air pollution, the street address or even the floor level may be most informative<sup>51</sup>. Our approach also assumes uniform exposure within the same postal code, ignoring that people are exposed to environmental pressures when they commute, are at work, go to gym and meet with friends. Future research should explore ecological momentary assessment with passive global positioning system tracking on smartphones to capture the dynamic relationship between the environment and well-being<sup>52</sup>.

The selected PGSSs were for traits previously (genetically) associated with well-being (for example, childhood maltreatment, resilience and loneliness), but also for more distal traits (for example, circadian rhythm and smoking cessation). Yet, collectively, the selected PGSSs showed limited predictive power in the independent test set, which may be due to several factors. First, PGSSs are based on genome-wide association studies (GWAS) that only capture the tiny effects of a restricted set of common variants; by increasing the variant coverage of PGSSs and GWAS sample sizes, out-of-sample prediction may increase<sup>53–55</sup>. Relatedly, our genetic sample ( $N = 5,874$ ) was perhaps still too limited for the PGSSs to become predictive out-of-sample. With respect to the number of PGSSs, we included a large number (60) covering multiple domains, but more extensive sets may be needed. As genotyping costs decrease, hopefully it will soon become possible to create PGSSs for even larger numbers of traits and individuals, rendering the combined predictive power of PGSSs for well-being high enough to be practically useful.

The findings of the current study can be used to formulate hypotheses on nonlinear pathways to well-being. Our SHAP values showed that the relation between some items and environmental exposures on the one hand and well-being on the other differed across groups of participants. If replicated, a next logical step would be to identify possible moderators of the effects of these items/symptoms. This can for example be achieved by associating individuals' SHAP values with external variables such as gender and age, or other traits such as depression or educational level. For the environmental exposures, it would entail in-depth analyses as to why some aspects of the built environment (for example, low number of rented business premises in adolescence) affect well-being positively for some and negatively for others. The geographical context will have a large influence on this<sup>48</sup>: naturally, both neighborhoods with low and high SES, or highly urbanized or rural areas, can have a low numbers of rented businesses. In the

end, risk stratifications based on the exposome may pave the way for personalized prevention and treatment options.

Our results have important implications both for future investigations of well-being and the practical utility of well-being prediction models. In line with well-being's complexity, predictive features included early childhood circumstances and behavior, SES, substance use, personality, life events, psychopathology, and lifestyle behaviors (Supplementary Table 1). Future studies should embrace this complexity, avoiding a narrow focus on individual risk/protective factors. Several highly predictive features align with prior research (for example, social support, personality and self-rated health)<sup>4,41</sup>, suggesting potential targets for interventions. The current study also shows how a data-driven approach can generate new hypotheses on the importance of features. For example, parental exercise behavior around age 10 was highly predictive of well-being later in life. Future research should investigate through which pathways this may occur, for example, through social learning<sup>56</sup> or health communication<sup>57</sup>. Further, the importance of consistent long-term monitoring is signified by multiple wave models outperforming single data wave models. Importantly, single waves of data were predictive of adult well-being from as young as age 7 onward. However, we did find proximal features to be more predictive than more distal features. This may be partly due to changes in rater effects, since childhood predictors were based on parent ratings and adolescence/adult ratings on self-reports. Additionally, the lack of incremental prediction of childhood predictors beyond adult predictors may be due to downstream effects of childhood differences. By controlling for those effects by including proximal adulthood features, inclusion of the childhood features would not contribute significantly to the model anymore. Still, the feature selection algorithm did choose (early) childhood features, and inclusion of more features from multiple waves childhood/adolescence boosted prediction considerably, underlining the importance of longitudinal information.

From a policy perspective, our general exposome results highlight the importance of initiatives targeted at housing. Although previous studies have shown associations between housing characteristics and well-being<sup>49</sup>, the current study shows that neighborhood housing stock in adulthood but also in childhood/adolescence can predict adult well-being at the individual level. These results are especially relevant given the current housing crisis in the Netherlands with long waiting lists for public housing for lower income residents, and surging housing prices in the private sector<sup>49,58</sup>. Our results suggest the importance of (local) governments prioritizing affordable housing, as it may have direct or downstream consequences on individuals' well-being.

Some limitations of this study are worth mentioning. First, given our data-driven approach, we decided to only remove direct indicators of well-being from our feature set, and no other features based on conceptual or statistical considerations. This decision led to the inclusion of features that conceptually overlap with well-being (for example, optimism and self-esteem), perhaps somewhat artificially boosting the performance of our specific exposome models. Yet, a model excluding all mental health features altogether reached similar  $R^2$  values (Supplementary Material 2). Second, in each wave, only two or three hedonic well-being measures were available to identify well-being at that time point. Future studies should include more indicators, preferably including eudaimonic measures, to improve identifiability of the used factor model. Moreover, our results should be interpreted in relation to hedonic (subjective) well-being; although strongly (genetically) associated with eudaimonic and social well-being<sup>1,59,60</sup>, it is not certain that our results generalize to these well-being concepts. Third, our results are restricted to a Western Educated Industrialized Rich Democratic sample, limiting generalizability to other countries or contexts. Our general exposome results should also be interpreted in relation to the specific Dutch context (for example, high population density and high well-being), which may have influenced general exposome–well-being relationships<sup>17</sup>. For example, air pollution levels

are comparatively low, and because of the Netherlands' small size, regional differences are small<sup>61,62</sup>. In addition, many different, possibly opposing or indirect effects at different geographical levels may be operating simultaneously; neighborhoods' housing prices are typically associated with improved mental health, but such neighborhoods are often found in Dutch urban areas associated with reduced mental health<sup>16,17</sup>. Similarly, in the Dutch context, blue space has been associated with reduced mental health due to the larger cities having many canals<sup>17</sup>. Together, this may have resulted in low predictive power of the general exposome at the individual level. Our study should thus be replicated in other countries, which differ in terms of, for example, population decomposition, neighborhood characteristics and urbanization grades. Fourth, women were overrepresented and people with a migration background underrepresented in our sample compared with the Dutch population. Fifth, due to attrition, our predictive models were applied to somewhat less happy and higher educated individuals compared with those who dropped out of the study, although differences were small. This may have affected our results if pathways to well-being differ across these groups. Further, our results may not generalize to clinical samples. External validation of our algorithms in external, longitudinal and more heterogeneous cohorts in terms of outcomes, predictors and demographics (gender and migration background) is thus needed. Finally, although we used longitudinal data, our results are still based on associations; causal interpretations of our findings are therefore unwarranted. Given that we included indicators associated with well-being in childhood and adolescence, bidirectional patterns and reverse causality may be responsible for adult well-being. For optimal prediction, the aim of this study, this is less of an issue given that the task is to predict individuals' well-being, regardless of its origins. For understanding of how individuals develop high or low well-being levels, however, causality is more important; although our SHAP values analyses provide some insights, novel causal machine learning<sup>63</sup> approaches are needed in the future.

To conclude, based on extensive, longitudinal assessments of the exposome, we are able to predict adult well-being levels with high and modest accuracy based on the specific and general exposome, respectively. Currently, there is no incremental prediction of general exposome and genome beyond the specific exposome, but the future may hold more promising results as more sophisticated methods for their assessment are rapidly developing. Combining them with the internal exposome (for example, microbiome and metabolome) may pave the way to personalized interventions based on one's genetic background and exposome.

## Methods

### Sample

Data from ten study waves of the NTR were used, collected between 1991 and 2022. In the NTR, every 2/3 years, longitudinal surveys about lifestyle, personality, psychopathology and well-being in twins and their families are collected<sup>64</sup>. The NTR is a population-wide, nonclinical sample and distinguishes between the YNTR and the ANTR. The YNTR comprises children rated by their parents, and standardized surveys are sent out around the ages of 1, 2, 3, 5, 7, 9 and 12. For several years (2004–2014), YNTR twins aged 14, 16 and 18 were invited for self-report surveys. When participants in the YNTR reach the age of 18, they are invited to take part in the ANTR in which adult participants provide self-reports.

Data from the YNTR waves 3, 5, 7, 10, 12, 14 and 16 were used because these included the most consistently administered variables. We use a fully data-driven approach: features are included on the basis of availability, not on substantive considerations. Only mother reports were included, because mothers were most compliant in responding, thus maximizing sample size. Additionally, aggregated NTR variables were included as these include important variables for the prediction of well-being (for example, parental SES). For our outcome (well-being), we used ANTR waves 8, 10 and 14 based on the availability of relevant

well-being variables. Signed informed consent, also for record linkage, was obtained from all individual participants included in the study. Participants did receive compensation for participation. For more information on data collection within the NTR, see refs. 64–67. Data was obtained by completing the data sharing request forms and approval of the NTR Data Access Committee.

In total three unimodal (specific exposome, genome and general exposome) and four multimodal (specific exposome/genome, specific/general exposome, genome/general exposome and all three) datasets were created. Demographic information on gender, age and migration background for all datasets are reported in Table 1. In childhood, gender was assessed by the parents, and in adolescence and in adulthood through a self-report question. Since biological information was not used to determine sex, the term 'gender' is relevant for this study. The migration background of the participant was determined following the official definition of the Dutch statistics bureau: if at least one parent is born outside of the Netherlands, the background of the participant is 'non-Dutch'. In the majority of the cases (93% for mothers and 96% for fathers), parents' birth countries were assessed through self-reports, otherwise information was based on Dutch census statistics. Because we did not use a self-identification question on ethnicity, we do not report the specific ethnic backgrounds of the non-Dutch participants. In the unimodal specific exposome sample, the mean number of study waves present was 8.96 (standard deviation (s.d.) 0.73), ranging between 6 and 10. Most people had data on nine study waves (53% of the sample). Across all waves, the average years between consecutive waves was 2.78 (s.d. 1.37), with a minimum average years of 1.60 (s.d. 0.96, range 0–5 years) between YNTR16 and ANTR8, and maximum average years of 6.17 (s.d. 0.43, range 5–8 years) between ANTR10 and ANTR14. A small number of participants completed two surveys within the same year; the largest number (51 or 3.6%) was found for the YNTR16 and ANTR8. In the unimodal general exposome sample, neighborhoods were spread all over the Netherlands (Supplementary Fig. 2). Across waves, the number of different six-digit postal codes ranged from 287 (YNTR3) to 2,511 (ANTR8). The mean number of participants per neighborhood (four-digit postal codes) across waves was 1.39 (s.d. 0.54), ranging between 1 and 15. Multimodal datasets were created by merging the preprocessed unimodal datasets.

Attrition analyses showed that the participants included (versus not included) in the analyses reported somewhat lower levels of well-being and were more highly educated, although effect sizes were small (Supplementary Table 2). Gender differences in attrition were negligible.

### Measures

**Outcome: well-being.** A continuous well-being score was created on the basis of the Satisfaction with Life Scale<sup>68</sup>, Subjective Happiness Scale<sup>69</sup> and a measure of quality of life, the Cantril Ladder<sup>70</sup>. The Satisfaction with Life Scale consists of five items answered on a seven-point scale (1 = 'strongly disagree' to 7 = 'strongly agree'). An example item is: 'I am satisfied with my life'. The Subjective Happiness Scale consists of four items rated on a seven-point scale (1 = 'strongly disagree' to 7 = 'strongly agree'). An example item is 'On the whole I am a happy person'. The Cantril Ladder requires participants to indicate the step on a ladder at which they place their lives in general on a ten-point scale (10, best possible life, and 1, the worst possible life).

We fitted a latent trait-state-occasion model<sup>71</sup> using structural equation modeling (in lavaan<sup>72</sup>) to compute an overall adult well-being factor score based on all available well-being measures from ANTR8, ANTR10 and ANTR14, for each respondent. For factor score estimation, the regression method was used with a transformation so that the covariance matrix of the estimated factor scores matched the model-based latent factor covariance matrix<sup>73,74</sup>. The outcome represents individuals' stable, overall well-being levels while controlling for time-specific states and age at time of measurement (by regressing out age at each wave within the model; Supplementary Fig. 3). In this model,

all data are preserved: even those with missing values on any of the outcomes will receive a general well-being score. Our outcome variable was restricted to adults ( $\geq 18$  years). Participants were included if they had two or more well-being measure available (either within or across waves). To prevent ‘data leakage’ known to affect model performance<sup>75</sup>, the latent trait–state–occasion model was fitted in the training set and the model parameters based on this set were used to create factor scores in both the training and test set. Model fit (degrees of freedom, 4;  $\chi^2$ , 488.836;  $P < 0.001$ ), Comparative Fit Index 0.985, Tucker-Lewis Index 0.976, Root Mean Squared Error of Approximation 0.028, Standardized Root Mean Squared Residual 0.029) was good according to common standards<sup>76</sup>. Correlations between the overall well-being score and its constituent traits were high, between 0.694 (quality of life, ANTR14) and 0.796 (satisfaction with life, ANTR14); Supplementary Fig. 4.

**Features.** For full lists of initially considered features, included features after preprocessing and feature selection, see Supplementary Table 1.

**Phenotypic data (specific exposome).** We included individual items/symptoms of scales rather than sum scores in our models (for example, individual items of the NEO Five Factor Inventory for extraversion rather than the extraversion scale score). Some summary measures (within measurement wave) based on count variables were calculated for this study’s purposes (for example, calculating ‘total conditions present’ from individual medical conditions).

**Genetic data (genome).** Genetic predictors were included through PGSSs. PGSSs are aggregate scores of an individual’s genetic predisposition for a given trait, calculated by summing risk alleles weighted by their effect sizes from discovery GWAS<sup>77,78</sup>. Given our data-driven approach, PGSSs for all traits readily available in the NTR were included. Supplementary Table 3 presents a list of the 60 PGSSs included in this study and their original GWASs. The PGSSs covered traits from a wide range of domains (for example, personality, childhood health, childhood psychopathology, substance use, SES and exercise behavior).

PGSSs were calculated in the standardized pipeline within the NTR. In short, variants for which the effect allele frequency was  $0.01 \leq$  effect allele frequency  $\leq 0.99$  are retained. The allele frequencies and effect sizes of the variants are aligned with the NTR reference for the 1,000 genomes variants. Variants not included in this reference were excluded. The processed summary statistics serve as input for the LDpred 0.9 software. For estimating the target linkage disequilibrium structure a (1) selection of unrelated individuals in the NTR sample are used and (2) a set of well-imputed variants in the NTR sample are selected. The parameter `ld_radius` is set by dividing the number of variants in common (from the coordination step output) by 12,000. For the coordination step, the median sample size is used as the input value for  $N$ . The plink2 software package generates the PGSSs (using the `-score` option) to the input weighted effect sizes and the genotype dataset. We include PGSSs based on the infinitesimal model, that is, including all SNPs. In all models including PGSSs, as covariates the first ten genomic principal components to control for population structure and genotyping platform dummies were included.

**Environmental exposures (general exposome).** For the general exposome, we leveraged the Geoscience and health Cohort Consortium database<sup>79</sup>, which has previously been used in combination with the NTR<sup>7</sup> and in other studies on mental health<sup>17</sup>. This database includes a wide range of objective environmental exposures at the neighborhood (postal code) level (for example, liveability, safety, population composition, air pollution and neighborhood SES indicators) based on several data sources<sup>80</sup>. A full list of available exposures can be found elsewhere (<https://www.gecco.nl/exposure-data-1/>), with more information on the included features is reported in Supplementary Tables 1 and 4. All available exposures for the time period 1990–2022 were requested

through the data access form. Geoscience and health Cohort Consortium database includes measurements of exposures per year, and for each wave of data collection within the NTR, the postal code of the respondent was available. When participants register for the NTR, they are asked to provide their address. Participants are asked to contact the NTR when they move so that their address can be updated. In addition, addresses are regularly updated by cross-checking with the Dutch Personal Records Database (in Dutch: Basis Registratie Personen). For some participants, postal code information may not be accurate if they move without notification. We checked this in a subset of participants for which self-report postal codes (four digits) were available in the child and adolescent waves: in ~90% of the cases (range 88–94%), the postal codes in the database matched with self-reported values. For each individual, the postal code at the year of survey completion was matched with the exposure of that year. Exposures were linked to the NTR data either using participants’ six- or four-digit postal code depending on which level a given feature was available. For features with different spatial resolutions available (for example, number of supermarkets within a 500 m, 1,000 m or 1,500 m radius), we chose a resolution of 1,000 m (for example, ref. 18).

## Analyses

The reported analyses are part of a larger preregistration (<https://osf.io/msbvp>), for additions and deviations see <https://osf.io/tuhgs>. Before any analyses were done, the dataset was divided into two independent datasets in an 80/20% split, the former ( $N_{train}$ ) being used for feature selection, training and validation using tenfolds cross-validation, while the second withheld dataset ( $N_{test}$ ) was used for final model testing (Supplementary Fig. 1). The NTR includes family data with relatives being non-independent observations. Therefore, members of the same family were not split across training and test set in order to prevent overfitting that could occur when they are separated between datasets. In addition, grouped tenfolds cross-validation was used in training to take relations between family members into account.

We ran all machine learning models for each unimodal dataset, all possible pairwise combinations, and a model including all three data modalities, resulting in a total of seven sets of analyses. For each set, we maximized the sample size, leading to different sample sizes across analyses (Table 1). However, in the multimodal analyses, we compared the multimodal model performance (for example, specific exposome + genome) with the performance of a unimodal specific exposome model based within this same (reduced) sample for optimal comparability.

The feature selection step was only applied in the unimodal analyses: the multimodal analyses were based on the features that were selected in each of the three unimodal analyses. This made the unimodal and multimodal models readily comparable: if we would have entered all three data modalities into the feature selection algorithm, then only the phenotypic features would have been selected since the predictive power of PGSSs and objective environmental exposures is much lower compared with the phenotypic features. In addition, the unimodal analyses were based on larger sample sizes than the multimodal analyses ensuring reliability and generalizability of the features.

The availability of longitudinal specific exposome data allowed predicting adult well-being from previous time points by training and testing models based on different slices of features. In the unimodal specific exposome dataset, we first trained models based on features from each wave separately (Fig. 4), on features only from childhood/adolescence (that is, YNTR3 until YNTR16), and on features only from adulthood (all three adult waves jointly), and compared them with the full phenotypic model including all childhood/adolescence and adulthood features described earlier (Fig. 5).

**Preprocessing.** All preprocessing steps were informed on the training data and based on this the test data were preprocessed to avoid data leakage.

**Removing and transforming features.** Features were screened for nonsensible values, and if present, set to missing. Features with zero variance and character (that is, text) features were removed. All twin-specific features were removed since we aimed at building prediction models for the general population. All direct well-being (that is, outcome) indicators and features including the words ‘happy’ or ‘happiness’ were excluded from the adult surveys, as this would overlap too much with the outcome. Models including direct well-being indicators in childhood and adolescence are reported in Supplementary Material 3. All continuous variables were first standardized and subsequently normalized (range 0–1), while categorical variables were dummy encoded.

As reported elsewhere<sup>7,17</sup>, the general exposome features showed strong correlations with each other and many were highly skewed. In our main analyses, in our general exposome dataset, we iteratively removed variables with correlations >0.95 with other exposures until no more of these features remained. Using the e1071 (ref. 81) package, all general exposome features were transformed by cubing them (skewness <1) or taking their cube root (skewness values >1).

**Missing value imputation.** The postal codes of the participants showed a relatively large number of missing values. To increase the general exposome sample size, we imputed missing values as follows. For many individuals, gaps in available postal codes between waves were found, even though no changes occurred in recorded postal codes. As an example, postal code 1111AA would be recorded for an individual at YNTR5, then two missing values for waves YNTR7 and YNTR10, and then a record of postal code 1111AA at wave YNTR12. In that case, we assumed the individual did not move in between waves and missing values were imputed with the recorded values (that is, 1111AA). For the year of exposures to link the imputed postal codes to, linear interpolation was used: in the previous example, if YNTR5 was completed in 2002, and YNTR12 in 2008, then exposures were selected from 2004 (YNTR7) and 2006 (YNTR10). Given the high stability of objective environmental exposures over time<sup>17</sup>, this linear interpolation scheme can be expected to give reliable results.

In the unimodal datasets, participants and features with more than 55% missing values were excluded (deviating from our preregistered 50% as this would have reduced our sample sizes too much). Remaining missing values were imputed using the *k*-nearest neighbors<sup>82</sup> method, a relatively simple nonparametric multivariate imputation technique more time efficient than more complex imputation techniques<sup>83</sup>. Following common practice, we chose *k* by taking the square root of the number of observations in each of the respective datasets.

**Feature selection.** Because of the relatively large number of features compared with the number of individuals, feature selection was applied to reduce the number of features before prediction. We applied elastic net regression<sup>84</sup> (with tenfolds cross-validated tuning on the training set), a combination of ridge and Least Absolute Shrinkage and Selection Operator (LASSO) regression, previously shown to provide more stable results than either methods separately<sup>85</sup>.

Out of 2,615 specific exposome features, 212 features from several domains (Supplementary Table 1) were selected. These included 91 (43%) features from childhood and adolescence (each wave contributing at least five features), the remaining 121 (57%) being measured in adulthood (Supplementary Table 5). For the genome, 13 out of 60 features were selected (PGSs for agreeableness, asthma, childhood bodymass index, childhood maltreatment, circadian rhythm, educational attainment, household income, loneliness, moderate to vigorous physical activity, pubertal growth, resilience, smoking cessation and well-being). Finally, for the general exposome, 29 out of 732 exposures were selected in our dichotomized feature models (Results), including 15 features (52%) from adolescence (3 from YNTR10, 9 from YNTR12, 2 from YNTR14 and 1 from YNTR16). These features mostly related to housing stock (for example, number of newly built houses, house transactions, number and area of

rented business premises, and percentage public housing), but also to population decomposition (for example, number of education pupils, percentage of inhabitants with non-western migration backgrounds), amenities (for example, kernel density local food shops in 1,000 m radius), and income (multi-person household income of housing benefit receivers). In our nondichotomized general exposome model, 36 features were selected with 18 of them (50%) from adolescence (4 from YNTR10, 8 from YNTR12, 2 from YNTR14 and 4 from YNTR16). These exposures related to housing stock (for example, apartment transactions), population decomposition (for example, percentage of inhabitants with non-western migration background and absolute number of single-person households), neighborhood statistics (for example, mortality and divorce rates), urbanization grade, amenities and neighborhood SES indicators (number of social security beneficiaries and percentage of inhabitants with a high (top 20%) income).

**Model training.** In line with previous studies<sup>34,37</sup>, we created a stacked ensemble model based on commonly used XGB, RF and SVM models. Stacked ensemble models train a second-level meta-learner based on the first-level (that is, XGB, RF and SVM) model predictions to arrive at an overall prediction. As each algorithm has its own advantages and limitations, the overall meta-learner tends to outperform each individual model by combining the inputs into a best guess<sup>34,40,86</sup>. For the final, second-level model we used the XGB model as it tends to perform well when predicting mental health<sup>4,29</sup>. Standard hyperparameters were tuned in a random grid search with 100 searches for each model, with grid ranges largely determined by previous studies<sup>34,36,37,87</sup>. Searched parameter grids and optimal (that is, tuned) combinations are reported in Supplementary Table 6. A standard ordinary least squares (OLS) regression based on the feature selected set was conducted as a baseline comparison.

**Model evaluation. Performance.** To evaluate model performance, the  $R^2$  of the optimal model (Supplementary Table 5) in the independent test set is reported. This metric was chosen for comparability with previous work<sup>4</sup>. Note that negative  $R^2$  values can be obtained if the model performs worse than simply predicting the outcome mean—typically a sign of overfitting in the training data. To compare performance across models, 95% CIs were estimated using nonparametric bootstrapping at the family level<sup>188</sup> (10,000 samples).  $R^2$  values are deemed significant if the confidence intervals do not include zero. Nonparametric clustered Wilcoxon signed rank tests<sup>89,90</sup> on the squared model prediction errors (that is,  $(y_{\text{pred}} - y_{\text{test}})^2$ ) were used to test whether the performance of two models (unimodal versus multimodal) differed significantly from each other. Throughout all analyses, all *P* values reported are based on two-tailed tests, using a conservative *P* value of 0.005 (ref. 91).

**Feature importance.** To derive meaning from our trained models, feature importances for each of the lower-level models were investigated using SHAP<sup>43</sup> values. SHAP allows for the estimation of each feature’s importance for each individual prediction (that is, participant), by comparing model performance with and without each feature. The overall importance of each feature can then be calculated by taking the mean of the absolute SHAP values across all individual predictions. Features with the highest SHAP values (indicating higher importance) are identified and the top 15 are reported. Since we trained three different models (XGB, RF and SVM), we report the top 15 based on the mean feature importance value across these three models. Because different feature importance methods have their own strengths, we also report permutation importances, which measure the degree to which each algorithm relied on a given variable in making its predictions<sup>4</sup>.

All preprocessing was done in the R environment<sup>92</sup>. All machine learning analyses were conducted in Python using PyCharm in combination with Anaconda3 as the user interface; the RF and SVM models were conducted using the scikit-learn package<sup>93</sup> and the extreme gradient boosting models were conducted using xgboost<sup>94</sup>. The TRIPOD

Checklist for Prediction Model Development is reported in Supplementary Table 7.

### Ethics statement

All procedures conducted in studies involving human participants adhered to the ethical standards of the institutional and/or national research committee, as well as the 1964 Helsinki Declaration. Data collection was approved by the Central Ethics Committee on Research Involving Human Subjects of the University Medical Centers Amsterdam. Informed consent was obtained from all individual participants included in the study. Only participants who consented to record linkage were included. Ethical approval numbers are as follows: YNTR3/YNTR5/YNTR7/YNTR10/YNTR12 (94/105, 21-06-1994; 96/205, 14-01-1997; 99/068, 11-08-1999; 2003/182, 18-12-2003; and 2010/359, 18-02-2011), YNTR14/YNTR16 (2003/182, 18-12-2003; 16-12-2010; 12-09-2012; and 30-10-2013), ANTR8 (NL25220.029.08/2008-244), ANTR10 (2011/334; 12-10-2011, 2012/433; and 26-02-2013), ANTR14 (2018/389; 25-07-2018, VCWE-2018-124; and 16-08-2018) and genetic data (04.001.98; 25-05-2007).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Being part of a national prospective cohort study, the Netherlands Twin Register data cannot be made publicly available for privacy reasons, but they are available for legitimate researchers via the data access procedure at [https://tweelingenregister.vu.nl/information\\_for\\_researchers/working-with-ntr-data](https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data). Data of the Geoscience and health cohort consortium (GECCO) can be requested via the data access request form at <https://www.gecco.nl/exposure-data-1/>.

### Code availability

Python scripts for the machine learning models can be found at <https://osf.io/zphw8/>.

### References

- Keyes, C. L. M. The mental health continuum: from languishing to flourishing in life. *J. Health Soc. Behav.* **43**, 207–222 (2002).
- Diener, E. Subjective well-being. *Psychol. Bull.* **95**, 542–575 (1984).
- Ryan, R. M. & Deci, E. L. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annu. Rev. Psychol.* **52**, 141–166 (2001).
- Oparina, E. et al. Human wellbeing and machine learning. Preprint at <https://arxiv.org/abs/2206.00574> (2022).
- Wild, C. P. The exposome: from concept to utility. *Int. J. Epidemiol.* **41**, 24–32 (2012).
- Wild, C. P. Complementing the genome with an ‘exposome’: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
- van de Weijer, M. P. et al. Expanding the environmental scope: an environment-wide association study for mental well-being. *J. Expo. Sci. Environ. Epidemiol.* <https://doi.org/10.1038/s41370-021-00346-0> (2021).
- von Stumm, S. & d’Apice, K. From genome-wide to environment-wide: capturing the environome. *Perspect. Psychol. Sci.* **17**, 30–40 (2022).
- van de Weijer, M. P. et al. Capturing the well-being exposome in poly-environmental scores. *J. Environ. Psychol.* <https://doi.org/10.3389/fpsyg.2021.671334> (2024).
- Rutter, M., Kim-Cohen, J. & Maughan, B. Continuities and discontinuities in psychopathology between childhood and adult life. *J. Child Psychol. Psychiatry* **47**, 276–295 (2006).
- Lahey, B. B., Zald, D. H., Hakes, J. K., Krueger, R. F. & Rathouz, P. J. Patterns of heterotypic continuity associated with the cross-sectional correlational structure of prevalent mental disorders in adults. *JAMA Psychiatry* **71**, 989–996 (2014).
- Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J. & Wood, J. K. Predicting psychological and subjective well-being from personality: a meta-analysis. *Psychol. Bull.* **146**, 279–323 (2020).
- Chu, P., Sen, Saucier, D. A. & Hafner, E. Meta-analysis of the relationships between social support and well-being in children and adolescents. *J. Soc. Clin. Psychol.* **29**, 624–645 (2010).
- Mann, F. D., DeYoung, C. G., Tiberius, V. & Krueger, R. F. Social-relational exposures and well-being: using multivariate twin data to rule-out heritable and shared environmental confounds. *J. Res. Personality* <https://doi.org/10.1016/j.jrp.2019.103880> (2019).
- Uher, R. & Zwicker, A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry* **16**, 121–129 (2017).
- Klijns, B. et al. Neighborhood income and major depressive disorder in a large Dutch population: results from the LifeLines Cohort study. *BMC Public Health* **16**, 1–13 (2016).
- Generaal, E., Timmermans, E. J., Dekkers, J. E. C., Smit, J. H. & Penninx, B. W. J. H. Not urbanization level but socioeconomic, physical and social neighbourhood characteristics are associated with presence and severity of depressive and anxiety disorders. *Psychol. Med.* **49**, 149–161 (2019).
- De Vries, S. et al. Local availability of green and blue space and prevalence of common mental disorders in the Netherlands. *BJPsych Open* **2**, 366–372 (2016).
- Gong, Y., Palmer, S., Gallacher, J., Marsden, T. & Fone, D. A systematic review of the relationship between objective measurements of the urban environment and psychological distress. *Environ. Int.* **96**, 48–57 (2016).
- Yang, T., Wang, J., Huang, J., Kelly, F. J. & Li, G. Long-term exposure to multiple ambient air pollutants and association with incident depression and anxiety. *JAMA Psychiatry* **80**, 305–313 (2023).
- Liao, P., Shaw, D. & Lin, Y. Environmental quality and life satisfaction: subjective versus objective measures of air quality. *Soc. Indic. Res.* **124**, 599–616 (2015).
- Baselmans, B. M. L. et al. A genetic investigation of the well-being spectrum. *Behav. Genet.* **49**, 286–297 (2019).
- Thorp, J. G. et al. Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-021-01094-9> (2021).
- Kim, S. et al. Shared genetic architectures of subjective well-being in East Asian and European ancestry populations. *N. Hum. Behav.* **6**, 1014–1026 (2022).
- Meng, X. et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat. Genet.* **56**, 222–233 (2024).
- Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
- Routledge, K. M. et al. Shared versus distinct genetic contributions of mental wellbeing with depression and anxiety symptoms in healthy twins. *Psychiatry Res.* **244**, 65–70 (2016).
- Bzdok, D., Varoquaux, G. & Steyerberg, E. W. Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry* **78**, 127–128 (2021).
- Habets, P. C. et al. Multimodal data integration advances longitudinal prediction of the naturalistic course of depression and reveals a multimodal signature of remission during 2-year follow-up. *Biol. Psychiatry* <https://doi.org/10.1016/j.biopsych.2023.05.024> (2023).

30. Rutter, M. & Silberg, J. Gene-environment interplay in relation to emotional and behavioral disturbance. *Annu Rev. Psychol.* **53**, 463–490 (2002).
31. Dunn, E. C. et al. Genome-wide association study (GWAS) and genome-wide by environment interaction study (GWEIS) of depressive symptoms in African American and Hispanic/Latina women. *Depression Anxiety* **33**, 265–280 (2016).
32. Assary, E., Vincent, J. P., Keers, R. & Pluess, M. Gene-environment interaction and psychiatric disorders: review and future directions. *Semin. Cell Dev. Biol.* **77**, 133–143 (2018).
33. Abdellaoui, A. et al. Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* **3**, 1332–1342 (2019).
34. Kourou, K. et al. A machine learning-based pipeline for modeling medical, socio-demographic, lifestyle and self-reported psychological traits as predictors of mental health outcomes after breast cancer diagnosis: An initial effort to define resilience effects. *Comput. Biol. Med.* **131**, 104266 (2021).
35. Taliaz, D. et al. Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. *Transl. Psychiatry* **11**, 1–9 (2021).
36. Kearns, M. et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl. Psychiatry* **9**, 1–9 (2019).
37. Tate, A. E. et al. A Genetically informed prediction model for suicidal and aggressive behaviour in teens. *Transl. Psychiatry* <https://doi.org/10.1038/s41398-022-02245-w> (2022).
38. Macalli, M. et al. A machine learning approach for predicting suicidal thoughts and behaviours among college students. *Sci. Rep.* **11**, 1–8 (2021).
39. Yang, H., Liu, J., Sui, J., Pearson, G. & Calhoun, V. D. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Front. Hum. Neurosci.* **4**, 192 (2010).
40. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118 (2018).
41. Chilver, M. R., Champaigne-Klassen, E., Schofield, P. R., Williams, L. M. & Gatt, J. M. Predicting wellbeing over one year using sociodemographic factors, personality, health behaviours, cognition, and life events. *Sci. Rep.* **13**, 5565 (2023).
42. Runeson, B. et al. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. *PLoS ONE* **12**, e0180292 (2017).
43. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process Syst.* **30**, 6785–6795 (2017).
44. Snep, R. P. H., Klostermann, J., Lehner, M. & Weppelman, I. Social housing as focus area for Nature-based Solutions to strengthen urban resilience and justice: lessons from practice in the Netherlands. *Environ. Sci. Policy* **145**, 164–174 (2023).
45. Musterd, S. Public housing for whom? Experiences in an era of mature neo-liberalism: the Netherlands and Amsterdam. *Housing Studies* **29**, 467–484 (2014).
46. Hoekstra, J. Social housing in the Netherlands: the development of the Dutch social housing model. In *2nd Multinational Knowledge Brokerage Event’ Sustainable Housing in a Post-Growth Europe’* (Univ. Barcelona, 2013).
47. Clair, A. Housing: an under-explored influence on children’s well-being and becoming. *Child Indic. Res.* **12**, 609–626 (2019).
48. Burger, M. J., Morrison, P. S., Hendriks, M. & Hoogerbrugge, M. M. Urban-rural happiness differentials across the world. *World Happiness Rep.* **2020**, 66–93 (2020).
49. Hoogerbrugge, M. & Burger, M. J. in *Housing and Urban-Rural Differences in Subjective Wellbeing in The Netherlands* 97–118 (Edward Elgar Publishing, 2024).
50. Groenewegen, P. P., van den Berg, A. E., de Vries, S. & Verheij, R. A. Vitamin G: effects of green space on health, well-being, and social safety. *BMC Public Health* **6**, 1–9 (2006).
51. Gao, Y., Wang, Z., Liu, C. & Peng, Z.-R. Assessing neighborhood air pollution exposure and its relationship with the urban form. *Build. Environ.* **155**, 15–24 (2019).
52. De Vries, L. P., Baselmans, B. M. L. & Bartels, M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. *J. Happiness Studies* **22**, 2361–2408 (2021).
53. Henches, L. et al. Polygenic risk score prediction accuracy convergence. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.06.27.546518> (2023).
54. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: realizing the promise. *Am. J. Hum. Genetics* **110**, 179–194 (2023).
55. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
56. Mitchell, J. et al. Physical activity in young children: a systematic review of parental influences. *Early Child Dev. Care* **182**, 1411–1437 (2012).
57. Grey, E. B. et al. A systematic review of the evidence on the effect of parental communication about health and health behaviours on children’s health and wellbeing. *Prev. Med.* **159**, 107043 (2022).
58. Aalbers, M. B., Hochstenbach, C., Bosma, J. & Fernandez, R. The death and life of private landlordism: how financialized homeownership gave birth to the buy-to-let market. *Housing Theory Soc.* **38**, 541–563 (2021).
59. Baselmans, B. M. L. & Bartels, M. A genetic perspective on the relationship between eudaimonic –and hedonic well-being. *Sci. Rep.* **8**, 1–10 (2018).
60. Gallagher, M. W., Lopez, S. J. & Preacher, K. J. The hierarchical structure of well-being. *J. Pers.* **77**, 1025–1050 (2009).
61. *Healthy Environment, Healthy Lives—how the Environment Influences Health and Well-Being in Europe* (European Environment Agency, 2020).
62. Schmitz, O. et al. High resolution annual average air pollution concentration maps for the Netherlands. *Sci. Data* **6**, 1–12 (2019).
63. Richens, J. G., Lee, C. M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* **11**, 3923 (2020).
64. Ligthart, L. et al. The Netherlands twin register: longitudinal research based on twin and twin-family designs. *Twin Res. Hum. Genet.* **22**, 623–636 (2019).
65. Van Beijsterveldt, C. E. M. et al. The Young Netherlands Twin Register (YNT): longitudinal twin and family studies in over 70,000 children. *Twin Res. Hum. Genet.* **16**, 252–267 (2013).
66. Willemsen, G. et al. The Netherlands twin register biobank: a resource for genetic epidemiological studies. *Twin Res. Hum. Genet.* **13**, 231–245 (2010).
67. Willemsen, G. et al. The adult netherlands twin register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.* **16**, 271–281 (2013).
68. Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The satisfaction with life scale. *J. Pers. Assess.* **49**, 71–75 (1985).
69. Lyubomirsky, S. & Lepper, H. S. A measure of subjective happiness: preliminary reliability and construct validation. *Soc. Indic. Res.* **46**, 137–155 (1999).
70. Cantril, H. *The Pattern of Human Concerns* (Rutgers Univ. Press, 1965).
71. Cole, D. A., Martin, N. C. & Steiger, J. H. Empirical and conceptual problems with longitudinal trait-state models: introducing a trait-state-occasion model. *Psychol. Meth* **10**, 3–20 (2005).
72. Rosseel, Y. *Lavaan: an R package for structural equation modeling and more*. *J. Stat. Softw.* **48**, 1–36 (2012).

73. Devlieger, I. & Rosseel, Y. Factor score path analysis. *Methodology* **13**, 31–38 (2017).
74. Croon, M. in *Latent Variable and Latent Structure Models* (eds Marcoulides, G. and Moustaki, I.) 195–223 (Erlbaum, 2002).
75. Verstynen, T. & Kording, K. P. Overfitting to ‘predict’ suicidal ideation. *Nat. Hum. Behav.* **7**, 680–681 (2023).
76. Hu, L. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* **6**, 1–55 (1999).
77. Wray, N. R. et al. Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).
78. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
79. Timmermans, E. J. et al. Cohort profile: the geoscience and health cohort consortium (GECCO) in the Netherlands. *BMJ Open* **8**, e021597 (2018).
80. Lakerveld, J. et al. Deep phenotyping meets big data: the Geoscience and hEalth Cohort COnsortium (GECCO) data to enable exposome studies in The Netherlands. *Int. J. Health Geogr.* **19**, 1–16 (2020).
81. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien (R Forge, 2021).
82. Troyanskaya, O. et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525 (2001).
83. Mohammed, M. B., Zulkafli, H. S., Adam, M. B., Ali, N. & Baba, I. A. Comparison of five imputation methods in handling missing data in a continuous frequency table. In *AIP Conference Proceedings* vol. 2355 (eds. Phang, C. et al.) 40006 (AIP Publishing LLC, 2021).
84. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320 (2005).
85. Nogueira, S., Sechidis, K. & Brown, G. On the stability of feature selection algorithms. *J. Mach. Learn. Res.* **18**, 6345–6398 (2017).
86. Papini, S. et al. Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *J. Anxiety Disord.* **60**, 35–42 (2018).
87. Tate, A. E. et al. Predicting mental health problems in adolescence using machine learning techniques. *PLoS ONE* **15**, e0230389 (2020).
88. Field, C. A. & Welsh, A. H. Bootstrapping clustered data. *J. R. Stat. Soc. B* **69**, 369–390 (2007).
89. Jiang, Y., Lee, M.-L. T., He, X., Rosner, B. & Yan, J. Wilcoxon rank-based tests for clustered data with R package clusrank. *J. Stat. Softw.* **96**, 1–26 (2020).
90. Rosner, B., Glynn, R. J. & Lee, M.-L. T. The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* **62**, 185–192 (2006).
91. Benjamin, D. J. et al. Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2018).
92. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
93. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
94. Chen, T. & Guestrin, C. Xgboost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (eds. Krishnapuram, B. & Shah, M.) 785–794 (2016).

## Acknowledgements

D.H.M.P. is funded by an Amsterdam Public Health AI and Machine learning grant and ERC consolidation grant (WELL-BEING 771057, M. Bartels). The NTR data collection was supported by the following grants: NWO large investment grant (NTR: 480-15-001/674), ZonMW Addiction program (31160008), Spinozapremie (NWO/SPI 56-464-14192),

Twin-family database for behavior genetics and genomics studies (NWO 480-04-004), genetic influences on stability and change in psychopathology from childhood to young adulthood (NWO/ZonMW 91210020), Genetic and Family influences on Adolescent psychopathology and Wellness (NWO 463-06-001), A twin-sib study of adolescent wellness (NWO-VENI 451-04-034), The US National Institute of Mental Health as part of the American Recovery and Reinvestment Act of 2009: Genomics of Developmental Trajectories in Twins (1RC2MH089995-01), Determinants of Adolescent Exercise Behavior (NIH-1R01DK092127-01), and part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health (NIMH, MH081802). M.B. is funded by an NWO VICI grant (VI.C.211.054). Geo-data were collected as part of the Geoscience and Health Cohort Consortium (GECCO), which was financially supported by the Netherlands Organisation for Scientific Research (NWO), the Netherlands Organisation for Health Research and Development (ZonMw) and Amsterdam UMC. More information on GECCO can be found at [www.gecco.nl](http://www.gecco.nl). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank all Netherlands Twin Register participants who provided data for this study. GECCO (Geoscience and Health Cohort Consortium) is acknowledged for gathering and combining existing data into the GECCO repository and maintaining the infrastructure necessary for these data. We thank A. Wagtendonk in particular for providing the data for the present study.

## Author contributions

D.H.M.P. designed the study, with input from P.C.H., C.H.V. and M.B. D.H.M.P. analyzed the data, with support from P.C.H. for the machine learning models. D.H.M.P. designed the figures and tables and drafted the paper. L.L. and C.E.M.v.B. were responsible for providing and support with the NTR data, R.P. was responsible for the polygenic scores. All authors contributed to and approved the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s44220-024-00294-2>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44220-024-00294-2>.

**Correspondence and requests for materials** should be addressed to Dirk H. M. Pelt.

**Peer review information** *Nature Mental Health* thanks Elham Assary, Jurriaan Hoekstra and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence,

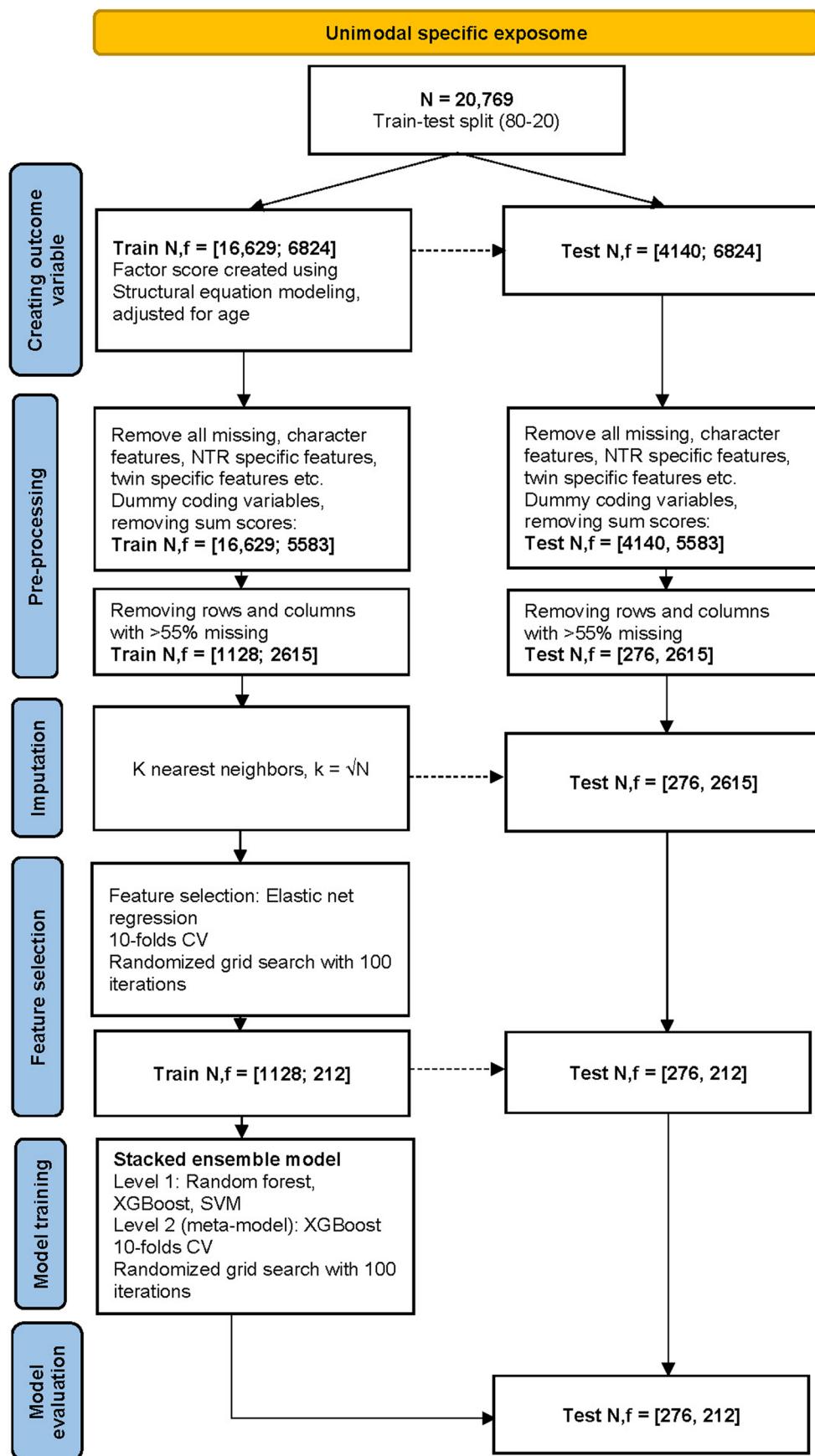
unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from

the copyright holder. To view a copy of this licence, visit  
<http://creativecommons.org/licenses/by/4.0/>.

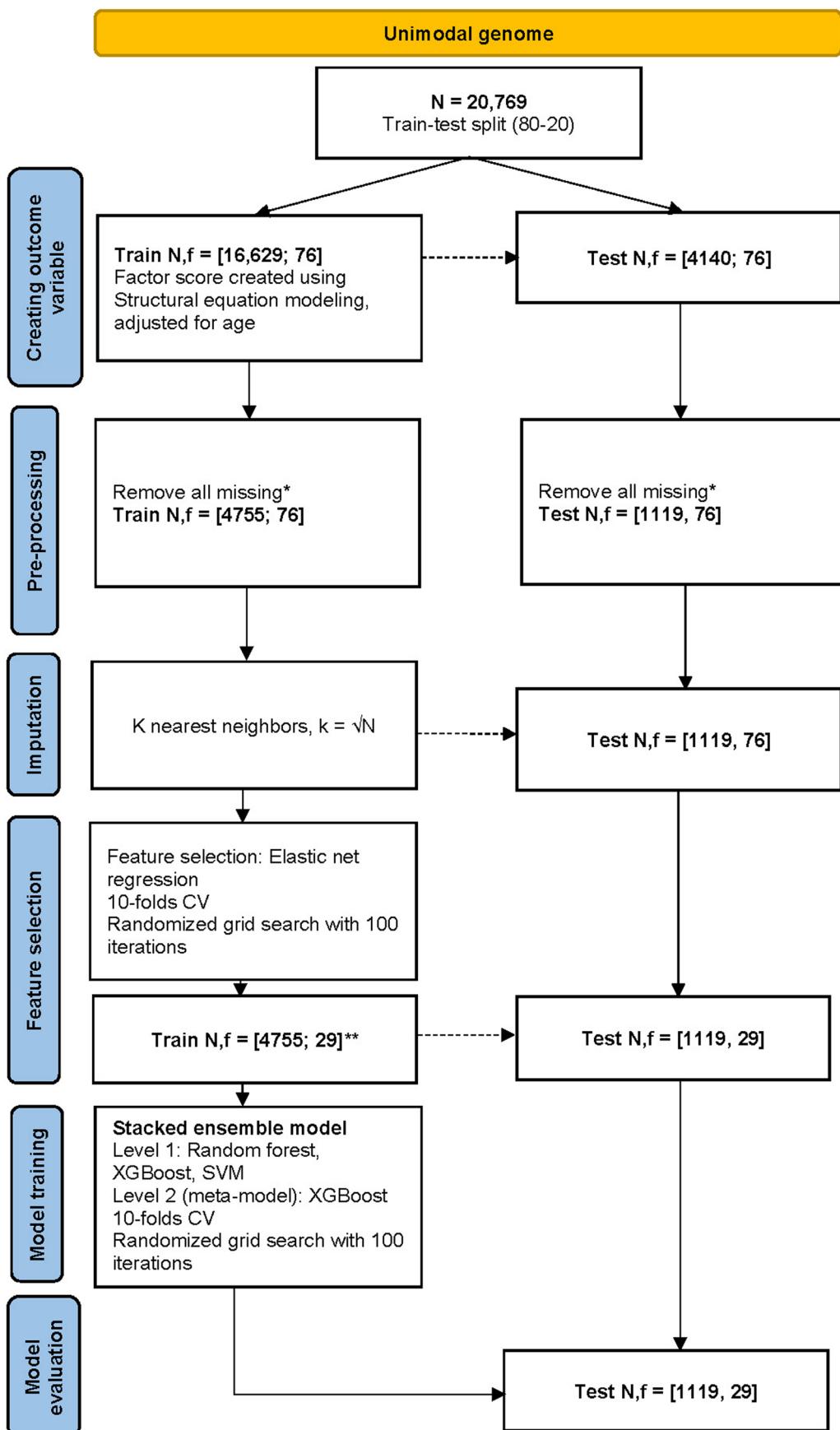
© The Author(s) 2024, corrected publication 2024

<sup>1</sup>Department of Biological Psychology, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

<sup>2</sup>Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Amsterdam, The Netherlands. <sup>3</sup>Department of Psychiatry and Anatomy and Neurosciences, Amsterdam University Medical Center location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>4</sup>Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep and Stress Program, Amsterdam, The Netherlands. <sup>5</sup>GGZ inGeest Mental Health Care, Amsterdam, The Netherlands. ✉ e-mail: [h.m.pelt@vu.nl](mailto:h.m.pelt@vu.nl)

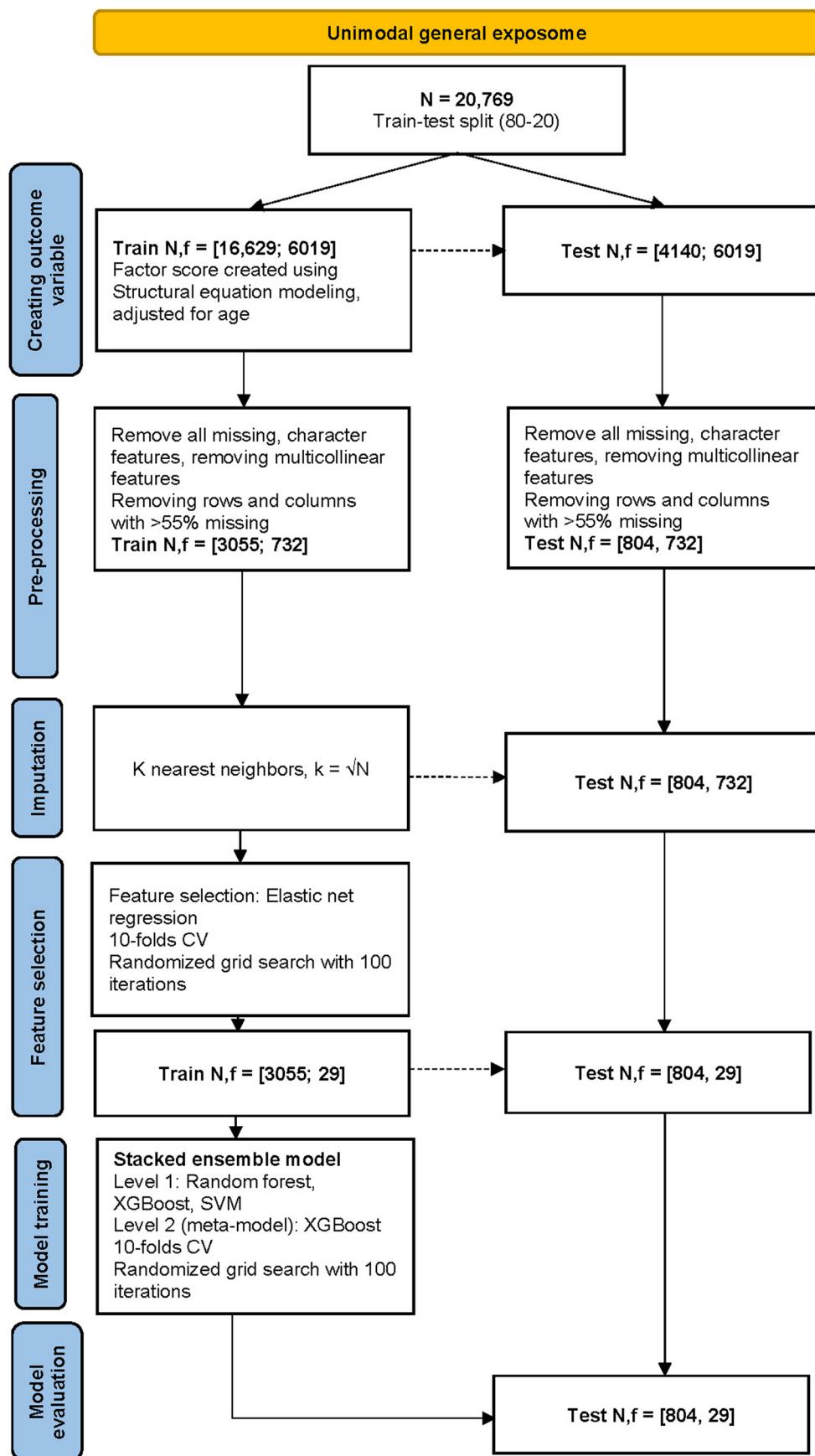


**Extended Data Fig. 1 | Flow chart of full data preparation and machine learning pipeline – unimodal specific exposome.** Note. N,f = sample size, number of features, XGBoost = extreme gradient boost, SVM = support vector machine. Dotted lines represent transformations/selections in test set based on train set.



**Extended Data Fig. 2 | Flow chart of full data preparation and machine learning pipeline – unimodal genome.** Note. N,f = sample size, number of features, XGBoost = extreme gradient boost, SVM = support vector machine.

Dotted lines represent transformations/selections in test set based on train set.\* Participants either had all or no genomic data available, \*\*13 polygenic scores, 10 principal components, 6 platform dummies.



**Extended Data Fig. 3 | Flow chart of full data preparation and machine learning pipeline – unimodal general exposome.** Note. N,f = sample size, number of features, XGBoost = extreme gradient boost, SVM = support vector machine. Dotted lines represent transformations/selections in test set based on train set.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection In more recent cohorts, data was collected through online survey tools (SurveyMonkey, NetQ).

Data analysis All pre-processing steps were conducted in the R environment (version 4.0.5), packages heavily used were tidyverse (2.0.0), lavaan (0.6-15), e1071 (1.7-13), and fastDummies (1.6.3). The figures were made using ggplot2 (3.4.3). LDpred 0.9 and plink2 were used for the construction of the polygenic scores. The machine learning analyses were conducted in Python using PyCharm in combination with Anaconda3, most notable with the scikit-learn (1.3.0) and XGBoost (1.7.6) packages. See <https://osf.io/zphw8/> for Python code including all loaded packages.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Being part of a national prospective cohort study, Netherlands Twin Register data cannot be made publicly available for privacy reasons, but they are available for

legitimate researchers via the data access procedure ([https://tweelingregister.vu.nl/information\\_for\\_researchers/working-with-ntr-data](https://tweelingregister.vu.nl/information_for_researchers/working-with-ntr-data)). Data of the Geoscience and health cohort consortium (GECCO) can be requested through the data access request form (<https://www.gecco.nl/exposure-data-1/>).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

### Reporting on sex and gender

In childhood, parent's reported on their child's gender by checking a box after the statements "Gender of oldest twin" ("Geslacht oudste van de tweeling") and "Gender of youngest twin" ("Geslacht oudste van de tweeling"), with the boxes labeled "boy" ("jongen") and "girl" ("meisje"). In adulthood, participants checked a box next to the statement "Gender" ("Geslacht"), with labels "male" ("man") and "female" ("vrouw"). These variables were entered as features in our machine learning models: gender specific analyses were not conducted because we were interested in building wellbeing prediction models for the entire population.

### Reporting on race, ethnicity, or other socially relevant groupings

Our phenotypic set of variables included questions on the birth country of the participants and the birth country of their parents. In all our models including genetic predictors (polygenic scores), we included the first ten principal components to correct for population stratification.

### Population characteristics

See below.

### Recruitment

At the start of the NTR, a commercial 'birth felicitation' bureau that visited parents of newborns at home and through city councils. Additional recruitment is done with the support of the Dutch Society of Parentof Multiples (Nederlandse Vereniging van Ouders van Meervelingen: NVOM; <https://www.nvom.nl>), through (online) newsletters, and events.

For more information on NTR recruitment, see: 10.1375/twin.5.5.401, 10.1017/thg.2019.93, 10.1017/thg.2012.118

### Ethics oversight

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration. Data collection was approved by the Central Ethics Committee on Research Involving Human Subjects of the University Medical Centers Amsterdam. Signed informed consent was obtained from all individual participants included in the study. Only participants who consented to record linkage were included. Participants were not compensated for participation.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

This study involves quantitative, longitudinal survey and genetic data (polygenic scores) of voluntary participants of the Netherlands Twin Register collected between 1991 and 2022, linked with environmental exposures at the postal code level provided by the Geoscience and hEalth Cohort COnsortium (GECCO).

### Research sample

The Netherlands Twin Register is a population-based cohort established in 1986. The NTR is a population-wide, non-clinical sample and distinguishes between the Young NTR (YNTR) and the Adult NTR (ANTR). The YNTR comprises of children rated by their parents, and standardized surveys are sent out around ages 1, 2, 3, 5, 7, 9, and 12. For several years (2004 - 2014), YNTR twins aged 14, 16, and 18 were invited for a self-report survey study. When participants in the YNTR reach the age of 18 years they are invited to take part in the ANTR. The ANTR consists of adult participants providing self-reports. The NTR collects DNA in their biobank. The sample was chosen for availability of wellbeing data, psychosocial data, genetic data, and possibility of record linkage. Gender distributions varied across the different datasets but were predominantly female [range 66%-73%]. Age ranged between 18 and 77. Participants were mostly higher educated. Full demographic information is reported in Table 1. The sample is not fully representative.

### Sampling strategy

Sample size was determined based on availability of data. As pre-registered, we requested all NTR variables with sample sizes  $N>3000$  to ensure a large enough initial sample size for each variable. For child data, only mother-reports were requested to maximize sample size. The NTR is not a random sample of the Dutch population (more female and higher educated participants, see also Table S2 on attrition). Rather, in the NTR, purposive sampling is used to recruit multiples and their family members. Once participants are recruited, sampling is based on availability. For the genetic data, criterion-based sampling is used: sufficient phenotypic (i.e., survey) data needs to be available before becoming eligible for genetic data collection.

### Data collection

In the ANTR, self-reports were paper and pencil until ~2009 (ANTR8), when surveys were partly distributed online as a pilot. Since then, data collection has gradually moved to fully online (present). In the YNTR, the surveys were paper and pencil until ~2016, after which parents filled in surveys about their children online. No researcher was present during any of the data collection. Since all data

has been collected before the design of this study, all researchers involved with data collection were blind to the study hypotheses.

#### Timing

The childhood waves were collected from 01/90 until present (periodically), the adolescent waves between 05/04 and 12/13. The adult waves were collected between 01/09 and 12/12 (ANTR8), between 05/13 and 12/15 (ANTR10), and between 06/19 and 02/20 (ANTR14). For more information on NTR study waves, see: 10.1017/thg.2019.93, 10.1017/thg.2012.118

#### Data exclusions

For the outcome variable (wellbeing), we excluded participants younger than 18 years old ( $N = 129$ ) because we were interested in predicting adult wellbeing. Those with only one single wellbeing measurement available were also excluded ( $N = 1,094$ ). Participants and features with more than 55% missing values were excluded (Unimodal specific exposome:  $N_p = 19,365$ ,  $N_f = 2,968$ ; Unimodal genome:  $N_p = 14,895$ ; Unimodal general exposome:  $N_p = 16,910$ ,  $N_f = 5,287$ ). We pre-registered 50% but this would have reduced our sample sizes too much.

#### Non-participation

In general, the response rate in the NTR is estimated to be between 40% and 50% (see 10.1017/thg.2012.140 and 10.1017/thg.2012.118).

#### Randomization

This study did not involve random assignment of participants to experimental groups. All important covariates (age and gender) were included in the analyses.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involved in the study                                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants                        |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |